VOLUME 46, NUMBER 6

June, 1958

published monthly by The Institute of Radio Engineers, Inc.

Proceedings of the IRE®

contents

published monthly by The Institute of Radio Engineers, Inc.

Proceedings of the IRE®

continued

COVER

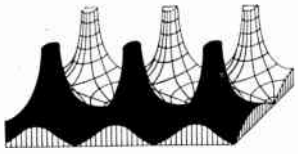
A junction transistor, flanked by its graphical symbol and basic operational equations, signify the special nature of this issue, published on the tenth anniversary of the transistor.

PROCEEDINGS OF THE IRE, published monthly by The Institute of Radio Engineers, Inc., at 1 East 79 Street, New York 21, N. Y. Manuscripts should be submitted in triplicate to the Editorial Department. Responsibility for contents of papers published rests upon the authors, and not the IRE or its members.
Fifteen days advance notice is required for change of address. Price per copy: members of the Institute of Radio Engineers, one additional copy $1.25; non-members $3.00. Yearly subscription price: to members $9.00, one additional subscription $13.50; to non-members in United States, Canada, and U. S. Possessions $18.00; to non-members in foreign countries $19.00. Entered as second class matter, October 26, 1927, at the post office at Menasha, Wisconsin, under the act of March 3, 1879. Acceptance for mailing at a special rate of postage is provided for in the act of February 28, 1925, embodied in Paragraph 4, Section 412, P. L. and R., authorized October 26, 1927.

June, 1958
Vol. 46 No. 6

Proceedings of the IRE

Poles and Zeros



We Do It Again. This page is the introduction to another IRE PROCEEDINGS Special Issue. These tomes have now, we hope, achieved such a reputation that were the covers not already a healthy IRE blue, we would be justified in binding them in a truly royal purple to signify the special contents. While continuing to pat ourselves on the back, may we also point out to all bill-paying members that the January Special Issue on Radio Astronomy contained 402 editorial pages, and here is another of equivalent size and import in the same year, yet.

The subject selected for this issue—Transistors—made choice of the date of issue automatic. June, 1958 marks the tenth anniversary of the first announcement of the invention of the transistor by the Bell Telephone Laboratories. Production of these devices will this year exceed 45 million in the United States; in 1916, ten years after the invention of the vacuum triode, the tube was still far from any form of quantity production. The transistor cannot take credit for the rapidity of its own development; it arrived in an era already catalyzed by prior vacuum tube art. It still speaks well for the breadth of scientific training of the electronic engineer that, rather than being put out of business by the transistor, he has taken it in stride, learned to think of energy levels and holes and active circuits, and has made the transistor the basis for a whole new era in electronics.

Many of our members may not be aware of the steps undertaken in the conception, gestation, and birth of a Special Issue. Certainly these brain children do not just arrive by stork nor do they burgeon as a result of fortuitous circumstances. They are actually the result of a great amount of work by a great many people, and the occasion of the birth of this latest Brobdingnagian infant seems propitious to review the procedure and give a nod of thanks to various hard working people along the way.

Conception of this issue occurred during an early 1957 meeting of the IRE Editorial Board, where it was agreed that progress since the November, 1952 issuance of Special Issue Transistor I amply warranted another compendium of scientific and engineering thought on the subject. Experience has shown that the very necessary first step in the process of gestation is to identify a sponsoring group and more particularly, an organizer of the issue or an issue editor. In the case of this number the joint IRE-AIEE Committee on Semiconductor Devices undertook the sponsorship, and its chairman, Stephen J. Angello, of the Westinghouse Research Laboratories, became the issue editor. He and his committee, which operates within the IRE as a subcommittee of the IRE Committee on Solid-State Devices, planned the contents of the issue to provide appropriate coverage of the field, and to insure that the issue would conform to requirements as a classic reference in the field, as had Transistor I.

Now it is a well recorded fact in the publishing field that all authors will submit their papers on time, and there are only infrequent delays due to typing difficulties or the mails which delay the arrival of papers beyond the promised dates, or even beyond the editor's true expectations. In such few isolated cases editors have been known to write letters of considerable solicitude fearing for the author's future state of health and general fitness, always couched in language which remains polite, or at least mailable. Finally, somehow, the desired papers are in hand and the presses roll, bringing us to the glorious natal day of another Special Issue—Transistor II.

Such an issue could not properly be introduced without a few words from the transistor's co-inventors and Nobel Prize winners, Shockley, Bardeen, and Brattain, and so the lead articles go to them. There follow 37 additional papers covering the field as of today.

To Steve Angello, James Early, and the others who contributed, many thanks for a fine job. To those who wrote, and wrote, and did meet the deadlines, no matter how many times removed—more commendations. In the winnowing process 20 other papers fell by the wayside, but these authors also contributed because they made the first team work the harder, and their time will come. Rejection of 20 papers from a field of 57 submitted is a good score these days. Would that some other journals could also make that claim.

Engineering Surplus? Since 1951 our eyes and ears have reacted to the ceaseless repetitions of the engineering shortage. In publicizing the situation the ivory hunters have all too often used a shot gun rather than a rifle, and have attracted to the engineering colleges a considerable percentage of ill-prepared and non-adapted students, rather than the selectively chosen students of high intellect who will contribute to the advancement of our scientific world.

We are now seeing the other team prowling the jungle—those using similar publicity devices and statistics to prove there is now a surplus of engineers, and we are going to see the results of their work in the next several years' input to our engineering schools.

The subject of guidance of qualified students into engineering careers is a psychological and public relations matter, and we feel it has been handled without sufficient recourse to experts in those fields. We hope that we have not thereby created a system whose damping constant is less than unity.

—J.D.R.

William H. Doherty

Director, 1958-1960

William H. Doherty was born on August 21, 1907 at Cambridge, Mass. A student at Harvard University, he received its B.S. degree in electric communication engineering in 1927, and the M.S. degree in engineering the following year.

His early work at Bell Laboratories began in 1929 and was in high-power transmitter development for transoceanic telephone service and broadcasting. This work led to the invention of a high-efficiency power amplifier now extensively used in broadcasting, for which the IRE awarded him the Morris Liebmann Memorial Prize in 1937.

Later he participated in pioneering work in the fire-control radar field, and throughout World War II supervised a development group which was responsible for the design of a number of radars used on naval surface ships and submarines for gunfire and torpedo control.

Mr. Doherty continued in military electronics work until 1949, when he became Director of Electronic and Television Research. He was appointed Director of Research in Electrical Communications in 1951 and continued in that post until he went to the American Telephone and Telegraph Company in 1955 as an Assistant Vice-President. Last year he returned to Bell Telephone Laboratories as Assistant to the President.

On June 1, 1958, Mr. Doherty assumed the post of Manager of Government Sales with the Radio Division of Western Electric Co., Inc.

He is a member of Tau Beta Pi and Sigma Xi, and he holds an honorary Doctor of Science degree from the Catholic University of America. He also is President of the Harvard Engineering Society.

Mr. Doherty joined the IRE as an Associate member in 1929. He became a Member in 1936, a Senior Member in 1943, and Fellow in 1944. He served on the following IRE Committees: Editorial Review, during 1954–55; Membership, from 1943 to 1946; Nominations, during 1952–53; Policy Development, in 1951; Policy Advisory, during 1952–53; and Professional Groups, from 1948 to 1951. He was chairman of the Nominations Committee in 1953, and an IRE Director from 1951 to 1953.

Scanning the Transistor Issue*

STEPHEN J. ANGELLO†, SENIOR MEMBER, IRE

THIS special issue of PROCEEDINGS commemorates the tenth anniversary of the announcement of the invention of the transistor by the Bell Telephone Laboratories. Since the publication of the first technical news of this invention,[1] the field has expanded truly high and wide. Diffused base transistors made at Bell Laboratories are oribiting around the earth in the Explorer and Vanguard satellites, and there is scarcely a country in the world not making some kind of effort on transistor and semiconductor device development.

The most exciting thing about the invention is that the operation of the transistor depends upon a new concept in the physics of the solid state. It was known for many years before the transistor that solids can be found which conduct electricity by positive electronic charge carriers as well as by negative electrons. There is some interesting historical reading in the development of the concept. Experimenters who normally were very careful made mistakes in determining the sign of the Hall effect, because it was hard to believe that an electron could have a positive charge. About 1939, W. Schottky in Germany gave a complete theoretical discussion of the equilibrium of electrons and holes in semiconductors, particularly in the region of a semiconductor adjacent to a large area metal contact.

W. H. Brattain and J. Bardeeen discovered that a metal point upon a germanium crystal when carrying forward current could influence the reverse current in a similar point contact nearby. This was interpreted as injection of minority current carriers from the metal into the germanium. These injected charge carriers were collected by the other probe, hence the names "emitter" and "collector" which are still in general use. Since the collector had to be close to the emitter, it was evident that the emitter influence region was small. These phenomena were interpreted correctly by W. Shockley who proceeded to develop a new theory of conduction in semiconductors of the germanium (valence crystal) type, introducing the concept of finite minority carrier density above the equilibrium density and the associated lifetime. Furthermore, he postulated a *p-n* junction rectifier and a type of transistor which was two *p-n* junctions separated by a narrow base region. This was a radical departure from the original point-contact device, and many months passed before the theory was checked by experimental devices.

The Solid-State Devices Committee of the IRE and AIEE which organized this issue felt it proper to honor the co-inventors of the transistor. To this end we have asked each gentleman to write a short essay, and these are given first place in the body of this issue. We delight in paying tribute to the superb experimental technique of Dr. Brattain, the deep theoretical understanding of Dr. Bardeen, and the creative genius of Dr. Shockley.

Our appreciation of this discovery will be enhanced if we imagine for the moment that the lifetime of an excess of minority charge carrier density is essentially zero. We would be restricted then to the phenomenon of a charge carrier density depletion layer in the boundary region adjacent to a metal or semiconductor contact. We dare not presume to name all possible devices which can be created with the space-charge depletion layer as a basis, but we can enumerate devices known at this time. This list may be compared with the existing devices based upon minority carrier injection with useful lifetime. The size and importance of the latter list makes the point of this discussion, namely, that the discovery of the transistor introduced a fertile concept into solid-state device development. Not only is the left-hand list longer, but also the commercial market for devices in this list eclipses the right-hand list. Only metal-semiconductor rectifying cells have been important in the depletion layer type devices, and this market is being decimated by *p-n* junction rectifying cells.

TABLE I

COMPARISON OF DEVICES DEPENDING UPON SPACE-CHARGE DEPLETION LAYER AND MINORITY CARRIER INJECTION

Minority Carrier Injection	Space-Charge Depletion Layer
p-n junction diodes	Metal-semiconductor rectifying cells
p-n-p and *n-p-n* junction transistors	Field effect transistor
Junction tetrode transistor	Field effect resistor
p-n-i-p and drift types	Variable capacitance diode
Double base diodes	Voltage regulator diode
p-n-p-n diode	
p-n-p-n triode	
Stepping transistor	

Requests have come to us from a number of people prominent in semiconductor research and development to take this opportunity to reaffirm a basic truth. This is that all device development is dependent upon the properties of the starting materials which are available. Therefore, device development ought to be coupled with a vigorous research and development program upon basic materials. Consider again that the transistor depends upon the minority current carrier lifetime. The discovery of the transistor was made only when the careful metallurgical work of W. G. Pfann had resulted in germanium of sufficient quality to show the

* Original manuscript received by the IRE, April 29, 1958.
† Westinghouse Research Labs., Pittsburgh 35, Pa.
[1] J. Bardeen and W. H. Brattain, "The transistor, a semiconductor triode," *Phys. Rev.*, vol. 74, pp. 230–231; July 15, 1948.

necessary conduction effects. Diffusion, alloying, etching, magic surface coatings, point contact forming, or any other device fabrication technique is of no avail to produce a transistor if the basic material is not of sufficient quality. Perhaps other exciting basic discoveries await careful research upon new semiconducting materials. Along this line, we believe that the field would benefit if someone who was closely associated with the Bell Laboratories' transistor discovery and development would describe the historical events in detail. We already know from Dr. Bown's preface to Dr. Shockley's classic text that there was a vigorous program underway with the broad aim of developing a solid-state amplifier. We suspect that their success was related to the close relation between basic materials research and device development. The results of device development can feed back valuable information to materials research to give them a *raison d'être* and to guide them into fruitful channels of thought and experiment.

Before scanning the papers of the issue in detail, it is fitting to explain that there are sixteen invited papers. The honor to the authors of these papers is enhanced by the method by which they were chosen. A letter survey was conducted asking nationally prominent men in the field to suggest authors for topics we considered desirable to cover in this issue. In most cases the authors were selected by popular vote in this survey. Our list of topics also was enhanced by the same survey. Many thousands of man hours have gone into the preparation of this issue and we hope it will serve the field for many years to come as an important reference work.

The first invited paper following the essays of Brattain, Bardeen, and Shockley is by J. A. Morton and W. J. Pietenpol. They have surveyed the technological impact of the transistor during the past decade. In an interesting narrative they trace the development of transistor types and applications. This article is required reading for the executive responsible for the guidance of solid-state development in his organization.

Although the commercial exploitation of the transistor has been mainly with germanium and silicon, we have asked Dietrich Jenny to survey the interesting results which have been obtained in the III–V compounds and other compound semiconductors. These new materials already have resulted in a new device feature —the wide-band gap emitter. This summary will serve to orient anyone interested in the future possibilities of this field.

There are other semiconductor devices in addition to transistors and diodes, and S. J. Angello has surveyed these with the idea of organizing the large number known by certain logical classifications. Criteria for "important" devices are attempted and descriptions and references are given for some devices which we predict will become important.

Dr. Shockley was invited to contribute an original paper, and appropriately he chose to push back further the frontier of knowledge of "Electrons, Holes, and Traps." This knowledge will be useful because certain devices such as the four-layer diode require trap level control in fabrication.

Because of the importance of recombination in device technology, we asked G. Bemski to prepare a tutorial article on this subject. The paper covers theory, basic experimental methods, and discussion of results.

Two experts were asked to survey the difficult field of electrical noise. K. M. van Vliet and A. van der Ziel divided the field between them into bulk semiconductors and junctions, respectively. It is our belief that these papers will serve as the textbook on electrical noise in semiconductors for a long time in the future.

An important source of recombination centers is the dislocations in crystals caused by neutron irradiation. G. C. Messenger and J. P. Spratt have contributed a paper which gives a theoretical discussion of the effect of neutrons upon the grounded emitter current gain of a transistor. Observed changes in transistor parameters are explained and some basic quantities of the recombination theory for germanium are given.

Another variation of the general theme of recombination is provided by a paper contributed by R. Gremmelmaier on irradiation of a p-n junction by γ rays. The diffusion length of minority carriers can be estimated by the effects of irradiation and results are given for Si, GaAs, and InP.

As Morton and Pietenpol state, the development of diffusion techniques in silicon and germanium was an important breakthrough in the technique of semiconductor device fabrication. F. M. Smits, who has contributed much to the development of this subject, was asked to summarize the state of the art. He has favored us with a very complete and valuable paper.

Major techniques such as diffusion are always fruitful sources for valuable modifications. Herbert Nelson has contributed one such modification in a paper on preparation of semiconductor devices by lapping and diffusion techniques.

Another important variation of the original diffusion technique is "outdiffusion." J. Halpern and R. H. Rediker describe this technique for the fabrication of narrow base germanium computer diodes. Feasibility of the process has also been shown for n-p-n graded-base transistors.

We proceed to the very important topic of the single p-n junction. John Moll has responded to our invitation and provided us with an historical survey of the theory of the p-n junction. This will be a reference paper for students entering the field.

Some extensions of the theory involve complicated depletion layer configurations. L. J. Giacoletto has contributed an analog model solution to such problems which is an aid to visualization.

Only a few years passed from the theoretical description of p-n junctions to commercial application of silicon and germanium rectifying cells. H. Henkels was asked to summarize the state of the art with respect to these

devices. The result is a rich mine of device development information, part of which was generated by Dr. Henkels' original work.

At the low-power end of the scale *p-n* junction devices often are referred to as "diodes." Recently, some very interesting developments have been carried out upon diodes in microwave systems. Of special interest is the feasibility of obtaining power gain. We asked A. Uhlir, Jr., who has contributed much original work to this subject, to prepare a status report. The number of applications possible for the new *p-n* junction diode which are found in this paper will surprise those who have not had occasion to follow developments in this field.

A contribution by G. C. Messenger rounds out the discussion of diodes by describing new approaches to extend the frequency capability of radar crystal mixers and to improve sensitivity.

A special use of diodes is in detection of light. D. E. Sawyer and R. H. Rediker have made noteworthy contributions with respect to speed of response of such diodes. Their paper describes the operation of 75-mμsec germanium diodes.

The natural extension from single *p-n* junctions is the configuration of two *p-n* junctions separated by a narrow base region. R. L. Pritchard has been invited to survey the progress of *p-n* junction triode theory to which he has contributed much. This paper is one reason why we feel that this transistor issue will be a reference work of some importance.

Electrical engineers often feel that they have a better understanding of a theoretical model if they can form equivalent circuits which approximate the theory. J. G. Linvill has made a contribution to this topic which may appeal to many device engineers.

One of R. L. Pritchard's original contributions to transistor theory is presented in his paper on current flow in junction transistors at high frequencies. He extends the theory to a two-dimensional model.

Having covered the theory of *p-n* junction triodes, we move on to the technology for providing real representations of theoretical models. An alloy-diffusion method for high-frequency germanium transistors is described by P. J. W. Jochems, *et al.* C. C. Thornton and J. B. Angell contribute to the technology of micro-alloy diffused transistors. D. E. Thomas and J. L. Moll have contributed a very interesting paper concerned with measurement of transistor characteristics. They prove that a junction transistor fits a certain general network theorem given by Bode. This enables complicated frequency characteristics to be determined by relatively simple measurements.

A special class of transistors which is enjoying increasing application in industry is the group capable of controlling power. We have asked M. A. Clark to survey this field. Workers interested in design theory and practice in the areas of high current densities and high voltage will find much valuable information here.

Two contributed papers concerned with thermal resistance measurements upon transistors are pertinent to power transistors and the general problem of rating transistors. These are by B. Reich, and J. T. Nelson and J. E. Iwersen.

Extension of power transistors to high frequency is highly desirable. J. T. Nelson *et al.* describe a 5-watt 10-megacycle transistor which is fabricated in silicon by the diffusion technique.

Emeis and Herlet have checked parts of the theory of power transistors by fabricating an extensive series of alloy transistors with a wide range of basic material parameters. Measurements upon these transistors are compared with theory. Emeis, Herlet, and Spenke describe an alloyed power transistor of cylindrical geometry which provides an emitter and base with very long perimeter. This perimeter can be varied in length, and further checks of power transistor theory are obtained.

The papers now progress into more complicated transistor structures with a paper contributed by I. M. Mackintosh on *p-n-p-n* triodes. These devices are rapidly becoming important in switch-type application. A complementary paper on this same subject is contributed by R. W. Aldrich and N. Holonyak, Jr.

Computers represent a very important use of transistors which will become more important as time progresses. R. A. Henle and J. L. Walsh were asked to survey the kinds of application of transistors in computers. Their review will be found very educational by those who wish to be brought up to date in this field.

It is probably fair to say that there is not a facet of electronics circuits which has not been affected by transistors. In an invited survey, D. D. Holmes describes one of these facets: the field of commercial communication equipment.

Two papers were contributed which are strictly circuits design papers. These are by J. J. Suran on the analysis and design of transformerless pulse generators, and by D. F. Page, who discusses oscillator circuit design and a unified approach to the design of instability.

Finally, a popular demand prompted us to ask Esther Conwell to bring her 1952 Transistor Issue paper up to date to provide the field with an accurate compendium of silicon and germanium properties. In addition to tabular data, she has supplied a lucid explanation of the concepts applicable to this field.

I wish to take this opportunity to thank certain persons who contributed much to the preparation of this issue. It was a pleasure to work with E. K. Gannett, the Managing Editor, and we acknowledge his cordial cooperation. James E. Early acted as co-organizer and contributed many valuable ideas to the planning of the issue.

The members of the IRE-AIEE Solid-State Devices Committee provided an able team of reviewers to supplement the normal IRE Review Board. Credit for any success which this issue might achieve belongs directly with the authors whose papers are recorded herein.

Comments on Implications of Transistor Research*

JOHN BARDEEN†

IN looking back over the past ten years of research associated with transistor development, I am most impressed by the tremendous strides made in our understanding of the effects of minute physical imperfections on the physical properties of solids. In large part, this is an unanticipated by-product, made possible by the production and control of nearly perfect single crystals of germanium, silicon, and related materials.

This control has not only made it possible to design and fabricate devices with remarkable properties but has greatly aided the scientific study of crystal imperfections.

Ten years ago we felt that prospects for commercial development of transistors were excellent, even though great problems had to be overcome. However, both the large scale of the effort required to achieve this and the great degree of success which has been obtained in overcoming frequency, power, and other limitations were not foreseen, at least by me. The junction transistor has of course played a key role in these developments.

Many properties of solids of scientific and technological interest other than semiconductivity are dependent on impurities or other imperfections present in parts per billion or less. These include, for example, photoconductivity, luminescence, and even plastic flow. Controlled very perfect single crystals have provided, in a sense, a laboratory for study of such processes. Research on the physics of surfaces has also been stimulated. Advances in theory have kept pace with experiment so that in many cases a detailed quantitative understanding is possible. Thus research stimulated by transistor applications is having an impact in many other areas of science and technology.

The growth of solid-state physics, already well underway when the transistor was discovered, has undoubtedly received a strong impetus. There has been rapid expansion in both the number of laboratories doing research in this area and the number of scientific publications.

We were very lucky to be in on a major discovery of the sort that cannot be predicted in advance. But it is the cumulative advance in scientific understanding, made by many people in many countries, which has made it possible to realize the potentialities of semiconductor devices and to make striking progress in many other areas dependent on crystal imperfections.

* Original manuscript received by the IRE, April 29, 1958.
† Dept. of Physics, University of Illinois, Urbana, Ill.

Essay on the Tenth Anniversary of the Transistor*

WALTER H. BRATTAIN†

AS far as I am concerned the advent of the transistor is one of those developments that come along when general background knowledge has developed to such a stage that human minds are prepared to take a new step in the understanding of phenomena that have been under observation for a long time. In the case of a device with such important consequences to technology it is notable that the break-through came from work dedicated to the understanding of fundamental physical phenomena rather than from the empirical cut and try efforts of producing a useful device.

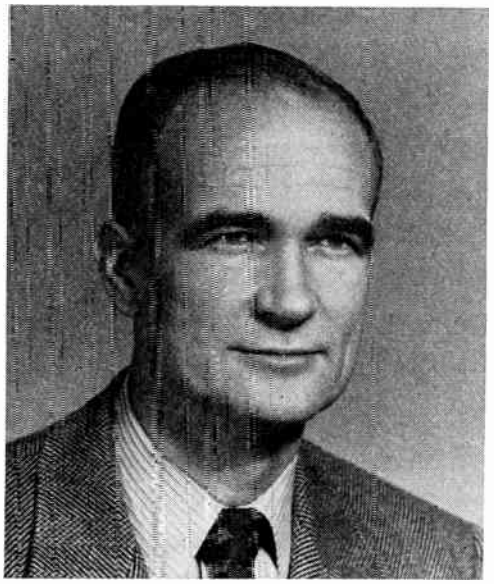
This general understanding of the physics of solids, semiconductors in particular, was dependent on the development of quantum mechanics. To understand how electrons and nuclei react together to form the aggregates we call solid crystals we had to understand first how they react individually to form atoms. The type of solid binding most characteristic of the semiconducting crystals—the covalent bond—was just the type that can only be understood from a quantum mechanical standpoint.

The picture of a semiconductor as a medium in which electrons and holes, as dissociated covalent bonds, can exist in thermodynamic equilibrium was the necessary conceptual idea. It is closely analogous to the quantum mechanical concept of electrons and positrons in vacuum. Another close analogy is to the theory of dissociation in electrolytes. On hindsight one wonders why this essential concept was not grasped sooner. Why did physicists and chemists have to wait until they could understand "defect" as well as "excess" conductivity from first principles before seeing the light?

In conclusion, I would like to point out that while a well thought out experiment may always give good results, nevertheless the really important experiment is the one that leads to new and unexpected results regardless of the original reason or expectation that inspired it. While lightning only strikes occasionally one should be cautious in discouraging the urge to do an experiment no matter how hazy the reason or how sure one is that the result may be trivial.

* Original manuscript received by the IRE, May 5, 1958.
† Bell Telephone Labs., Murray Hill, N. J.

An Invited Essay on Transistor Business*



WILLIAM SHOCKLEY†, FELLOW, IRE

IN 1950, I finished writing a book[1] by placing at the end of the last chapter a prediction about the future of transistor electronics. In the interest of establishing my position as a prophet of transistor business, I shall quote this prediction:

> "It may be appropriate to speculate at this point about the future of transistor electronics. Those who have worked intensively in the field share the author's feeling of great optimism regarding the ultimate potentialities. It appears to most of the workers that an area has been opened up comparable to the entire area of vacuum and gas-discharge electronics. Already several transistor structures have been developed and many others have been explored to the extent of demonstrating their ultimate practicality, and still other ideas have been produced which have yet to be subjected to adequate experimental tests. It seems likely that many inventions unforeseen at present will be made based on the principles of carrier injection, the field effect, the Suhl effect, and the properties of rectifying junctions. It is quite probable that other new physical principles will also be utilized to practical ends as the art develops."

* Original manuscript received by the IRE, April 28, 1958. This research has been supported in its entirety by E. L. Shockley, who has contributed major portions of one weekend towards its completion.

† Shockley Semiconductor Lab., Beckman Instruments, Inc., Mountain View, Calif.

[1] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y.; 1950.

Now for the future! In the course of carrying out the research for this essay, I have discovered a new and significant law of growth in the transistor field. This is the Shockley frequency-production index. In effect, it is a measure of the simplicity of the transistor, compared to the man. Men produce transistors, and transistors in their purest action (sinusoidal oscillation) produce cycles. The ratio is the frequency-production index and is defined by

$$f_{\alpha co}/P_y \equiv I_f P \tag{1}$$

where $f_{\alpha co}$ is the alpha cutoff frequency of the highest frequency transistor produced in a given year, and P_y is the volume of transistor production. The value of the ratio, when expressed in units of years/sec, lies between 50 and 100 as may be seen from Table I.

TABLE I

Year	$f_{\alpha co}$ sec^{-1}	P_y year^{-1}	$I_f P$ year/sec
1949		10^4	
1950	7×10^6	10^5	70
1951		2×10^5	
1952	5×10^7	4×10^5	100
1953		6×10^5	
1954	1.2×10^8	1.3×10^6	90
1955		3.5×10^6	
1956	8×10^8	1.3×10^7	62
1957	1.5×10^9	2.9×10^7	50
1958	5×10^9	7×10^7	70
1959	10^{10}	1.5×10^8	70
1960	1.7×10^{10}	2.5×10^8	70

The frequency-production index can be expressed in dimensionless form and is then the number of oscillations at the alpha cutoff frequency of the highest frequency transistor during the average time required to produce a transistor in the U.S.A. In dimensionless form I_fP is 2×10^{11} cycles per unit. This measures the simplicity of the transistor compared to man.

Since transistor production depends on years and men, I have used in Table I the well established ratio[2] of 70 between them in predicting frequency from volume and frequency-production index. The production estimates are based on conventional methods of estimating growth.

[2] This number will be recognized as 3 (10) in the customary score system.

No business essay is complete without a reference to money. Assuming that the cost per unit varies as $P_y^{-0.33}$, my estimate of average transistor prices for 1958, 1959, and 1960, is \$1.80, \$1.40, and \$1.20. The corresponding sales volumes are \$125, \$210, and \$300 million.

In closing this essay, I should like to acknowledge the assistance that has made it possible to carry out the extensive research involved. The significant data published by the Electrical Industries Association[3] has been essential. Special thanks are due to the encouragement and assistance of my friends at Bell Telephone Laboratories; without it, the completion of the investigation would have been difficult.

[3] *Wall St. J.*, p. 11; October 14, 1957.

The Technological Impact of Transistors*

J. A. MORTON†, FELLOW, IRE, AND W. J. PIETENPOL†, MEMBER, IRE

Summary—**During the past ten years the transistor has invaded every phase of the electronics industry. Its important features are its high efficiency at low power levels, its reliability, and its potential low cost.**

Presented here are the major milestones and problems which were overcome and led to devices that cover a broad field of electronic technology, and to the growth of a new industry.

ELECTRONICS is an increasingly important part of modern technology. It merits this distinction because it pervades our economy as water does a sponge. Since electronics has to do with the high speed transmission and processing of information, it becomes a wonderful extension of man's mind.

Electronics has the potential of affecting every aspect of our modern industrial world through communication, entertainment, transportation, power, manufacturing, and business. If we can measure the transistor's effect in expanding the breadth and versatility of electronics, by imaginative implication, we can define its impact on the whole of modern technology.

The bulk of present day electronics has to do mostly with the *transmission* of information. In some 45 years it has grown from the original deForest vacuum triode and modulation theory to the fifth largest American industry comprised largely of consumer, military, and industrial transmission functions.

* Original manuscript received by the IRE, March 5, 1958.
† Bell Telephone Labs., Inc., Murray Hill, N. J.

For some time now, through applications of information and switching theory to functions such as pulse-code modulation, memory, and logic, electronic man has known how *in principle* to extend greatly his visual, tactile, and mental abilities to the digital transmission and processing of all kinds of information. However, all these functions suffer from what has been called "the tyranny of numbers." Such systems, because of their complex digital nature, require hundreds, thousands, and sometimes tens of thousands of electron devices. The large amount of power used inefficiently and the high cost of reliability of the electron tube have prevented these expansions of electronics in all but a few cases where the high cost could be tolerated, even though not desired.

This is where the transistor comes in. The really important aspects of the transistor are its very high-power efficiency at low-power levels and its potential reliability. Because of its relatively simple mechanical features, it is potentially low cost—another basic requirement of "the tyranny of numbers."

So the story of the transistor to date has really been the struggle to realize these potentialities of high performance with high reliability and low cost. Let us take a quick look at the major milestones and problems that have been passed leading to today's transistor technology.

To measure the present and future stature of transistor electronics, we will glance at a few of the many system functions which transistors are now performing

or are committed to do. These examples cannot be inclusive—there are too many. Many of them will be drawn from the Bell System since the facts are more readily available to the authors. But some military, consumer, and industrial applications must be mentioned.

History

The era from the announcement of the invention of the point-contact transistor in mid 1948 until mid 1951 can be likened to early childhood. In this era, key items in advancing the device technology and research understanding came through the development of pulling and zone-melting methods for growing single crystals of germanium of unprecedented purity and uniformity.

Work by the inventors of the transistor and their colleagues during this era, resulted in improved theoretical understanding and physical realization of the junction transistor with greatly improved performance. Fabrication of junction diodes, triodes, and phototransistors was accomplished by both growing and alloying techniques. The micropower, high gain, and low noise predicted for such devices was proved.

From mid 1951 to mid 1952, a period reminiscent of adolescence was experienced. Gains in performance encouraged further applications. However, as for an adolescent youngster, so too for the transistor, from day to day we were impressed with new performance possibilities—yet, at the same time, new reliability difficulties were brought to light.

These reliability problems were shown to be surface dependent, research and development programs on surface physics and reliability were initiated throughout the industry. This heralded the "young manhood" period of 1952 to mid 1955.

During this period, larger responsibilities and definite commitments were undertaken. For example, point transistors went to work in an operator tone-dialing trial in October of 1952. The first over-the-counter sales of alloy and grown-junction-transistor hearing aids took place around the end of 1952. A few years later no manufacturer produced tube hearing aids.

In March, 1953, point-contact phototransistors and transistors were employed in the card translator portion of direct distance-dialing equipment. In a reliability study over 100 million transistor hours of operation were attained in this application with an indicated failure rate of less than 0.04 per cent per 1000 hours. TRADIC, the first transistorized military computer, was demonstrated in January, 1954, and later in the same year, the first all-transistor personal radio became available in time for the Christmas market.

Other equipment developed during this era included the first all-transistor industrial computer, and transistorized telephone sets for the hard of hearing and for use in noisy locations. Transistorized rural carrier and line concentrator tests were successfully undertaken.

On the military side, digital data transmission and processing systems, computers, and missile-control systems relied wholly or in part on transistors for their successful development and trial.

This sudden expansion in applications called for a wide variety of types of transistors. Techniques for germanium were pushed to their limits. Germanium crystals of almost perfect physical and chemical properties were developed to improve manufacturing yields.

Diodes, triodes, tetrodes, and power units of a wide variety were developed and put into manufacture. Silicon entered the field as an alloy-junction diode of greatly improved properties; silicon became an important bread and butter business.

Gains in understanding of the surface effects in reliability were put to work successfully in the form of improved surface treatment and capsulation techniques. Reliability surpassed that of all but the best telephone tubes. In fact, one hearing aid transistor manufacturer complained bitterly and frequently over the disappearance of his replacement business.

Where then did transistor electronics stand near the end of this era of early manhood?

The wide variety of applications called for a staggering range of characteristics. Point, grown, alloy, and surface-barrier techniques on germanium and silicon were close to their capability limits. Each structure and technique had serious limitations. Transistor designers and manufacturers were forced to develop a large number of compromise designs—about 200. The resulting manufacturer's nightmare inhibited large-scale economical manufacturing of any one prototype. This was reflected in still high costs, from $2.00 to $45.00 per transistor.

Meantime, research people were busy pushing ahead the frontiers of understanding of materials, structures, and techniques. Toward the end of this era they made a triple breakthrough of such magnitude that it started a new era—the "Era of Maturity." This work demonstrated the possibility of removing most of the limitations which forced design compromises to a multiplicity of types.

By proving the feasibility of purifying silicon to transistor requirements, by solving the problems of solid-state diffusion as a technique, and by devising the diffused-base structure, they demonstrated the possibility of making diffused silicon transistors with cutoff frequencies of 50–100 mc, 5 to 10 times faster than the older germanium transistors while at the same time retaining the silicon advantages of high power, temperature, and efficiency. Application of similar structures and techniques to germanium pushed the frequency frontier into the low-microwave region.

The magnitude of these research accomplishments dictated a re-evaluation of most transistor development-application programs starting in mid 1955. It was realized that further refinements of growing, alloying, and surface barrier techniques were possible but only at large development and manufacturing costs. However,

the dramatic demonstrations of potentiality of these new diffusion techniques, based as they were on more complete scientific understanding, clearly showed that maximum effort should be devoted to their development. The eventual economies of a single prototype diffusion technique and diffused structure, with heavy emphasis on silicon, were the goals.

Present Status

Where then has this diffusion breakthrough brought us today? Either in, or just entering, production are the following diffused devices:

1) Silicon-power rectifiers ranging from 0.5 ampere up to 100 amperes which will handle reverse breakdown voltages up to several hundreds of volts.
2) Silicon-voltage limiters with closely controlled breakdown voltages ranging from 4.2 volts up to 200 volts.
3) Diffused-base silicon transistors with frequency cutoffs in the 50–100-mc range. As switching transistors they will provide 10-mc switching rates with attendant power dissipation up to one-quarter watt.
4) Diffused-base germanium transistors with frequency cutoffs as high as 1000 mc. One code of this prototype designed as an oscillator has a minimum rating of 50 mw output power at 250 mc. Other codes designed for very high speed switching provide switching rates as high as 50–100 mc.

Today the circuit designer has at his command a broad range of structures and characteristics. These new devices in combination with improved versions of the older structures greatly extend the range of performance formerly possible. It is not possible here to compare all of their electrical characteristics, but it is helpful to show in Fig. 1 the frequency range as a function of power dissipation covered by presently available types. The frequencies plotted are the grounded-base-alpha cutoff frequencies for each prototype. For video or broad-band amplifiers, the useable range would be somewhat below the cutoff frequency depending upon the mode of application. However, for oscillators, useful power outputs can be achieved to well above the frequency cutoff.

The availability of such complete performance range from present semiconductor devices has profoundly affected the kinds and numbers of system applications.

In transmission systems, pulse-code modulation carrier, personal radio paging, and VHF communications are depending upon diffused germanium devices. Although transistors have not yet found their way into commercial television receivers, laboratory developments indicate this to be highly probable with attendant savings in size, power, and maintenance costs.

In power systems, the new diffused silicon rectifiers, voltage regulators, and lightning protectors are essential

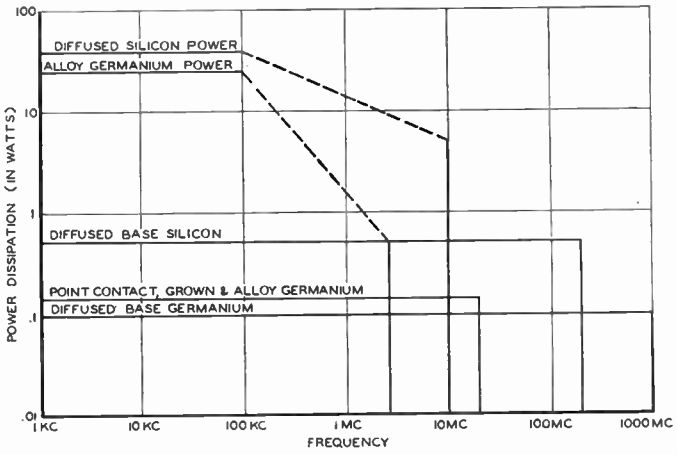


Fig. 1—The curves indicate the power-frequency spectrum covered by various prototypes of transistor structures.

components for success. By applying diffusion to cheap silicon, it has been possible to design a much smaller and better telephone click reducer than the copper-oxide unit now in use for many years but dependent upon an uncertain supply of unique copper.

In electronic switching, computers, and digital data processing, the new diffused diodes and triodes are greatly extending the speeds and reliability while at the same time greatly reducing the size and power required for such large and complex machines.

In the military area too this new line of devices is gaining acceptance in computers, missile systems, servo-systems, fuses, radar, communications, and power supplies. In fact, diffused germanium transistors are now circling the earth in the EXPLORER and VANGUARD satellites.

All in all today's variety of old and new transistors are finding their way into a staggering variety of tube and nontube replacement equipment. To round out the list with a few more, there should be mentioned binaural hearing aids, portable radios, phonographs and dictating machines, auto radios and fuel-injection systems, portable cameras, paging receivers and instruments, machine-tool controls, clocks and watches, toys, and even a guidance system for a chicken feeding cart.

As of this date, there are approximately seventy domestic and foreign manufacturers of transistors and related diodes. Production figures abroad are not available but it is known that essentially every major Western nation, as well as Japan, is very active. It is believed, however, that they are somewhat behind the United States in application and production. We have no authoritative technical information on Russia's status but *Pravda* assures us that Russia is well in the forefront—as usual?

New Designs

In the laboratory stage there are a number of new designs which will extend the range of electrical performance of the devices presently in manufacture. One, which is now entering production, is the four-region

p-n-p-n silicon switching diode. The electrical characteristics of this device are similar to those of a cold-cathode gas tube. However, the silicon device requires a great deal less power and can operate at speeds one thousand times faster than the gas tube. Thousands of these diodes will be used in the switching network of all-electronic telephone systems. Other major applications will doubtless be found in military and commercial digital computers.

It is interesting to note that this small device requires precise control of almost every bulk and surface property known to semiconductors. It is necessary to control accurately the density of impurities throughout the bulk material, the width of the various layers, and the density of the imperfections in the bulk material, which in turn controls the lifetime of minority carriers. It is necessary to control not only the density of these imperfections, but also the type of imperfections (the energy level within the forbidden gap). On the surface one must control and add impurities in such a manner that the density and type of surface states are within reasonably narrow limits. The surface must be carefully cleaned and oxidized so that the device will be electrically stable over long periods of time. In addition, the atmosphere around the device must be controlled so that there are no ions present to alter the electrical properties of the diode. Recent technological developments have made possible such precise control of each of these properties.

With the ever-advancing speeds of electronic computers go requirements for faster and faster computing diodes. By a controlled reduction of carrier lifetime, minority carrier storage and recovery times have been reduced to less than 2 mμsec. These designs, which are ready for pilot manufacture, will extend the speed of computer diodes by a factor of 10.

New diffused transistor structures are being made in the laboratory by reducing the thickness of the base layer and by reducing electrode spacing and cross-sectional area. In this way the frequency range of germanium and silicon transistors can be extended another factor of 10 over that of present designs. It is expected that diffused germanium transistors will soon be made which will oscillate at frequencies as high as 10,000 mc per second.

For some time it has been known that a variable reactance can serve as an amplifier, but it is only recently that this principle has been put to practical microwave use. The dependence of the capacitance of a *p-n* junction upon the voltage across it makes possible a rapidly variable capacitance, when an appropriate high-frequency voltage drive is used. Furthermore, theory predicts that this device should have extremely low noise. In actuality, an amplifier of exploratory design has a measured gain of 15 db at a frequency of approximately 6000 mc per second and a measured noise figure of 4.5 db. In this case the driver was a 12-kmc reflex klystron. It now seems possible that amplifiers can be designed for operation at frequencies as high as, or possibly higher than, those possible with advanced designed vacuum tube and traveling wave structures with properties competitive in many ways.

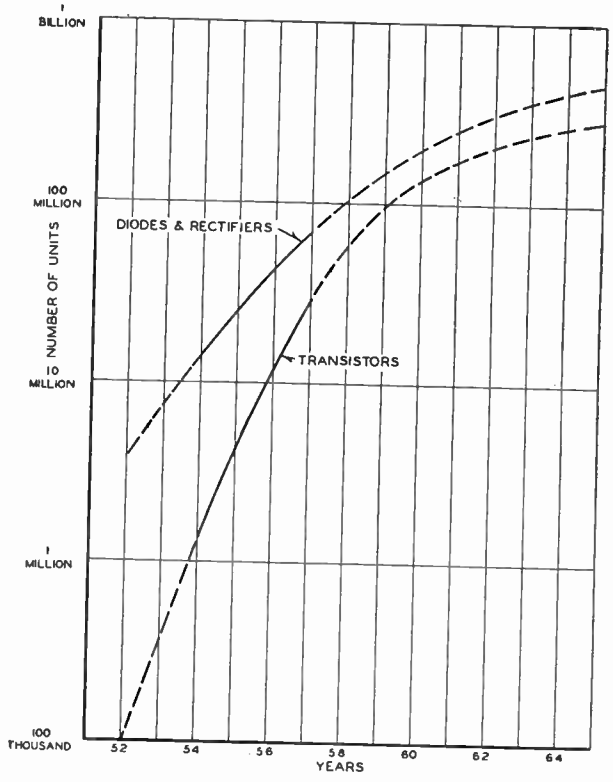


Fig. 2—The solid portions of the curves indicate the number of units sold in the United States by the year. The dotted portions of the curves represent estimates.

Growth of an Industry

The growth of the transistor from its birth in 1948 to its present maturity has been rapid indeed. With the critical need for electronic components in today's world and with the many advantages offered by transistors, it is not surprising that such a revolution has occurred. The management of nearly every major electronic laboratory has directed teams of their most competent research and development people toward the design of these semiconductor devices.

Through the years, meetings and symposia have been held for scientists to exchange information on the most recent technology developments. By concerted effort, many of the basic problems associated with semiconductor devices have been solved; today, transistors are being used in many industrial and military systems, particularly where low power, small size, and high reliability pay off. In the United States the sale of transistors alone has risen from a level of essentially nothing in 1952 to 29 million units in 1957. Forecasters predict that these sales will go to over 250 million units by 1965. If one adds to this the sale of semiconductor diodes, it is predicted that combined sales will reach 600 million by 1965. (See Fig. 2.) In 1957, the dollar volume of transistors and semiconductor diode sales

was 69 million dollars and 103 million dollars, respectively. By 1965, it is expected that the dollar volume of semiconductor sales will exceed that of the older electron tube. A further indication of the growth can be seen by the fact that the Joint Electron Tube Engineering Council had issued 600 transistor and 1300 diode industry codes by the end of 1957.

During its short ten years of life, the transistor, through its inherent low-power requirements, small size, and high reliability has permeated the entire electronics industry. It has already captured large sections of the market. With materials, structures, and techniques presently in the laboratory and currently in manufacture, the transistor will play an increasingly important part in modern electronic technology.

By basic scientific contributions and imaginative inventions, scientists and engineers have laid the foundation on which is being built a truly great technological industry. It may well be that the extension to man's mind, that transistor electronics makes possible, will yet have a greater impact on society than the nuclear extension of man's muscle.

The Status of Transistor Research in Compound Semiconductors*

DIETRICH A. JENNY†, MEMBER, IRE

Summary—**New semiconductors capable of competing with germanium and silicon in transistor applications must be looked for among the compound semiconductors, and more specifically among the III-V and IV-IV compounds. Gallium arsenide and indium phosphide are the most promising all-round materials for high-frequency as well as high-temperature performance. Indium antimonide and indium arsenide may be of interest for extremely high-frequency transistors operating at low temperatures The aluminum compounds, gallium phosphide and silicon carbide, are potentially useful for very high operating temperatures at the cost of high-frequency performance. Some of the unusual properties of the compound semiconductors have led to novel methods of junction preparation and new junction structures, such as the surface-diffusion and the wide-gap junction. Bipolar and unipolar surface-diffusion transistors have been demonstrated in indium phosphide, and the wide-gap emitter principle for high injection efficiency has been experimentally verified in gallium arsenide transistors. Electron lifetimes in these two compound semiconductors are estimated from the transistor results.**

Introduction

THE ADVENT of the germanium transistor in 1948, and the silicon transistor shortly thereafter, raised the inevitable question whether there are other semiconductors capable of exhibiting transistor action. Research in this direction was primarily stimulated by the hope of finding a semiconductor with superior properties for transistor applications. A glance at the periodic table and the electrical properties of the elements shows immediately that such a material would scarcely be found among the elemental semiconductors, except possibly diamond which has some obvious disadvantages. Therefore, the search for a competitor to germanium and silicon in the transistor field was concentrated on the *compound semiconductors*.

Compound semiconductors, in contrast to the elemental semiconductors, are true chemical compounds of two or more elements with characteristic stoichiometric compositions. Representatives of this vast class of semiconductors are found throughout the entire range of chemical compounds from the simple binaries to the most complex organic structures. The elemental semiconductors, such as α-tin, tellurium, selenium, germanium, and silicon are, in effect, only special cases of the compounds. Although by far the major systematic research efforts, both theoretical and experimental, have heretofore been concentrated on germanium and silicon, the compounds have played an important role in semiconductor research from the beginning. In fact, the earliest evidence for a conduction mechanism different from that in metals was Faraday's observation of a negative temperature coefficient of the resistivity in the compound, silver sulfide, in 1833. Rectification at a contact between dissimilar materials was discovered by Braun with pyrites and galena, and almost simultaneously by Schuster with "tarnished" copper, or copper oxide, in 1874. Silicon carbide and lead sulfide attained some importance as detectors in the early radio days, but they were soon displaced by the vacuum tube. Copper oxide has been one of the most important solid rectifier materials to this day, particularly in power applications. The practical importance of these compounds stimulated some early fundamental research, but, due to the lack of a satisfactory model of semiconduction, little significant information was gained. In 1931, after quantum mechanics had come into its own, Wilson laid

* Original manuscript received by the IRE, April 10, 1958. The work reported in this paper was supported in part by the U. S. Air Force, Wright Air Dev. Center, Dayton. Ohio.

† RCA Labs., Princeton, N. J.

down the groundwork for modern semiconductor theory in the form of the energy band model.[1] Sporadic experimental research was subsequently carried out with many compound semiconductors in an effort to study them in terms of the Wilson theory, but technological and reproducibility problems precluded satisfactory conclusions. The renewed interest in point contact diodes as radar and general high-frequency detectors during World War II marked the beginning of intensive theoretical and experimental research on germanium and silicon. Tremendous progress in the understanding and application of semiconductors was made during this period as evidenced by the discovery of the transistor by Bardeen and Brattain in 1948[2] and the *p-n* junction transistor by Shockley in 1949.[3]

Stimulated by these events, compound semiconductor research received renewed attention which resulted in the demonstration of a point contact transistor with lead sulfide by Gebbie, Banbury, and Hogarth.[4] However, it was Welker's contribution in 1952, pointing out the semiconducting properties of aluminum antimonide and indium antimonide, that sparked extensive and systematic research in the compound semiconductor field.[5] It became clear that the study of compound semiconductors could contribute much to the fundamental understanding of semiconductors in general, and that new applications as well as improvements of the devices already known would ultimately result. Systematic compound semiconductor research is still in its infancy and much remains to be done until the understanding of the compounds has reached a stage comparable to that of germanium and silicon. However, a host of fundamental information is already available in certain compound semiconductor classes, such as the III-V compounds, the II-VI compounds, and the only known IV-IV compound, silicon carbide. Work is progressing rapidly towards the realization of certain practical devices with some of these new semiconductors. Galvanomagnetic devices of indium antimonide and indium arsenide, utilizing the Hall and magnetoresistance effect, have recently appeared on the market. Diodes and rectifiers made from gallium arsenide, indium phosphide, aluminum antimonide, gallium phosphide, and silicon carbide are under development. Compound semiconductor transistors are still in the research stage, but their feasibility has been experimentally established in indium phosphide and gallium arsenide.

The following is a brief summary of the evaluation of compound semiconductors for transistor applications based on information from the literature and original work by the author.

Semiconductor Properties and Transistor Performance

In the search for new transistor materials certain fundamental semiconductor properties must be evaluated in comparison with those of germanium and silicon. The forbidden energy band gap (band gap), the impurity activation energies of donors and acceptors, the charge carrier mobilities of electrons and holes, and the dielectric constant are of primary importance. Other physical and chemical properties have to be taken into account due to their effect on the material and device technology, and device stability. The following shall be restricted to the discussion of the electronic properties directly affecting the performance of transistors, whereas other physical and chemical properties of secondary importance will be largely neglected.

The first-order relations between transistor performance and the important semiconductor properties are summarized in Table I, where the unipolar as well as the bipolar transistor types are considered. The unipolar relations are equally applicable to high-frequency diodes and general rectifiers with minor modifications.[6] The upper frequency limit of unipolar transistors is determined by the RC time constant of the junction capacitance and the series resistance, as shown by Dacey and Ross.[7] The upper frequency limit relation for bipolar transistors is that derived by Giacoletto.[8]

A high band gap is conducive to high-temperature operation, whereas low impurity activation energies are of importance for low-temperature operation. High mobilities and a low dielectric constant are desirable for high-frequency performance. However, it must be borne in mind that these properties are to some extent interrelated, so that certain compromises are unavoidable. For instance, the dielectric constant affects not only the high-frequency performance through its effect on device capacitances, but enters also into the impurity activation energy (hydrogenic model)[9] and the impurity scattering mobility (Conwell-Weisskopf and Brooks-Herring relations).[10] Without going further into these details, it

[1] A. H. Wilson, "The theory of electronic semiconductors," *Proc. Roy. Soc.*, vol. 133, pp. 458–491; October, 1931.

[2] J. Bardeen and W. H. Brattain, "The transistor, a semiconductor triode," *Phys. Rev.*, vol. 76, p. 459; July, 1958.

[3] W. Shockley, "The theory of *p-n* junctions in semiconductors and the *p-n* junction transistor," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

[4] H. A. Gebbie, P. C. Banbury, and C. A. Hogarth, "Crystal diode and triode action in lead sulphide," *Proc. Phys. Soc.*, vol. 63B, p. 371; February, 1950.

P. C. Banbury and H. K. Henisch, "On the frequency response of lead sulphide transistors," *Proc. Phys. Soc.*, vol. 63B, pp. 540–541; April, 1950.

C. A. Hogarth, "Crystal diode and triode action in lead selenide," *Proc. Phys. Soc.*, vol. 64B, pp. 822–823; June, 1951.

H. K. Henisch, "Transistor experiments on binary lead compounds," *Phys. Rev.*, vol. 91, p. 213; July, 1953.

[5] H. Welker, "Ueber neue halbleitende Verbindungen," *Z. Naturf.*, vol. 7a, pp. 744–749; November, 1952.

[6] D. A. Jenny, "A gallium arsenide microwave diode," Proc. IRE, vol. 46, pp. 717–722; April, 1958.

[7] G. C. Dacey and I. M. Ross, "Unipolar field-effect transistor," Proc. IRE, vol. 41, pp. 970–979; August, 1953.
——, and ——, "The field-effect transistor," *Bell Sys. Tech. J.*, vol. 34, pp. 1149–1189; November, 1955.

[8] L. J. Giacoletto, "Comparative high-frequency operation of junction transistors made of different semiconductor materials," *RCA Rev.*, vol. 16, pp. 34–42; March, 1955.

[9] H. A. Bethe, "Theory of the Boundary Layer of Crystal Rectifiers," M.I.T. Rad. Lab., Cambridge, Mass., Rep. No. 43-12; November, 1942.

[10] E. Conwell and V. F. Weisskopf, "Theory of impurity scattering in semiconductors," *Phys. Rev.*, vol. 77, pp. 388–390; February, 1950.
H. Brooks, "Scattering by ionized impurities in semiconductors" (abstract), *Phys. Rev.*, vol. 83, p. 879; August, 1951.

TABLE I

SEMICONDUCTOR PROPERTIES AND TRANSISTOR PERFORMANCE

Transistor Performance		Pertinent Semiconductor Properties	First-Order Relation
Operating temperature range	Upper temperature limit T_u	Band gap E_g	$T_u \propto E_g$
	Lower temperature limit T_l	Impurity activation energy E_i	$T_l \propto E_i$
Upper frequency limit F	Unipolar transistors	The higher of the two mobilities μ_{high}; dielectric constant κ	$F \propto \frac{\mu_{high}}{\kappa^{1/2}}$
	Bipolar transistors	Electron and hole mobility, μ_n and μ_p; dielectric constant κ	$F \propto \sqrt{\frac{\mu_n \times \mu_p}{\kappa^{1/2}}}$

TABLE II

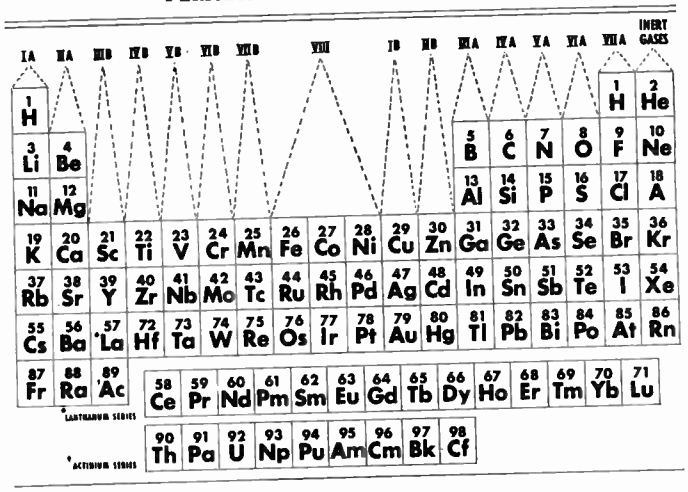
EXAMPLES OF COMPOUND SEMICONDUCTORS

Compound Class	Compound	Band Gap (ev)	Electron Mobility (cm^2/volt per second)
Elemental	α-Sn	0.08	~3000
	Ge	0.7	3900
	C (diamond)	6.7	1800
Binary	InSb	0.18	65,000
	GaAs	1.35	>5000
	SiC	2.8	>100
Ternary*	$AgTlTe_2$	0.1	—
	$CnInSe_2$	0.9	~1000
	$CnAlS_2$	2.5	—
Organic*	Cynanthron	0.2	10^{-8}
	Indanthracene	0.66	10^{-12}
	Anthracene	1.64	10^{-2}

* U. Winkler, "Die elektrischen Eigenschaften der intermetallischen Verbindungen Mg_2Si, Mg_2Ge, Mg_2Sn and Mg_2Pb," *Helv. Phys. Acta*, vol. 28, pp. 633–666; December, 1955.

TABLE III

PERIODIC TABLE OF THE ELEMENTS



is now possible to evaluate new semiconductors in comparison with germanium and silicon on the basis of the first-order relations between transistor performance and semiconductor properties of Table I.

THE NEW COMPOUND SEMICONDUCTORS

Semiconductors are not confined to the elements, but a vast number of representatives are found among the chemical compounds. A few representative examples of compound semiconductors with increasing chemical complexity and some of their properties are listed in Table II.

Each class of compounds contains a large number of semiconductors with band-gap values from a small fraction of an electron volt to several electron volts. On the other hand, the mobilities seem to fall within characteristic ranges for each compound class, whereby the binary compounds exhibit the highest absolute mobility values as well as the highest mobilities for a given band gap. The favorable band-gap and mobility combinations together with the relatively simple chemistry of the binary compounds have led to concentration of the research effort on this compound class.

The evaluation of binary compound semiconductors for transistors and related applications has indicated that most of the numerous binary compound groups must be ruled out for various reasons. Table IV shows the results of this evaluation which was compiled from the literature and from an experimental investigation of over 200 different compounds carried out at RCA Laboratories. This table is arranged according to the columns of the periodic table of the elements of Table III. The reasons for discarding most of the binary compound groups are indicated in the respective boxes of Table IV, where the symbols and abbreviations are: low melting point (low MP); low band gap (low E_g); low mobilities (low μ), chemical, thermal, or mechanical instability (unstable); and technological difficulties in preparation, purification, and crystal growth (technology).

The compound groups in the two boxes framed with heavy lines contain at present the most promising representatives, and are the well known column IV elemental semiconductors including germanium, silicon and the compound silicon carbide, and the III-V compound semiconductors. The boxes framed in heavy dashed lines contain compound groups which could not be investigated satisfactorily because of technological difficulties in preparing materials suitable for measurements. The

TABLE IV
BINARY COMPOUND SEMICONDUCTOR EVALUATION

COLUMNS OF THE PERIODIC TABLE	IIIA B Al Ga In Tl	IVA C Si Ge Sn Pb	VA N P As Sb Bi	VIA O S Se Te Po	VIIA F Cl Br I (At)
IA Li Na K Rb Cs Fr	low MP unstable metallic	low MP unstable	low MP unstable	low μ, MP unstable	low μ
IIA Be Mg Ca Sr Ba Ra	low E_g metallic	technology low μ, E_g	technology low μ, E_g	low μ	low μ
IIIB Sc Y La RARE EARTHS	low E_g metallic	technology	technology	low μ	low μ
TRANS. ELEMENTS ACTINIUM SERIES	low E_g metallic	technology low E_g metallic	technology low E_g metallic	low μ	low μ
IB Cu Ag Au	low E_g metallic	low E_g metallic	low μ	low μ, E_g	low μ
IIB Zn Cd Hg	low E_g metallic	no compounds	low μ	low μ, E_g	low μ
IIIA B Al Ga In Tl	low E_g metallic	technology metallic	III-V COMPOUNDS	low μ, E_g	low μ
IVA C Si Ge Sn Pb		C Si Ge α-Sn SiC	technology unstable	low μ, E_g	low MP
VA N P As Sb Bi			low MP	low μ, E_g	low MP
VIA O S Se Te Po				low μ, E_g	low MP
VIIA F Cl Br I (At)					low MP

pertinent properties of the members of the two interesting compound groups are shown in Table V opposite, as far as they are known at present.

This table contains measured values of the various properties with indications for the future trends expected from theory, and experience with germanium and silicon. The mobility values for the compounds in Table V are Hall mobilities, whereas the mobilities for germanium and silicon are the drift mobilities.

Gallium arsenide and indium phosphide clearly stand out as the most promising representatives in terms of the first-order relations between transistor performance and semiconductor properties of Table I. For very high-temperature devices the aluminum compounds, gallium phosphide and silicon carbide, are of interest due to their high band gaps, if a sacrifice in high-frequency performance from low mobilities can be tolerated. Very high frequencies should be attainable with indium arsenide and indium antimonide because of their high electron mobilities, but the low band gaps require operation below room temperature.

Besides the interesting electronic properties of some of these single compounds, it is well worth mentioning that many representatives form solid solutions throughout the entire mixture range with a monotonic band gap transition between the two band gaps of the components. Table VI shows three examples of such mixtures with their band gap coverage which extends from 0.33 ev to 2.25 ev without interruption.

The mixture of gallium arsenide and gallium phosphide shown in Fig. 1 is of particular interest and will be discussed later.[11]

The III-V compounds crystallize in the zincblende structure which is geometrically identical to the diamond lattice of germanium and silicon. The atoms of each component element are contained in their own face-centered cubic sublattices. This similarity between the diamond and the zincblende structure is of significance in comparing the properties of the III-V compounds with elements of column IV (germanium, silicon, diamond) as Herman has indicated.[12]

THEORETICAL TRANSISTOR PERFORMANCE IN COMPOUND SEMICONDUCTORS

Experience with germanium and silicon has shown that single crystal material with extremely high purity and crystal perfection is imperative for successful transistor experiments. Therefore, the evaluation of a new semiconductor for transistor applications must be made at an early stage of material research, to avoid an unnecessary waste of effort on the technology of purification and crystal growth with possibly unsuitable materials. Transistor research has, fortunately, progressed to the point where such an evaluation can be made on the basis of the few fundamental properties pointed out earlier. The results of an evaluation of several promising compound semiconductors in comparison with germanium and silicon is briefly discussed below for the unipolar and the bipolar transistor types. The two device properties of major interest are high-frequency performance and operating temperature range with particular emphasis on the upper operating temperature limit.

The temperature limitations of the various semiconductors is best represented in comparison with germanium. Assuming that devices made from germanium are operable up to a temperature of 100°C, as established by practical experience, the upper operating temperature limit, T_u, of devices made from other semiconductors can be estimated from the expression for the intrinsic carrier concentration. Here the actual concentration is set equal to that of germanium at 100°C:

$$T_u = \frac{E_g}{k \log (N/n_i^{Ge})} = 533E_g - 273; (°C) \qquad (1)$$

where E_g is the band gap of the new semiconductor, k is the Boltzmann constant, N is the averaged density of states in the conduction and valence bands, and n_i^{Ge} is the intrinsic hole and electron density of germanium at

[11] O. G. Folberth, "Mischkristallbildung bei $A^{III}B^{V}$-Verbindungen," *Z. Naturf.*, vol. 10a, pp. 502–503; June, 1955.
[12] F. Herman, "Speculations on the energy band structure of zincblende-type crystals," *J. Electronics*, vol. 1, pp. 103–114; September, 1955.

TABLE V

III-V Compound and Column IV Semiconductor Properties*

Semiconductor	Band Gap (ev)	Electron Mobility (cm^2/volt sec)	Hole Mobility (cm^2/volt sec)	Dielectric Constant	Melting Point (°C)
α-Sn	0.08	~3000	—	—	18 (transition)
Ge	0.7	3900	1900	16	958
Si	1.1	1500	500	11.8	1414
SiC	2.8	>100	>20	7	2700
C (diamond)	6–7	1800	1200	5.7	>3500
InSb	0.18	65,000	~1000	15.9	523
InAs	0.33	20,000	~200	11.7	936
GaSb	0.68	~4000	~700	14	702
InP	1.25	>4000	>100	10.8	1060
GaAs	1.35	>5000	>400	11.1	1280
AlSb	1.52	≥400	≥400	10.1	1080
GaP	2.25	>100	>20	8.4	>1300
AlAs	~2.2	—	—	—	>1600
AlP	~2.5	—	—	—	—

* This table is compiled from a large number of literature references, and results obtained at RCA Laboratories. Reference is made to a few summary papers which contain comprehensive data and bibliographies in the field of III-V compound semiconductors.

H. Welker and H. Weiss, "Group III—group V compounds," in "Solid State Physics," F. Seitz and D. Turnbull, ed., Academic Press, Inc., New York, N. Y., vol. 3, pp. 1–78; 1956.

L. Pincherle and J. M. Radcliffe, "Semiconducting intermetallic compounds," in "Advances in Physics" (suppl. 5 to *Phil. Mag.*), vol. 5, pp. 271–322; July, 1956.

F. A. Cunnel and E. W. Saker, "Properties of III-V compound semiconductors," in "Progress in Solids," A. F. Gibson, ed., John Wiley and Sons, Inc., New York, N. Y., vol. 2, pp. 37–65; 1957.

E. Burstein and P. H. Egli, "The physics of semiconductor materials," in "Advances n Electronics and Electron Physics," L. Marton, ed., Academic Press, Inc., New York, N. Y., vol. 7, pp. 1–84; 1955.

TABLE VI

Band Gap Coverage of III-V Compound Mixtures

Mixture	Band Gap Coverage (ev)
InAs-InP[11]	0.33–1.25
GaAs-Ge*	1.25–1.35
GaAs-GaP[11]	1.35–2.25

* D. A. Jenny and R. Braunstein, "Some properties of Ge-GaAs mixtures," *J. Appl. Phys.*, vol. 29, pp. 596–597; March, 1958.

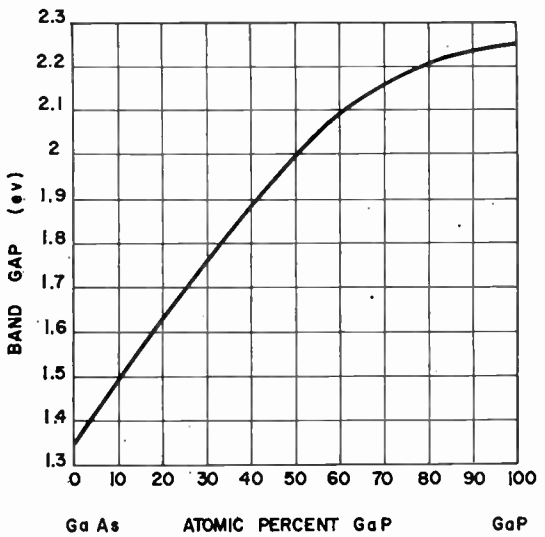


Fig. 1—Band-gap distribution in gallium arsenide-gallium phosphide mixtures (after Folberth[11]).

100°C. The numerical approximation neglects the effect of minor temperature and effective mass terms in the density of states. Accurate calculations show that, for instance, in silicon T_u is less than that obtained from (1) because of the relatively high effective masses of the charge carriers. The lower operating temperature limit could, in principle, be calculated from a similar relation by substituting the impurity activation energies for the band-gap energies. Unfortunately, very little information about the impurity activation energies in the compounds is available, although preliminary measurements of the Hall effect at low temperatures indicate that the impurity activation energies of indium phosphide and gallium arsenide are in the order of those in germanium. This is in agreement with the theoretical expectation derived from the low effective masses of the charge carriers in these two compounds (hydrogenic model).[9]

The high-frequency limit can be calculated from the first order relations of Table I and the material properties of Table V with the assumption that germanium transistors (unipolar and bipolar) will ultimately give useful performance at 1000 mc, which is a reasonable limit considering technological factors. This would require critical device dimensions, such as the base width in a bipolar transistor, of about 1 micron.

The mobilities for germanium and silicon, used in this calculation, are the intrinsic or lattice scattering values for the unipolar case, and the impurity scattering values at an ionized impurity concentration of 5×10^{16} cm^{-3} for the bipolar case. Design calculations by Dacey and Ross[7] have shown that very low impurity (majority carrier) concentrations of about 10^{14} cm^{-3} give optimum performance in unipolar germanium transistors, so that the use of intrinsic mobility values is justified in this case. The majority carrier concentration in the diffusion region of modern high-frequency transistors is in the order of 5×10^{16} cm^{-3}, so that this value is used for the ionized impurity concentrations in the bipolar case. The

mobilities for the compounds are measured Hall values, as pointed out previously, whereas the performance of devices based on conduction, diffusion, and drift mechanisms is determined by the drift mobilities. Experience has shown that the drift mobilities in germanium and silicon differ by less than a factor 2 from the Hall mobilities, so that the use of the latter is permissible for this preliminary evaluation. The fact that the ionized impurity concentration in the compounds evaluated here is still higher than 5×10^{16} cm^{-3} gives a pessimistic picture in comparison with germanium and silicon and an improvement can be expected when purer and more perfect single crystals of the compounds become available.

To obtain a realistic temperature dependence of the upper frequency limit of transistors, it is necessary to consider the effect of temperature on the mobilities. Since relatively little is known about the actual temperature dependence of the mobilities in compound semiconductors, preliminary estimates obtained for the electron mobilities in indium phosphide are used for all compounds considered here.[13] The approximate temperature function is $\mu_{n,p} \propto T^{-2}$. For germanium and silicon the experimentally determined temperature dependences are used, such as for germanium[14] $\mu_n \propto T^{-1.6}$ and $\mu_p \propto T^{-2.3}$, and for silicon[15] $\mu_n \propto T^{-2.5}$ and $\mu_p \propto T^{-2.7}$. These temperature dependences are those of the lattice scattering or intrinsic mobilities which are applicable to the unipolar transistor case. In the case of the bipolar transistor the negative temperature exponents have somewhat smaller absolute values at the lower temperatures due to the effect of the onsetting impurity scattering. However, the above approximations do not change the ultimate outcome of this evaluation, since they are on the pessimistic side for the compounds.

It is now possible to plot the numerical first-order relations between transistor performance and semiconductor properties for several compounds of potential interest for transistors. Figs. 2 and 3 show such curves in the form of log-log plots of the relative upper frequency limit in megacycles vs operating temperature in °C for unipolar and bipolar transistors, respectively.

These figures are self-explanatory and need little interpretation. In conclusion, it can be stated that gallium arsenide seems to be the only semiconductor, among those evaluated to date, which promises to be competitive with both germanium and silicon for transistor applications. In other words, the pertinent fundamental properties of gallium arsenide, as determined experimentally, indicate that transistors made from this material may ultimately not only exceed the temperature limits imposed by silicon, but also operate at higher frequencies than germanium transistors. Indium phosphide is a close runner-up with only a slight disadvantage in high-frequency and high-temperature potentialities. Aluminum antimonide, gallium phosphide, and silicon carbide can operate at even higher temperatures at the cost of high-frequency performance.

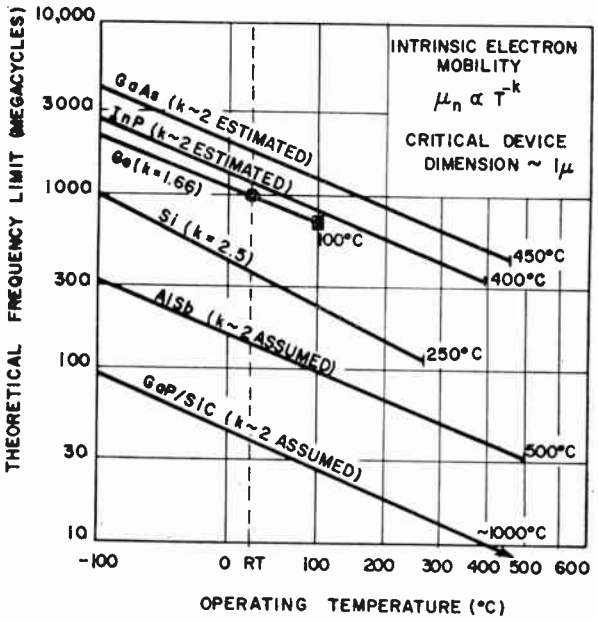


Fig. 2—Theoretical frequency limit vs operating temperature of unipolar compound semiconductor transistors.

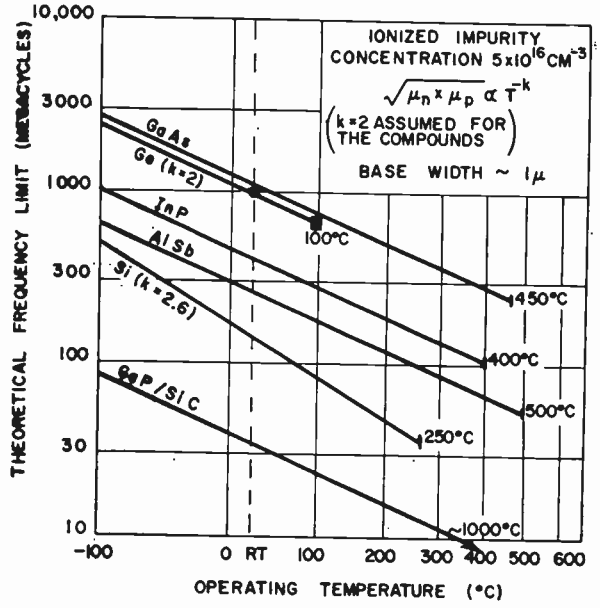


Fig. 3—Theoretical frequency limit vs operating temperature of bipolar compound semiconductor transistors.

Junctions in Compound Semiconductors

Most of the compound semiconductors lend themselves to basically the same junction preparation technology and yield the same general junction characteristics as germanium and silicon. Alloy, diffusion, and grown junctions, as well as metal-semiconductor barriers, have been investigated. The junction characteristics are qualitatively similar to those in silicon with

[13] M. Glicksman, RCA Labs., personal communication.
[14] F. J. Morin, "Lattice-scattering mobility in germanium," *Phys. Rev.*, vol. 93, pp. 62–63; January, 1954.
[15] G. W. Ludwig and R. L. Waters, "Drift and conductivity mobility in silicon," *Phys. Rev.*, vol. 101, pp. 1699–1701; March, 1956.

pronounced evidence for carrier generation and recombination in the space-charge region due to a trap level near the middle of the forbidden band.

However, some of the compounds possess certain properties which make possible new preparation techniques and new junction structures. For instance, in the process of growing single crystals of a compound from the melt, deviations from the melt-stoichiometry have a pronounced effect on the segregation coefficient of impurities, without affecting the stoichiometry of the solid.[16] This effect can be used to advantage for producing nonhomogeneous impurity distributions, and, therefore, for growing *p-n* junctions by changing the stoichiometry conditions during crystal growth. This is easily accomplished through variations in the pressure of the volatile component, such as the phosphorus pressure in the preparation of indium phosphide. Another novel method of preparing junctions in compound semiconductors makes use of the decomposition of the compound at temperatures appreciably below the melting point.[17] For instance, heating the surface of a doubly-doped (donors and acceptors) piece of indium phosphide beyond a certain temperature leads to evaporation of phosphorus leaving behind an indium rich melt. This surface melt "leaches" out certain impurities, so that a *p-n* junction results under favorable circumstances. An ohmic contact is automatically provided in the form of the indium metal at the surface. This method does not require the addition of an impurity during the process, nor does it necessitate heating to the melting point of the semiconductor. Among the many other junction preparation methods unique to the compound semiconductors, the two which were instrumental in demonstrating the feasibility of transistors in indium phosphide and gallium arsenide shall be described in some detail below.

The Surface-Diffusion Junction

Zinc, which is an acceptor impurity in III-V compounds, was found to adhere tenaciously to the surface of indium phosphide, so that surface diffusion can be effected within a certain temperature range without excessive loss by evaporation. Estimates of the surface-diffusion coefficient of zinc on indium phosphide show that it is several orders of magnitude higher than the volume-diffusion coefficient as expected. This observation opens an interesting possibility of producing large area *p-n* junctions near the semiconductor surface. The process is illustrated in three stages in Fig. 4. An impurity pellet containing zinc is placed on the surface of an *n*-type indium phosphide wafer [Fig. 4 (a)] and subsequently heated in a protective atmosphere, so that the pellet melts and forms a penetrating alloy melt similar to that in the preparation of an alloy junction. During this process zinc diffuses into the indium phosphide

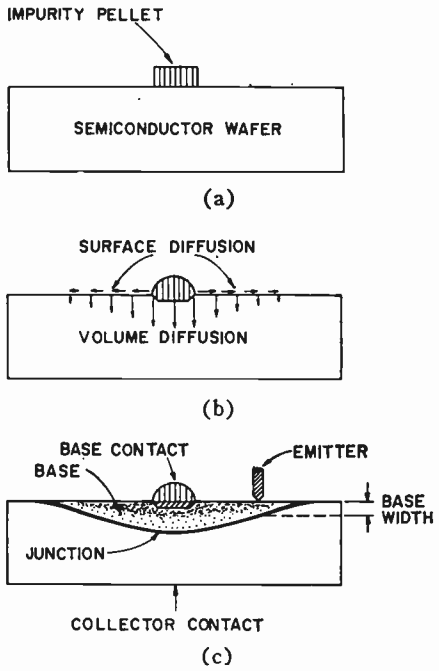


Fig. 4—Structure and processing steps of surface-diffusion junction and transistor.

underneath the alloy melt [Fig. 4 (b)]. At the same time zinc diffuses along the surface at a much higher rate and covers the surrounding surface area with a radially decreasing surface concentration. Volume diffusion towards the inside of the indium phosphide takes place simultaneously with a penetration depth dependent on the surface concentration. In this manner, after cooling to room temperature, a saucer-shaped junction is formed with an ohmic contact to the *p*-type side in the form of the solidified alloy dot [Fig. 4 (c)]. A very shallow large area junction is thus obtained, which lends itself ideally for photodiodes, photovoltaic junctions,[18] and particularly diffusion transistors. The method does not necessitate an impurity in the vapor phase, nor a sealed system, and the processing procedure is essentially equivalent to that for the preparation of alloy junctions. The automatically obtained ohmic contact to the diffused region is an added feature.

The Wide-Gap Junction

The effect of a change in the band gap of a semiconductor in the transition region of a *p-n* junction has been originally postulated by Shockley[19] and was later treated in detail by Kromer.[20] The most immediate application of such a junction structure is in the emitter of a transistor, where a higher band gap on the emitter side, as compared with the base side, can increase the injection efficiency tremendously over that of a homogeneous band-gap junction.

[16] K. Weiser, RCA Labs., personal communication.
[17] K. Weiser, "Decomposition method for producing *p-n* junctions," *J. Appl. Phys.*, vol. 29, pp. 229–230; February, 1958.
[18] D. A. Jenny, J. J. Loferski, and P. Rappaport, "Photovoltaic effect in Ga As *p-n* junctions and solar energy conversion," *Phys. Rev.*, vol. 101, pp. 1208–1209; February, 1956.
[19] W. Shockley, "Circuit Element Utilizing Semiconductive Material," U. S. Patent No. 2,569,347; September 25, 1951.
[20] H. Kromer, "Theory of a wide-gap emitter for transistors," Proc. IRE, vol. 45, pp. 1535–1537; November, 1957.

The possibility of producing solid mixtures or solid solutions of two semiconductors with different band gaps to attain intermediate band-gap values allows the realization of wide-gap junctions. The germanium-silicon system is one example, where the components are fully miscible. However, the gallium phosphide-gallium arsenide system, mentioned earlier (Fig. 1), has several technological advantages for the preparation of wide-gap junctions. It is possible to immerse gallium arsenide into an atmosphere of phosphorus and to take advantage of solid diffusion, so that phosphorus atoms are substituted for the arsenic atoms in the gallium arsenide lattice, thus converting the starting material into either gallium phosphide or any intermediate mixture of gallium arsenide and gallium phosphide. Arsenic is, of course, liberated in this process. Although the diffusion process does not produce an abrupt band-gap transition, it introduces a nonhomogeneous tapering band gap from the extreme of gallium phosphide (2.25 ev) at the surface to that of gallium arsenide (1.35 ev) in the interior. In addition to the band-gap transition a p-n junction must be produced in the region of the tapering band gap. This can be accomplished by diffusion doping with an appropriate impurity, such as sulfur in the case of p-type gallium arsenide, as was actually done in the example described here. Fig. 5(a) shows the energy-band representation of an ideal wide-gap junction with abrupt band-gap and impurity transitions, and Fig. 5(b) shows that of the described junction with graded band-gap and impurity transitions produced by the diffusion process.

The approximate band-gap value in the transition region of a graded wide-gap junction can be determined in an indirect way. The rectification mechanism of p-n junctions in all known low-lifetime semiconductors, except germanium, is invariably that of generation and recombination in the space charge region as described by Sah, Noyce, and Shockley,[21] instead of the diffusion mechanism originally introduced by Shockley for the ideal case with relatively long lifetimes. The reason for this is the presence of a trap or recombination-center level near the middle of the forbidden band. It is, therefore, reasonable to assume that the former mechanism is also active in the graded wide-gap junction under discussion. The rectification activation energy of both the reverse and the forward current corresponds in this case roughly to half the band-gap value in the junction transition region. Measurements of ordinary diffusion junctions in gallium arsenide (homogeneous gap junctions) and of the above described graded wide-gap junctions are shown in Table VII. These results indicate that the band-gap value in the transition region of the graded wide-gap junction is about 1.8 ev as compared to 1.35 ev for gallium arsenide alone. A cross check on this value from diffusion calculations and the band-gap distribution between gallium arsenide and gallium phosphide, shown in Fig. 1, have verified this result.

[21] C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in p-n junctions and p-n junction characteristics," PROC. IRE, vol. 45, pp. 1228–1243; September, 1957.

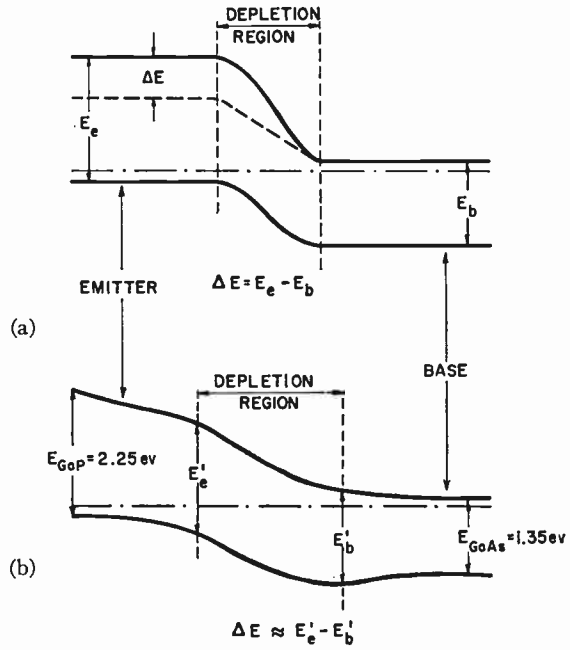


Fig. 5—Energy-band representation of (a) abrupt wide-gap junction and (b) graded wide-gap junction.

TABLE VII

RESULTS OF HOMOGENEOUS-GAP AND WIDE-GAP JUNCTIONS IN GALLIUM ARSENIDE

	Homogeneous-Gap Junction	Wide-Gap Junction
Rectification Activation Energy E_r (ev)	0.65	0.9
Estimated Band Gap in the Depletion Region $E_g = 2E_r$ (ev)	~1.3	~1.8
Actual Band Gap in the Depletion Region E_g (ev)	1.35	~1.8*

* Calculated from estimates of the diffusion coefficient of phosphorus in gallium arsenide and the band-gap distribution in gallium arsenide-gallium phosphide mixtures of Fig. 1.

Wide-gap junctions of the type described are of interest for diodes and rectifiers with even higher upper operating temperature limits than that for gallium arsenide, without necessitating the preparation of gallium arsenide-gallium phosphide mixtures as starting material in single crystal form. The use of this wide-gap junction as an emitter in gallium arsenide transistors for increased injection efficiency will be described below.

EARLY TRANSISTOR RESULTS IN COMPOUND SEMICONDUCTORS

As mentioned in the introduction, Gebbie, Banbury and Hogarth were the first to demonstrate point-contact transistor amplification in a compound semiconductor, lead sulfide.[4] However, lead sulfide has too low a band gap (0.35 ev) for useful transistors. Several reports of transistor action in new semiconductors which appeared in nontechnical publications are disregarded

here. Welker claimed to have observed transistor action in indium phosphide,[22] but Jenny showed that his criterion, although necessary, was not sufficient and the demonstration of power amplification is, at this time, the only satisfactory way of establishing the existence of the transistor mechanism.[23] In the following, two new transistor structures are described which were used in establishing the feasibility of transistors in indium phosphide and gallium arsenide.

The Surface-Diffusion Transistor in Indium Phosphide

Early transistor experiments in indium phosphide with point-contact structures, and a hybrid structure of a point-contact emitter and a large-area metal-semiconductor collector, have shown that transistor action with power gain is feasible in this material. However, the surface-diffusion transistor, described below, was the first transistor in a compound semiconductor to utilize a material property peculiar to indium phosphide and is, therefore, of particular interest.

The surface-diffusion junction described in the previous section was used as collector junction in this transistor. An emitter was applied to the p-type region near the periphery of the saucer-shaped junction, where the distance between emitter and collector is very small, as shown in Fig. 4. In this manner it was possible to provide an extremely small active base width, which was necessary because of the low electron lifetime and diffusion length. The emitter consisted of a small area metal-semiconductor contact made with indium, similar to the emitter in point-contact transistors.

The results of early measurements on surface-diffusion transistors in indium phosphide are summarized in Table VIII. The behavior of these transistors is qualitatively analogous to that of germanium and silicon transistors, so that a detailed presentation of their characteristics would not contribute to the information contained in the table. Surface stability to maintain low surface recombination velocities near the emitter is one of the major remaining problems. The emitter for practical transistors will ultimately consist of an alloy or diffusion junction rather than a metal-semiconductor contact.

The important contribution of the surface-diffusion transistor structure is that it made possible the reproducible demonstration of the feasibility of transistors in indium phosphide; furthermore it is, at present, the only effective tool for estimating electron diffusion lengths and lifetimes in this compound semiconductor.

It may be worth mentioning that unipolar surface-diffusion transistors have been demonstrated in indium phosphide, utilizing junction preparation techniques similar to those described. Preliminary measurements have indicated transconductance values of 0.1 ma per volt.

TABLE VIII
Early Results of Indium Phosphide Surface-Diffusion Transistors at 4000 Cycles

Power gain	36 db
Current gain β	13
Current gain α	~ 0.9
Base width w (measured)	$\sim 3\times10^{-4}$ cm
Electron diffusion length L_n	$\sim 3\times10^{-3}$ cm
Electron lifetime τ_n	$\sim 10^{-7}$ seconds

A Gallium Arsenide Transistor with a Wide-Gap Emitter

In the case of gallium arsenide it was necessary to utilize the wide-gap emitter principle to obtain sufficient injection efficiency for a demonstration of transistor amplification. The high carrier concentrations in the available gallium arsenide precluded the attainment of sufficient carrier injection through high carrier concentrations in the emitter, as is successfully accomplished in germanium and silicon. The graded wide-gap junction, described in the previous section (Fig. 5), was therefore used as emitter. A band-gap difference, ΔE, between emitter and base in a wide-gap emitter affects the injection efficiency, γ, of the homogeneous-gap case according to the relation:[15]

$$\gamma_{\Delta E} = \frac{1}{\left(\frac{1}{\gamma} - 1\right) e^{-\Delta E/kT} + 1} \tag{2}$$

where $\gamma_{\Delta E}$ is the wide-gap injection efficiency, T is the absolute temperature, and k is the Boltzmann constant. In the case of the graded wide-gap emitter the ΔE is taken as the band-gap difference between the two boundaries of the depletion layer in the emitter junction. This is actually only a pessimistic approximation, as the band-gap gradient in the emitter gives an additional improvement in injection efficiency. For instance, a ΔE of only 0.2 ev can yield an increase of the injection efficiency from 0.5 to 0.999 compared with the homogeneous band-gap case.

A transistor structure with the graded band-gap junction, described in the previous section, was used for the demonstration of transistor amplification in gallium arsenide. The fact that in this case the collector is also a graded wide-gap junction is purely incidental and of no further significance. The results of preliminary measurements on wide-gap and homogeneous-gap transistors in gallium arsenide are shown in Table IX. Difficulties in making low-resistance ohmic contacts to the gallium arsenide-gallium phosphide material of the emitter and collector precluded the attainment of the ultimate power amplification potential of this transistor type. Nevertheless, the results demonstrate not only the feasibility of transistors in gallium arsenide, but also the experimental verification of the theoretically predicted wide-gap emitter effect. The application of the wide-gap emitter to gallium arsenide transistor structures made possible the estimate of the electron diffusion length and lifetime in this compound semiconductor.

[22] H. Welker, "Ueber neue halbleitende Verbindungen II," *Z. Naturf.*, vol. 8a, pp. 248–251; April, 1953.

[23] D. A. Jenny, "Bemerkung zu einem von H. Welker gefundenen 'Transistor-Effekt' in Indium Phosphide," *Z. Naturf.*, vol. 10a, pp. 1032–1033; December, 1955.

Conclusions and Future Outlook

The foregoing status report on transistor research with new semiconductors indicates that certain compound semiconductors are potentially capable of extending transistor performance towards higher temperatures and possibly higher frequencies as compared to germanium and silicon. Transistor amplification has been demonstrated in indium phosphide and gallium arsenide, the two most promising all-round compound semiconductors. Gallium phosphide and silicon carbide may be useful at very high operating temperatures at the cost of high-frequency performance, whereas indium antimonide and indium arsenide could conceivably operate at very high frequencies but only below room temperature.

However, there are still numerous technological problems to be solved before the theoretical predictions for high-temperature and high-frequency performance can be realized in practical transistors. New preparation methods and new device structures have become possible with the compound semiconductors due to some of their unusual properties.

Perhaps the most intriguing aspect of the compound semiconductors is that gallium arsenide, and possibly indium phosphide, promise not only to combine, but even to exceed the favorable high-frequency and high-temperature transistor properties of germanium and silicon. Whether gallium arsenide or indium phosphide will ultimately replace germanium and/or silicon in transistors depends largely on technological and economic factors.

TABLE IX

Early Results of Homogeneous Gap and Wide Gap Emitters in Gallium Arsenide Transistors at Low Frequencies

	Homogeneous-Gap Emitter	Wide-Gap Emitter
Current gain α	~ 0.1	0.9
Power gain	negative	4 db*
Base width w (measured)		$\sim 10^{-4}$ cm
Electron diffusion Length L_n		$\sim 10^{-3}$ cm
Electron lifetime τ_n		$\sim 10^{-8}$ seconds

* Limited by high contact resistances in the emitter and collector.

Acknowledgment

The author gratefully acknowledges the cooperation and contributions of the following colleagues at RCA Laboratories: K. Weiser who prepared the indium phosphide, as well as P. G. Herkart and J. R. Woolston who prepared the gallium arsenide, used in the transistor experiments; J. J. Wysocki who did most of the junction measurements and analysis; and M. Glicksman who made available the results of Hall and conductivity measurements.

Review of Other Semiconductor Devices*

STEPHEN J. ANGELLO†, SENIOR MEMBER, IRE

Summary—**Semiconductor devices other than transistors and rectifying cells are useful and some are becoming important in solid-state electronic-systems technology. It is shown how one possible system of classification for all semiconductor devices can be set up. The system starts by listing a complete set of semiconductor properties and a complete set of external influences which can modify the semiconductor attributes. Devices are then classified in order of complication with respect to semiconductor attributes with external influences being applied singly, in pairs, and so on. It is not claimed that this system is a tool for invention of new devices, but it could serve to organize creative thought. An attempt is made to list criteria for important devices, and a selection is made of those devices which we predict will be important in the future. The selection is Hall effect devices, thermistor devices, photodevices, and thermoelectric devices. Each of these classes of devices is discussed briefly with respect to important properties, and some applications which have been described in the literature.**

* Original manuscript received by the IRE, February 27, 1958.
† Westinghouse Research Labs., Pittsburgh, Pa.

Introduction

A SURVEY of solid-state device development shows that any property of a solid material can be the basis for a device, and in many cases feasibility has been determined. Because of the extent of the field of solid-state devices, we shall restrict this discussion to devices depending upon the physical properties of semiconductors. The word "other" in the title implies, also, that we shall omit discussion of transistors and rectifying cells. These are discussed adequately in other articles in this issue of the PROCEEDINGS.

Some of the devices we shall discuss are, or shall become, important supplementary components to integrate with transistors and diodes into complete, reliable, static electronic systems.

Survey of Physical Properties of Semiconductors

The possibility of constructing solid-state devices arises from the fact that certain internal and external influences can modify the properties of a solid. In keeping with the IRE definition of a semiconductor, which specifies *electronic* conduction, we restrict our attention to *physical* properties. Ionic transport and chemical effects are excluded.

We attempt here to give a complete (but not unique) set of basic postulates which underlies the present theoretical description of the physical properties of semiconductors. Device characteristics result from alteration in detail of these attributes.

List of Postulates

1) A semiconductor is one of several classes of solids which are an ensemble of atoms and electrons that obey the laws of quantum mechanics.
2) Electrical conduction is by electrons (negative charges) and holes (positive charges). These have density n electrons/cm^3 and p holes/cm^3. The densities can be varied over a wide range by internal influences, such as the density of atoms foreign to the pure solid, or by external influences, for example, absolute temperature.
3) Current carriers have effective masses m_e and m_p. These masses are often anisotropic depending upon crystal direction.
4) Current carriers have mobility μ_n and μ_p cm^2/volt-second.
5) There exists in the solid a band of energy states which are unallowed for the conduction electrons. This gap separates energy states which are full of mobile electrons at absolute zero of temperature, and energy states comprising a conduction band.
6) Minority carriers in a given semiconductor region can exist in excess of equilibrium density for a finite lifetime τ seconds.

Semiconductor properties appropriate for the description of devices will be found to be related to these basic attributes. For example, electrical conductivity is

$$\sigma = nq\mu_n + pq\mu_p,$$

where q is electronic charge. Also,

$$\sigma = \sigma_0 \exp\left(-\frac{E}{2kT}\right),$$

where E is the unallowed band gap and T is absolute temperature. The exact form of the relation depends on the conditions relevant for the description, and the approximations introduced.

Internal influences which can modify the basic attributes are: 1) density of various impurities in a crystal, 2) spatial distribution of impurities, 3) density and distribution of crystal defects, and 4) crystalline phase changes. The external influences are: 1) applied electric field, 2) applied magnetic field, 3) absolute temperature, 4) spatial distribution of absolute temperature or the flow of heat, and 5) radiation.

To illustrate the use of the classifications above, we shall apply them in describing some well-known devices. For example, if the impurity densities in a germanium crystal are adjusted to give three regions in series p-n-p, and the spatial distribution is appropriate for minority carrier injection at the p-n transition and appropriate for supporting high blocking voltage at the n-p junction, we have a transistor. Application of two electrical fields will enable the device to act as an amplifier in the appropriate circuit. Visible radiation shining upon one junction with an electric field appropriately applied will result in a phototransistor. In the first case, only one type of external influence was applied. The second case involves two types of external influence applied simultaneously. In this way it may be seen that the number of possible semiconductor devices is very large. We shall try to select, for more complete discussion, devices which are new and show future promise as supplementary system components. Some well-established devices will be covered by references to review literature.

Criteria for Device Selection

In general, a device will be found to be important if the desired effect is "large." The term "large" is, of course, relative to the requirements of the receptor of the device response. In some cases an "efficiency" is the important factor and is judged relative to the efficiencies of competing devices, and often with respect to the economics of device operation.

We may illustrate the discussion above by citing two devices, each of which has been known for over 50 years, but only now are becoming interesting because new materials and techniques make large effects possible. These are the Hall effect generator and thermoelectric generator. More will be said about these devices later in this review.

With these ideas in mind, we may range through the various possible semiconductor devices and choose. The ranging will be more orderly if we start with homogeneous semiconductors and apply the external influences one at a time. The next more complicated devices will be homogeneous with a pair of external influences, a triple, and so on. A single p-n junction will be the simplest departure from homogeneity, and the external influences will be applied singly, in pairs, and so on.

Table I is designed to put in order some of the devices which can be made from a homogeneous semiconductor.

In Table II we shall consider one p-n junction in the homogeneous, isotropic semiconductor. To conserve space and patience we shall list only those influences which result in devices known to the art. The list is not intended to exhaust all possible devices.

TABLE I

HOMOGENEOUS ISOTROPIC SEMICONDUCTOR DEVICES

Input External Influence*	Number of Electrodes	Effect	Name of Device (Where Applicable)
E	2	$I=\sigma E$, Ohm's law	Resistor
H		no effect	
T		no effect	
$T(x, y, z)$	2	Seebeck, $E=\text{const}_1\Delta T$	
$h\nu$		no effect	
E, H	2	$I=\sigma(H)E$	Magnetoresistor
E, H†	4	Hall effect $V=f(H, E)$	Hall generator
E, T	2	$I=\sigma(T)E$	Thermistor
$E, T(x, y, z)$	2	Thomson Heat $Q_T=\text{const}_2 I\Delta T$	
$E, h\nu$	2	$I=\sigma(h\nu)E$	Photoresistor
E_1, E_2, H	3	Suhl effect	

* Other external influences are zero or constant, that is, they are treated as parameters but not as variables.

† Many other galvanomagnetic effects exist which are not listed here.

E = electric field strength,
H = magnetic field strength,
T = absolute temperature degrees Kelvin,
$h\nu$ = quantum of radiation: X-rays, γ-rays, β-rays, α-rays, infrared and other electromagnetic radiations,
Δ = change of the appropriate influence.

TABLE II

ISOTROPIC SEMICONDUCTOR DEVICES WITH ONE p-n JUNCTION

Input External Influence	Number of Electrodes	Effect	Name of Device
E	2	$I=f(E)$, $h\nu=f(E)$	Rectifying cell Electroluminescent cell
E_1, E_2	4	Minority carrier injection in a filament	Filamentary transistor or double base diode
$T(x, y, z)$	2	Seebeck effect	Thermocouple, Thermoelectric Generator
$h\nu$	2	Photovoltaic effect	Photocell, Solar battery
$E, T(x, y, z)$	2	Peltier effect $Q_P=\pi I(E)$	Peltier refrigerator
$E, h\nu$	2	$I=f(E, h\nu)$	Photodiode

It will now be clear to the reader that this is a possible scheme to classify all semiconductor devices. By extending the list of effects and influences, other solid-state devices may be classified similarly.

DISCUSSION OF NEW AND PROMISING DEVICES

Hall Effect Generators

It has been well known for over 60 years that an external magnetic field can influence the equipotentials of a conductor carrying current. In the usual configuration a plate of length L, width W, and thickness d is fitted with electrodes at each end so that a current flows along the length. Equipotential points across the width are moved by the influence of a magnetic field perpendicular to both the current and the equipotential lines. Terminals which are on equipotentials with zero magnetic field, have a "Hall voltage"

$$V_h = \frac{RHI}{d}$$

and appear when a field H is applied. The factor "R" is the Hall constant which depends upon the material under study. The usefulness of the effect depends on the factors H and I which appear as a product. Moreover, the Hall voltage appears as a linear function of this product.

In order for a material to be used in a practical device, three requirements[1] must be fulfilled:

1) The Hall voltage must be sufficiently large—at least several hundred millivolts, that is, the Hall constant must be large.
2) The Hall voltage should be only weakly temperature dependent.
3) It must be possible to take power from the Hall voltage terminals, that is, the resistance of the material must be low.

Only two materials are known at the present time which make practical utilization of the Hall effect possible; these are, Indium Antimonide[2] and Indium Arsenide.[3] It is possible that some design adjustments may be made between the parameters R and ambient temperature by specification of a mixed crystal of $InAs_xP_y$.[4]

According to our earlier discussion of classification, we have here a four-terminal device with an electric field applied to one pair of terminals, and an applied magnetic field with a useful output at the second pair of terminals. The Hall effect in a generator device has been discussed in detail by Kuhrt.[5] For practical applications the effect of loading the Hall terminals is important. The geometry of the Hall plate can be selected for best results in a given application. Practical designs must deal with temperature effects and the influence of the magnetic field upon the internal resistance of the generator.

Three basic applications of the Hall generator are indicated by the properties of the device:[1]

1) Multiplication for example:
   $V_h(t)=\text{const.}\ H(t)I(t)$. If $H(t)$ is proportional to the current in a circuit and $I(t)$ is proportional to the voltage, $V_h(t)$ will be proportional to instantaneous power.
2) Control of the Hall voltage by means of the magnetic field. If the primary current in a Hall generator is held constant, the Hall voltage can give a true picture of a magnetic field in space in both magnitude and direction (*i.e.* the Hall effect is proportional to the component of a magnetic field perpendicular to the current and equipotentials).

[1] W. Hartel, "Anwendung der Hallgeneratoren," *Siemens-Z.*, vol. 28, pp. 376–384; September, 1954.
[2] H. Weiss, "Über die elektrischen Eigenschaften von InSb," *Z. Naturforsch*, vol. 8a, pp. 463–469; August, 1953.
[3] O. G. Folberth, et al., "Die elektrischen Eigenschaften von InAs II," *Z. Naturforsch.*, vol. 9a, pp. 954–958; November, 1954.
[4] O. G. Folberth, "Mischkristallbindung bei $A^{III}B^{V}$-Verbindungen," *Z. Naturforsch.*, vol. 10, pp. 502–503; June, 1955.
[5] F. Kuhrt, "Eigenschaften der Hallgeneratoren," *Siemens-Z.*, vol. 28, pp. 370–376; September, 1954.

3) Control of the Hall voltage by means of the primary current. If the magnetic field is held constant, the variation of V_h with time is a true picture of the time variation of the current.

Multiplication of any two quantities can be accomplished if one can be made to influence proportionally the magnetic field, and the other to influence the Hall current. Since instantaneous quantities are involved, the product can take phase relations into account. With more than one generator, three or more quantities can be multiplied. In addition, circuits can be arranged to form the reciprocal of a quantity, and from this, quotients of quantities can be formed.

An interesting application to electrical machinery is the measurement of the torque of a dc motor under load.[6] A Hall generator plate is mounted in a pole face of the motor so that the motor flux is perpendicular to the Hall current and the plane of the Hall terminals. Hall current is made proportional to the armature current. The Hall voltage will then be proportional to the product of the motor flux and armature current or the inner torque. Several uses for such a Hall voltage will come to mind, for example, a relay could be actuated to act as protection against overloads on the motor. Also the Hall voltage could be used as a field current control with suitable amplification making it possible to develop a constant torque over a range of load values.

Variation of the Hall voltage with magnetic field; primary current constant is utilized in a device for measuring dc current in a heavy buss. A magnetic core is built around the current-carrying buss with a Hall generator in the air gap. In the arrangement there is a relation between the field H and the buss current. With a fixed primary current the Hall voltage is a measure of the buss current. It is possible in this way to measure thousands of amperes to an accuracy of one per cent or less without breaking into the buss structure. Another obvious application for this device is a small device for the measurement of magnetic fields[7] without the need of mechanical motion or variation of the magnetic field.

It appears that the Hall effect generator will have many applications in the future for sensing elements and simple analog computations such as the input for transistorized systems.

Thermistors

A thermistor is an electron device which makes use of the change of resistivity of a semiconductor with change in temperature. This was one of the first of the devices developed in the past 18 years of semiconductor development. The best-known review of these devices was written by Becker, et al.[8] In this, a number of interesting applications are outlined, all depending upon the large negative variation of resistivity of these devices. Recently, basic concepts of semiconductor physics have been applied to obtain new and useful modifications of the classical thermistor characteristics.

The classical thermistor has a large negative temperature coefficient of resistance of the order of a per cent per degree centigrade, that is, the resistance decreases with increasing temperature. It would be highly desirable from the standpoint of device design to be able to preassign the following factors:

1) resistivity,
2) magnitude of the temperature coefficient, and
3) sign of the temperature coefficient.

Some progress has been made in the ability to preassign resistivity and temperature coefficient by Heikes and Johnston[9] who have made a systematic study of valency control by lithium substitution in transition metal oxides. Resistivity is controlled by lithium composition, and temperature coefficient is controlled by forming mixed crystals of oxides. Very large values of negative temperature coefficient have been obtained[10] by proper doping of InSb. Two per cent per degree Fahrenheit is claimed. Interesting developments[10] in positive temperature coefficient thermistors have also been made. Large coefficients have been obtained in PbSe, for example. More interesting, however, are systems like barium-strontium titanate modified with lanthanum which have been made to exhibit positive values of eight per cent per degree Fahrenheit. Moreover, the increase is from 500 ohm-cm at 160°F to 10^6 ohm-cm at 300°F. The increase in resistivity in this case is thought to be due to a cubic-to-tetragonal phase change in barium titanate. A characteristic of this form could be used as a temperature actuated on-off switch.

It is clear that the requirements of electronic systems cannot be served by temperature-independent resistance elements alone. There is extensive need for resistance elements which will provide primary temperature sensing, and will act as control elements in solid-state device circuitry.

Photocells

According to the classification scheme of this paper, there are three main types of photocells: these are, 1) the photoconductive type in which an electric field is applied to a homogeneous semiconductor, and incident radiation causes a change in resistivity by influencing current carrier density, 2) the photodiode type in which

[6] F. Kuhrt and E. Braunersreuther, "Drehmomentnessung an einem Gleichstrommotor mit Hilfe des Halleffektes," *Siemens-Z.*, vol. 28, pp. 299–302; August, 1954.

[7] H. Hieronymus and H. Weiss, "Über die Messung kleinster magnetisher Felder mit Hallgeneratoren," *Siemens-Z.*, vol. 31, pp. 404–409; August, 1957.

[8] J. A. Becker, et al., "Properties and uses of thermistors-thermally sensitive resistors," *Trans. AIEE*, vol. 65, pp. 711–725; November, 1946.

[9] R. R. Heikes and W. D. Johnston, "Mechanism of conduction in Li-substituted transition metal oxides," *J. Chem. Phys.*, vol. 26, pp. 582–587; March 1957.

[10] R. K. Willardson, "New semiconductor materials," *Battelle Tech. Rev.*, vol. 6, pp. 8–14; August, 1957.

the leakage current of a p-n junction biased in the reverse direction is modulated by intensity of incident radiation, and 3) the photovoltaic type in which incident radiant results in hole-electron generation at the junction, and charge separation by virtue of the contact potential. Other types of photocells are possible by utilizing phenomena associated with incident radiation, for example, the photomagnetoelectric effect has been used as an infrared radiation detector.[11]

The photoconductive type of device is used extensively in infrared detectors and will continue to be used in this way because of the detection sensitivity which can be obtained. Further development of semiconductor materials will result in more sensitivity and faster response time.[12]

Photodiodes show a family of leakage vs reverse voltage curves with light intensity as the parameter, which are reminiscent of transistor collector characteristics with the emitter bias as a parameter.[13] The devices may be made with microsecond response and have been used to read punch cards and punched tapes.

The photovoltaic cell provides a primary-power-source converting incident radiation into electric current. The silicon solar cell is presently the most powerful of this class of device, and has been made to convert sunlight with 11 per cent efficiency. The present status has been reviewed by Prince.[14] The outlook is promising for utilization; the main problem is one of economics to compete with existing power sources.

We feel that there is a good future for application, not only as a power source, but also as a photoelectric link in electronic systems.

Thermoelectric Devices

It has been well known for over 100 years that a circuit consisting of two dissimilar electrical conductors can show several thermoelectric effects. If one junction is placed in a heat reservoir at a temperature different from the other junction, a current will flow in the circuit. The current is due to the generation of a Seebeck emf. Heat is absorbed at the hot junction and rejected at the cold junction. A fraction of this heat flowing through the system may be converted into electric power in a load. The inverse effect also occurs, namely, a battery inserted in the circuit causes current to flow through the junctions. Heat will be absorbed at one junction causing it to cool, and heat will be rejected at the other junction. The creation of a temperature gradient by application of an electric field to a junction is known as the Peltier effect.

Although the effects have been known for many years, little application has been made because the efficiency and size of the temperature gradient possible are small in the case of metals. Very carefully formulated semiconductors are being developed in which the conversion efficiency is approaching 10 per cent and the cooling power is approaching 90°C below the hot junction temperature.[15]

The maximum amount of work which a machine can extract from a hot reservoir at T_h°K is determined by

$$\epsilon_c = \frac{T_h - T_c}{T_h},$$

where ϵ_c is the Carnot efficiency and T_c is the heat sink temperature. The actual efficiency of a device working between these two temperatures is modified by the device efficiency so that

$$\epsilon = \epsilon_c \epsilon_{\text{device}},$$

where ϵ is the actual over-all conversion efficiency.

In the case of a thermoelectric circuit the over-all efficiency is

$$\epsilon = \epsilon_c \epsilon_{TC} \doteq \epsilon_c \frac{T_h \alpha^2}{4\rho k}.$$

The device efficiency has been replaced by a figure of merit

$$M = \frac{T_h}{4} \frac{\alpha^2}{\rho k},$$

where α is the thermoelectric power, ρ is resistivity, and k is thermal conductivity. The substitution

$$\epsilon_{TC} \doteq M$$

is good up to about 10 per cent efficiency. It will be noted that the semiconductor must have high α concomitant with small ρ and k. The value of M is not easy to maximize because α and ρ are related to oppose maximization of M. The best compromise is found in materials having of the order of 10^{19}/cm^3 current carriers, that is, very highly doped semiconductors or semimetals. An understanding of transport in semiconductors has been fruitful in decreasing k without increasing ρ.[16]

The same figure of merit applies to refrigerators and heat pumps. The maximum temperature difference between T_h and T_c is given by

$$\Delta T_{\max} = 2MT_c.$$

Values of $M = 0.2$ are now available and better values will result from intensive research now being carried on.

The effect upon electronic systems by the thermoelectric phenomena will be in three areas: 1) primary power source to compete with batteries, solar cells, and others, 2) cooling of electronic equipment in small, critical volumes throughout the system, and 3) devices depending directly upon the effects. The first requires no elaboration except to say that the usual heat sources will be supplemented by isotopes and nuclear reactors.

[11] O. H. Lindberg and H. C. Chang, "An indium antimonide PME infrared detector," presented at IRE-AIEE Semiconductor Devices Conf., Boulder, Colo.; July 17, 1957.
[12] A. Rose, "Maximum performance of photoconductors," *Helv. Phys. Acta*, vol. 30, pp. 242–244; August, 1957.
[13] J. N. Shive, "The properties of germanium phototransistors," *J. Opt. Soc. Amer.*, vol. 43, pp. 239–244; April, 1953.
[14] M. B. Prince, "Silicon solar cells," *Electronic Ind.*, vol. 16, pp. 60–61; March, 1957.
[15] A. F. Ioffe, "Semiconductor Thermoelements and Thermoelectric Cooling," Infosearch Ltd., London, Eng., 1957.
[16] A. F. Ioffe, "Heat transfer in semiconductors," *Can. J. Phys.*, vol. 34, No. 12A (supplementary number), pp. 1342–1355; December, 1956.

Under item 2 we point out that cooling of small volumes is not accomplished economically with mechanical systems. Under item 3 we mention a hygrometer depending upon the cooling of sample of air by a Peltier junction.[15]

Discussion

It can be seen from the foregoing development that it is possible to order all semiconductor devices (and all solid-state devices) by applying external influences to a semiconductor singly, in pairs, in triplets and so on. The semiconductor is first homogeneous, second, contains one *p-n* junction, and so on. We do not claim that such a system will serve to invent new semiconductor devices, but it will serve to give order to creative thinking along hese lines.

It is clear that already many applications have been found for semiconductor devices, and that more development will integrate these, and devices not yet invented, into reliable solid-state electronic systems.

Bibliography

It is not practical to include a complete bibliography, but the following papers have been cited by a reviewer as worthy of notice:

Goldsmid, H. J. "Thermoelectric Applications of Semiconductors," *Journal of Electronics*, Vol. 1 (1955–1956), pp. 218–222.
Ross, I. M., and Saker, E. W. "Applications of Indium Arsenide," *Journal of Electronics*, Vol. 1 (1955–1956), pp. 223–230.
Ross, I. M., and Thompson, N. A. C. "An Amplifier Based on the Hall Effect," *Nature*, Vol. 175 (1955), p. 518.
Saker, E. W., Cunnell, F. A., and Edmond, J. T. "Indium Antimonide as a Fluxmeter Material," *British Journal of Applied Physics*, Vol. 6 (1955), pp. 217–220.
Wick, R. F. "Solution of the Field Problem of the Germanium Gyrator," *Journal of Applied Physics*, Vol. 25 (June, 1955), pp. 741–756.

Electrons, Holes, and Traps*

WILLIAM SHOCKLEY†, FELLOW, IRE

Summary—**The statistics of recombination and of trapping of electrons and holes through traps of a single species are presented. The results of the Shockley-Read recombination theory are derived and more fully interpreted. A level of energy known as the equality level is introduced. When the Fermi level lies at this level, the four basic processes of electron capture, electron emission, hole capture, and hole emission all proceed at equal rates. Transient cases for large trap density are presented.**

I. Introduction

RECOMBINATION centers or traps seem destined to play an increasingly important role in semiconductor devices. The variation of alpha with current in silicon transistors appears to be dependent upon the presence of traps lying fairly near the center of the energy gap.[1] This effect of variation of alpha with current plays an essential role in *n-p-n-p* switching transistors[2] or 4-layer diodes.[3]

Studies of the properties of many chemical elements in silicon and germanium have been published.[4] Some of these can exist in states of multiple charge. For example, a gold atom in germanium apparently behaves much like an acceptor capable of binding three holes. Ionizing all the holes leaves an ion with three negative charges which can probably trap a hole effectively.

In this paper, we do not deal with the statistics of the complex cases in which the trap may make transitions between more than two conditions of charge. The discussion centers on the simpler case, in which the trap may exist in two states of charge only. The more negative of these will be referred to as *filled*, and the less negative or more positive as *empty*. The emphasis in the paper is on the statistics of the transition.

The statistics for transitions of a trap between its two charge conditions are basic to understanding the more complex cases of multiply charged traps. There is a rather disconcerting aspect to these statistics, and this accounts for the length of this article. Although a given type of trap can be described by only four interaction constants, the variety of situations is so great that it proves difficult to visualize the relationships involved. Considering the basic simplicity of the problem, working out the details proves to be surprisingly complex.

Underlying the treatment is the *principle of detailed balance*, which is of great importance in atomic process. One of the intuitively most appealing descriptions of it was, we believe, first given by J. C. Slater. His presentation is as follows: When a system has reached thermal equilibrium, it has run down and is no longer changing. Past and future are alike to it. Now imagine that a motion picture is made of the system, showing the atoms and electrons in detail. This film can be projected backwards in time and, since past and future are alike,

* Original manuscript received by the IRE, March 24, 1958.
† Shockley Semiconductor Lab., Beckman Instruments, Inc., Mountain View, Calif.
[1] C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in *p-n* junctions and *p-n* junction characteristics," PROC. IRE, vol. 45, pp. 1228–1243; September, 1957.
[2] J. L. Moll, M. Tanenbaum, J. M. Goldey, and N. Holonyak, "*P-n-p-n* transistor switches," PROC. IRE, vol. 44, pp. 1174–1182; September, 1956.
[3] W. Shockley, "Unique properties of the four-layer diode," *Electronic Ind. and Tele-Tech.*, vol. 16, pp. 58–60, 161–165; August, 1957.
W. Shockley and J. F. Gibbons, "Introduction to the four-layer diode," in "Semiconductor Products," vol. 1, pp. 9–13; January/February, 1958.
[4] See W. C. Dunlap, Jr., "An Introduction to Semiconductors," John Wiley and Sons, Inc., New York, N. Y.; 1957. Also, C. T. Sah and W. Shockley, "Electron-hole recombination statistics in semiconductors through flaws with many charge conditions," *Phys. Rev.*, vol. 109, pp. 1103–1115; February 15, 1958. (References.)

the observer will not be able to tell the difference. Now suppose, for example, the forward-running picture shows on the average (Cap. n) electrons being captured per unit volume per unit time on traps giving up the energy in the form of heat waves (phonons), and suppose it shows (Em. n) electrons being emitted per unit volume per unit time, acquiring the energy from heat waves. Then, if the picture is run backwards, each actual emission process will appear to be a capture, and there will appear thus to be (Em. n) captures per unit time. Thus, if (Em. n) is not equal to (Cap. n), forward and backwards running of the film can be distinguished, contrary to the assumption that the system is run down. Thus, the principle of detailed balance requires that each process and its reverse proceed at equal rates. Extensive use is made in this article of the principle of detailed balance.

Much of the treatment in the following sections repeats results derived from the original Shockley-Read[5] recombination theory. The case of transient disturbances, recently published by Sandiford,[6] is also presented. One new case, Section IX, deals with periodic effects such as minority carrier flow through a base layer.

A new concept, the *equality level*, is introduced. The understanding of this energy level makes it easier to visualize how the dependence of lifetime on electron and hole densities arises.

In some cases it is easier to visualize how certain effects occur in terms of equivalent circuits, and in other cases an atomic picture is more helpful. Both of these are presented, and an attempt is made to show their relationship.

II. Simple Derivation of the Lifetime Equation

In this section, we give a derivation of the equation for the lifetime of holes in n-type material and for electrons in p-type material. The emphasis is on the physical meaning of the resulting equations and on the steps in the derivation. Questions of mathematical rigor have been left for later sections. We also introduce two energy levels on the basis of physical interpretation; the mathematical definition of these levels are given later.

The treatment given here is restricted by several assumptions. The most basic is the assumption that we are dealing with nondegenerate semiconductors so that the velocity distribution of the electrons is independent of their density and the same is true for holes. The treatment is also restricted to small disturbances from equilibrium and to small densities of traps.

The basic processes with which we are concerned in this section are illustrated in Fig. 1. Conventional plus and minus symbols are used for holes and electrons, and donors are represented by encircled plus signs. The traps

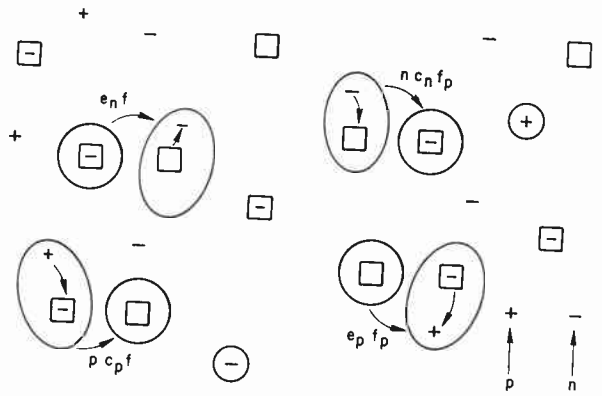


Fig. 1—The four basic trapping processes.

are represented by squares and are shown as being either neutral or possessing a minus charge. Entirely similar results will apply, however, to cases in which the trap may change from a plus charge to zero as a result of the transitions represented on the figure. Before considering the transitions represented by the symbols with arrows, let us introduce a few definitions. The hole and electron densities are represented by the customary symbols n and p, and the density in an intrinsic sample by n_i. In accordance with the assumption that the semiconductor is nondegenerate, the mass-action law holds:

$$\textbf{mass action law:}\quad np = n_i^2. \tag{1}$$

It is supposed that the density of traps is N_t:

$$N_t = \text{density of traps.} \tag{2}$$

We shall let the symbol f represent the fraction of the traps in the more negative state, and f_p, the fraction in the more positive state. The relationship between these quantities is evidently

$$f_p = 1 - f. \tag{3}$$

Under equilibrium conditions, f is simply the Fermi factor for the traps.

Returning to Fig. 1, consider the process represented by $e_n f$. This process corresponds to spontaneous emission of an electron from a negatively charged trap into the conduction band. The trap then changes to the neutral condition, as represented schematically on the diagram. (The trap does not, of course, move in space.) The rate at which this process goes on per unit volume is simply proportional to the total number of traps in the negative condition. (We do not consider the dependence of the process upon temperature or other factors here.) The symbol e_n represents the rate at which this process would go on per cubic centimeter in the presence of N_t traps if all of the traps were negatively charged. Thus, e_n is actually proportional to the specific emission property for an individual trap times the trap density.

The inverse process is represented by the symbol $nc_n f_p$ on the figure. In this case, the capture of electrons into neutral traps is evidently proportional to the number of electrons present, and the fraction of the traps which are in a condition to capture electrons. If all of the N_t traps were in the more positive condition of

[5] W. Shockley and W. T. Read, Jr., "Statistics of recombinations of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835–842; September 1, 1952.

[6] D. J. Dandiford, "Carrier lifetime in semiconductors for transient conditions," *Phys. Rev.*, vol. 105, p. 524; January 15, 1957.

charge, then the rate of capture of electrons would be simply nc_n. Evidently this should be multiplied by the fraction f_p of the traps which are in this condition.

Similar definitions apply to the quantities related to holes.

The model considered thus involves four processes: electron emission, electron capture, hole emission, and hole capture. The quantities describing these processes are not all independent, and we shall shortly derive a relationship among the four constants.

To achieve this derivation, we apply the *principle of detailed balance*, which is one of the most powerful tools in dealing with statistical mechanical problems. As discussed in Section I, it states that under equilibrium conditions every process and its reverse must proceed at exactly equal rates. In the case of Fig. 1, it requires the following two equalities:

$$nc_n f_p = e_n f \tag{4}$$

$$pc_p f = e_p f_p. \tag{5}$$

Each of these equations may be solved for the ratio f/f_p with the result

$$f/f_p = c_n n/e_n = e_p/c_p p. \tag{6}$$

We shall use this equation below in order to eliminate f in subsequent equations. At this point, however, we note that the ratio of the last two terms in (6) must be unity, and since, under the equilibrium condition, considered n times p satisfies the mass-action law, we obtain from (6)

the detailed-balance relationship: $e_n e_p = c_n c_p n_i^2$. (7)

This is the relationship among the four constants describing the trap. We shall refer to it frequently in following derivations.

We next define two important pairs of densities: n_1, p_1 and n^*, p^*. Both of these pairs of densities satisfy the mass-action law. The density pair n_1, p_1 corresponds to the condition in which half of the traps are in the more negative condition and half are in the more positive condition. Thus it corresponds to values of $\frac{1}{2}$ for f and f_p. This means that the Fermi level for this particular case lies at the energy level of the trap. The mathematics of this relationship are given in a subsequent section. Here we simply take the definition of the densities as being that just given. Referring to (4) and (5), we see that the required values are:

$$n_1 = e_n/c_n; \qquad p_1 = e_p/c_p. \tag{8}$$

Since by definition these two densities are possible equilibrium densities, the product of the two equations in (8) leads to n_i^2 and, again, to the detailed-balance relationship (7) among the four constants. It is convenient to express the ratio f/f_p in terms of n_1 and p_1:

$$f/f_p = n/n_1 = p_1/p. \tag{9a}$$

The other pair of densities, n^* and p^*, are defined in a somewhat similar way. They are also possible equilibrium densities and would prevail in the situation in which all four of the processes represented on Fig. 1 proceed at exactly the same rate. At first it may be surprising that a situation can exist in which all four processes proceed at the same rate. However, it should be recalled that from the principle of detailed balance, the two electron processes must always proceed at exactly equal rates under conditions of thermal equilibrium and so must the two hole processes. Furthermore, as the semiconductor is made progressively more n-type, the rate at which the electron processes proceed continually increases, and the rate at which the hole processes proceed continually decreases. Thus, in general there will be some density of donors or acceptors at which all four of the processes proceed at equal rates. The density n^* corresponding to this condition is easily found by requiring equality of the rates of the two processes which causes the traps to become more negatively charged, namely, electron capture and hole emission. Equating these two rates leads at once to the expression for n^*, and a similar process leads to the equation for p^*:

$$n^* = e_p/c_n; \qquad p^* = e_n/c_p. \tag{9b}$$

In accordance with their definition, these two densities correspond to equilibrium condition and thus satisfy the mass-action law. Multiplying the pair of equations in (9b) together thus leads once more to the detailed-balance relationship (7) among the four constants.

We refer to n^* and p^* as the *equality densities*. A semiconductor having these densities is said to be in the equality condition in regard to the particular traps involved.

[It should be noted that if we invent a new group of traps having the same c_n and c_p as the old group but having their emission constants interchanged, the new group will satisfy the detailed balance condition, and will have the role of (n_1, p_1) interchanged with that of (n^*, p^*). This is the basic reason for the confusion between E_t and E^* discussed below with Fig. 2. See also (163).]

We are now in a position to apply the concepts discussed to the calculation of the lifetime of injected carriers. For this purpose, suppose that light is falling on the specimen, such that hole electron pairs are being created at a rate of g_L per unit volume. Suppose that the only processes which can occur in order to restore equilibrium are those represented on Fig. 1. Under these conditions, the electron and hole densities will be increased from their equilibrium values n and p, to larger values $n+\delta n$ and $p+\delta p$. Let us suppose that we are dealing with an n-type specimen and concentrate our attention on the minority carriers or holes which are present. The definition of steady-state lifetime which we use in this section for the holes is given by

$$\delta p = \tau_p g_L. \tag{10}$$

This equation states that holes are being generated at a rate g_L and live, on the average, for a time τ_p, so that the total accumulation of additional holes is $\tau_p g_L$ and this is equal to δp. The presence of δp extra holes means that the detailed-balance between hole capture and hole emission is disturbed, and a net rate of hole capture arises, which is just equal to the rate of generation g_L.

In order to calculate the lifetime τ_p, we accordingly calculate the net rate of capture of holes due to δp.

Since we are dealing with small disturbances, we can assume that each of the extra holes in the added density δp finds itself in substantially the same environment as would a normal hole. In other words, it finds a fraction f, equal to the equilibrium value, of the traps ready to capture holes. Accordingly, the extra rate of capture produced by the added holes is given by

$$\text{extra rate of capture} = \delta p c_p f. \tag{11}$$

This capture is not 100 per cent efficient, however, because there is some probability that a captured hole will be re-emitted and have to be captured again. Thus we must correct the extra rate of capture given by (11) by the effectiveness of the capture process. This effectiveness may be calculated by considering the relative probability that a positively charged trap will be made negative by electron capture rather than by hole emission. This ratio is controlled by the electron density n. The effectiveness of capture is thus:

$$\text{effectiveness of capture factor} = c_n n/(c_n n + e_p). \tag{12}$$

In (12), it is evident that the numerator is proportional to the rate of capture of electrons by a trap in the more positive condition, and the denominator is proportional to the total rate at which the trap becomes more negative. Thus, (12) is the effectiveness with which the hole capture described by (11) must be multiplied to find the total rate of recombination of hole electron pairs.

(This method of treating the effectiveness of capture factor leads to the correct answer, but the reasoning is actually not perfectly valid. It should be noted that the capture of the extra holes δp disturbs the distribution of the traps. This has two effects on the interaction with holes. In the first place, the increased number of empty traps gives rise to an additional emission of holes, and this effect is included in (12). The disturbance has another effect as well; it decreases the capture rate of normal holes which are present. In considering the effectiveness factor, only the former term was considered. The neglect of the decrease in capture of the normal hole density is, however, exactly compensated by another neglected term. This compensating term is the decrease in emission of electrons due to the decrease in the number of negatively charged traps. These neglected factors are in the same ratio as those considered, and thus their inclusion does not alter the effectiveness of capture factor. Thus, the fact that the particular effects considered in (12) lead to the correct ratio turns out to be a consequence of the detailed-balance equations, (4) and (5), which hold in the equilibrium condition. We return to this point more fully in Section V, which deals more rigorously with the theory of small disturbances.)

This total rate of recombination must be equated to the rate of generation in the steady-state condition. Accordingly, we obtain

$$g_L = \delta p c_p f c_n n/(c_n n + e_p). \tag{13}$$

This equation may be rewritten in terms of the densities n_1 and n^*, discussed above. From (9b) and (3) we obtain

$$f/(1-f) = n/n_1 \tag{14}$$

$$f = [1 + (n_1/n)]^{-1}. \tag{15}$$

The effectiveness factor (12) may be written in the form

$$n/[n + (e_p/c_n)] = n/(n + n^*) = [1 + (n^*/n)]^{-1} \tag{16}$$

from which it is seen that if the semiconductor is in the equality condition, the effectiveness of capture is 50 per cent. If it is more n-type than the equality condition, the effectiveness of capture approaches 1; if it is less n-type, the effectiveness of capture becomes small, being equal to n/n^*.

Using the definition given earlier for lifetime and combining (13), (15), and (16), we obtain

$$\tau_p = \delta p/g_L = \tau_{po}\left(1 + \frac{n_1}{n}\right)\left(1 + \frac{n^*}{n}\right) \tag{17}$$

in which the limiting lifetime τ_{po} is

$$\tau_{po} = 1/c_p. \tag{18}$$

Evidently τ_{po} is the lifetime in a heavily doped n-type sample in which every trap is in the negative condition (*i.e.*, $n \gg n_1$), and there is no probability of a captured hole being re-emitted before it is eliminated permanently by an electron (*i.e.*, $n \gg n^*$).

Eq. (17) has an evident interpretation in terms of the mechanisms we have just discussed. The two parenthetical factors represent an increase in lifetime due to a decrease in effectiveness of the recombination process. The first factor corresponds to the average number of times that an added hole must approach traps before encountering a negatively charged trap. If n is substantially greater than n_1, the Fermi level lies well above the energy level of the trap, and all of the traps are negatively charged. In this case, the factor is unity. On the other hand, when $n = n_1$, half of the traps are in the empty condition, and only half of the traps can capture holes. Under these conditions, the number of approaches to traps is doubled, and so is the lifetime. For smaller values of n, the fraction of the traps which are in a condition to capture holes varies directly as n, and the number of times a hole must approach a trap in order to be captured thus varies inversely as n. This corresponds to the limiting form when n is substantially less than n_1.

The second factor in (17) plays a similar role, but its mechanism is different. It represents the number of times a hole must be captured before it finally recombines. If n is substantially greater than n^*, then electron interactions with the traps proceed at a much greater rate than hole interactions. Under these conditions, it is very likely that a hole captured by a trap will combine with an electron before it is reemitted, and in this case the factor of (17) is unity. On the other hand, if n is equal to n^*, an empty trap (*i.e.*, a trapped hole) is equally likely to become negative due to electron capture and hole emission, and consequently hole capture is

only 50 per cent effective in producing recombination with the consequence that the lifetime is doubled over what it would be if every captured hole were recombined.

The fact that the two parenthetical expressions of (17) are similar in form means that, by studying the dependence of lifetime upon doping level, it is impossible to distinguish, in some cases, between the energy level and the equality level of a trap. This point is discussed below in connection with the figures in this section, and a method of distinguishing is discussed in Section VI.

Precisely similar reasoning can be carried out for the case of electron capture in a p-type specimen. This leads to

$$\tau_n = \tau_{no}\left(1+\frac{p_1}{p}\right)\left(1+\frac{p^*}{p}\right) \tag{19}$$

$$\tau_{no} = 1/c_n. \tag{20}$$

We next consider how the lifetimes given by (17) and (19) depend upon the impurity density and hence the Fermi level of a specimen. For this purpose, we use diagrams which show how the logarithm of lifetime varies with Fermi level. In (19) three types of behavior are possible:

1) p may be greater than both p_1 and p^*; in this case, τ_n is constant and equal to τ_{no};
2) p may be greater than one of the two quantities and smaller than the other; in this case τ_n varies inversely as p;
3) p may be smaller than both of the two quantities p_1 and p^*; in this case, τ_n varies inversely as p^2.

Similar results apply to t_p in which case the effects depend upon the relative values of n, n^*, and n_1. We illustrate below how these three possibilities affect the appearance of the lifetime vs Fermi level curve.

We consider next the limiting values approached by τ_p and τ_n as an intrinsic specimen is approached. The assumptions on which the τ's are based are not valid for an intrinsic specimen; however, it is pertinent to consider the values obtained by extrapolating the formulas for τ_n and τ_p to the intrinsic case. Actually, the two extrapolated values are equal at the intrinsic condition for which both p and n are equal to n_i. This result may be readily established by carrying out the following series of manipulations which show the identity of the two limiting forms:

$$\begin{aligned}\tau_p &= \tau_{po}\left(1+\frac{n_1}{n_i}\right)\left(1+\frac{n^*}{n_i}\right)\\ &= \frac{n_1 n^*}{n_i^2 c_p}\left(\frac{n_i}{n_1}+1\right)\left(\frac{n_i}{n^*}+1\right)\\ &= \frac{e_n e_p}{c_n{}^2 n_i{}^2 c_p}\left(\frac{p_1}{n_i}+1\right)\left(\frac{p^*}{n_i}+1\right)\\ &= \tau_{no}\left(1+\frac{p_1}{n_i}\right)\left(1+\frac{p^*}{n_i}\right).\end{aligned} \tag{21}$$

The equality of the two τ's extrapolated to n_i is useful in seeing how τ varies over the entire range of conductivities.

In Fig. 2 the dependence of τ upon the Fermi level is illustrated for four cases. On the figure, two energy levels are introduced: E_t and E^*. These correspond to the position of the Fermi level, which leads to the densities $n_1 p_1$ and $n^* p^*$, respectively. A line of unit slope on the figure corresponds to an actual slope of $1/kT$. Thus, slopes of unity will be obtained whenever one of the τ's of (17) or (19) depends upon a first power of n or p. Slopes of 2 are obtained where the quantity depends inversely on the square of n or inversely on the square of p. The curves have been shown as straight lines with corners. This is accurate except near the intersections of the straight-line segments, in which there is a rounding off, extending approximately kT to either side of the limiting energy value.

From (17) and (19), it is seen that E_t and E^* enter the equations in an entirely symmetrical way. Thus, exactly the same set of diagrams will be obtained by interchanging the two energies. In every case, the horizontal lines in the region of heavy doping correspond to τ_{no} and τ_{po} as given in (18) and (20).

III. Lifetime and Diffusion Length

The definition of τ_p used in Section II is a steady-state value. Whenever the hole density in an n-type region continuously exceeds its equilibrium value by δp, each unit volume acts as a sink for holes of strength $\delta p/\tau_p$. It does not necessarily follow from this that τ_p is the lifetime of photoconductivity; there may be time constants associated with holes getting into equilibrium with the traps. We discuss these in Section VIII.

However, the steady-state value τ_p is the appropriate one to use for calculating diffusion lengths. If the current density I_p due to hole flow has a divergence div I_p, then each region loses holes at a net rate per unit volume.

$$\delta p/\tau_p + (1/q)\ \text{div}\ I_p = g_p \tag{22}$$

where g_p is the rate of generation due, for example, to generation by photon production of hole-electron pairs.

As is well known, minority carriers often flow almost entirely by diffusion; this is because any net current or charge density due to minority carriers flowing by diffusion can be compensated by very small electric fields acting on the majority carriers—electric fields so small that their disturbance of the diffusion of the minority carriers is negligible.[7] For the case of holes flowing by diffusion only in a region where g_p is zero, (22) reduces to

[7] This point has proved to be difficult to understand. Treatments will be found in W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y., sec. 12.5; 1950.
——, "Transistor electronics: imperfections, unipolar and analog transistors," Proc. IRE, vol. 40, pp. 1289–1313; November, 1952.
J. A. Hoerni, "Carrier mobilities at low injection levels," Proc. IRE, vol. 46, p. 502; February, 1958.

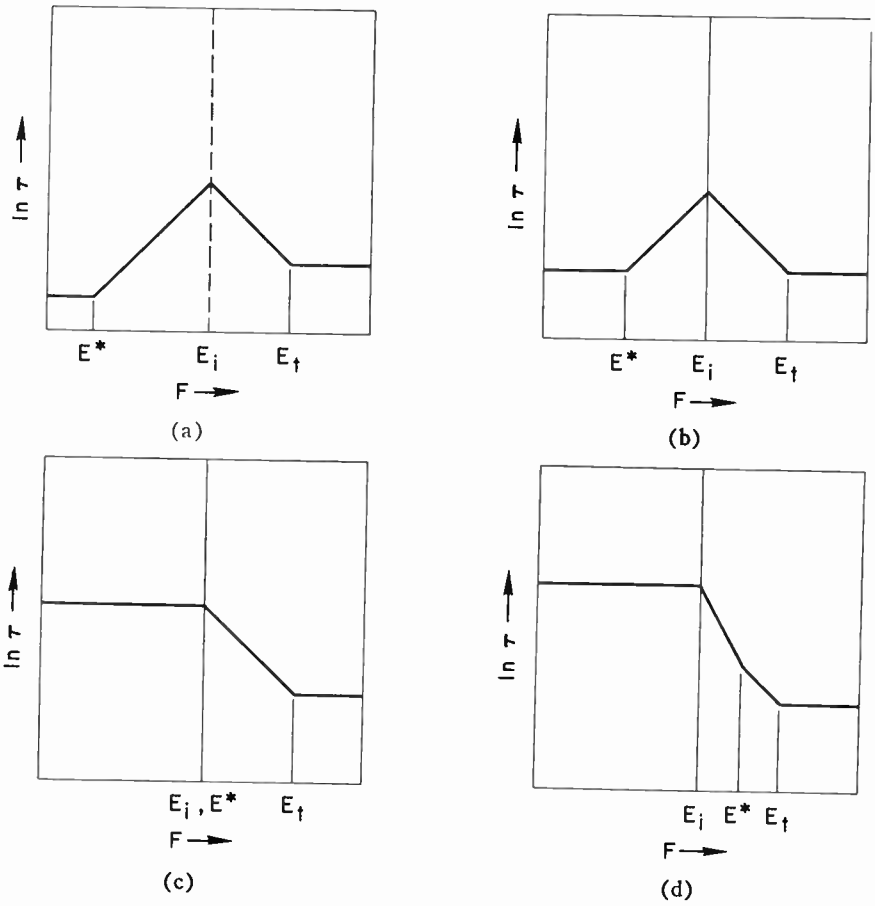


Fig. 2—Some possible forms for the dependence of lifetime upon Fermi level. (Lines of slope 1 on the figure correspond to slopes of $1/kT$.) (a) Corresponds to $c_p < c_n$; (b) to $c_p = c_n$; (c) $e_p c_p = e_n c_n$; (d) $1 < e_n/e_p < c_p/c_n$.

$$\delta p/\tau_p - D_p \,(\mathrm{div})^2\, \delta p = 0$$

for which the solutions for one-dimensional flow are of the familiar form

$$\exp\,(\pm x/L_p)$$

where the diffusion length L_p is

$$L_p = (D_p \tau_p)^{1/2}.$$

It is thus evident that the steady state τ_p is the correct one to use to calculate lifetime for use in an experiment in which the exponential decay in distance is measured. In calculating β, the transmission factor through a uniform base layer in a junction transistor, it is also evident that L_p calculated from τ_p is the appropriate diffusion length to use. (As discussed in Section IX, however, more complicated considerations apply to alpha-cutoff frequency.)

IV. The Energy Levels, Fermi Levels, Imrefs, Voltages, and Capacitances

The material of this section is largely analytical in nature and derives and collects equations needed in the subsequent interpretation. In accordance with the customary notation for semiconductors, the energy level for an electron at the bottom of a conduction band is E_c; at the top of the valence band it is E_v. We are concerned chiefly with the intrinsic level E_i which lies approximately half way between E_c and E_v. When the Fermi level lies at E_i, the electron and hole densities are equal and both have the value n_i. If the Fermi level is raised above E_i by an amount kT, the electron density increases by a factor of e and the hole density decreases by the same factor. It is thus evident that the ratio of electron density to hole density under equilibrium conditions increases by a factor of e whenever the Fermi level is raised by an amount $kT/2$.

Next we derive the expressions for the energy level of the trap E_t and the equality level E^*. This can easily be done in terms of the ratio of electrons to hole density for the case when the Fermi level falls at these two levels. When the Fermi level falls at E_t, the traps are half filled and this corresponds to the pair of densities n_1 and p_1, as discussed in connection with (8). The ratio of the two densities is

$$n_1/p_1 = e_n c_p/c_n e_p. \tag{23}$$

In accordance with the reasoning presented above, the trap level E_t lies above the intrinsic level by $kT/2$ times the natural logarithm of this ratio. Accordingly, we have

$$E_t = E_i + (kT/2) \ln e_n c_p/e_p c_n. \tag{24}$$

In (34) we give some alternative ways of writing this energy level which are related by the detailed balance requirement (7) among the four constants.

Proceeding in a similar way for the equality level, we note that from (9b)

$$n^*/p^* = e_p c_p/e_n c_n. \tag{25}$$

which leads to a value for E^* given by

$$E^* = E_i + (kT/2) \ln e_p c_p/e_n c_n. \tag{26}$$

It is instructive to compare the relationship of E_t and E^*. Subtracting (26) from (24) gives for the difference

$$E_t - E^* = kT \ln e_n/e_p. \tag{27}$$

The physical interpretation of this result is that when the two emission constants are equal, the balance requirement of equality for the rate of electron processes and the rate of hole processes is met when the traps are half filled with electrons; this is, of course, the condition that the Fermi level lies at E_t and thus shows why E_t and E^* are equal when the emission constants are equal.

It is also interesting to derive an expression for the average value of E_t and E^*. Averaging (24) and (26) gives

$$(E_t + E^*)/2 = E_i + (kT/2) \ln c_p/c_n. \tag{28}$$

It is evident that for this case, if the two capture cross sections are equal, E_t and E^* are symmetrically placed about the intrinsic level E_i.

The expressions for E_t and E^* can also be derived from the expressions for hole and electron densities in terms of Fermi levels. We carry out this derivation below and also introduce the expressions for n, p, and f in terms of their energy levels and imrefs. The purpose of presenting this second derivation is chiefly to obtain equations useful in subsequent discussions. The imref is in fact defined by these equations and is the value required for the Fermi level in order to give the prescribed density or Fermi factor. Introducing F_n, F_p, and F_t as the imrefs for holes, electrons, and traps, respectively, the relevant equations are as follows:

$$n = n_i \exp (F_n - E_i)/kT \tag{29}$$

$$p = n_i \exp (E_t - F_p)/kT \tag{30}$$

$$f = [1 + \exp (E_t - F_t)/kT]^{-1} \tag{31}$$

$$f/f_p = \exp (F_t - E_t)/kT. \tag{32}$$

The last equation is readily obtained from (31) and the definition that f_p is equal to $1-f$.

We now derive some alternative expressions for E_t in terms of (29) and (32). From (9) and (32), under equilibrium conditions in which all of the imrefs may be replaced by F, we have the following relationship

$$\begin{aligned} f/f_p &= \exp (F - E_t)/kT = n/n_1 \\ &= (n_i c_n/e_n) \exp (F - E_i)/kT. \end{aligned} \tag{33}$$

This may be readily solved for E_t.

$$\begin{aligned} E_t &= E_i + kT \ln e_n/c_n n_i \\ &= E_i + kT \ln c_p n_i/e_p. \end{aligned} \tag{34}$$

The first line of (34) comes directly from (33), and the second line is obtained by proceeding from the ratio p_1/p and (30). The equivalence between (34) and (24) readily follows from the detailed balance relation (7) which leads to the following equalities:

$$(e_n/c_n n_i)^2 = e_n c_p/e_p c_n = (c_p n_i/e_p)^2 \tag{35}$$

and these equalities lead readily to the equivalence of (34) and (24).

A similar procedure may be employed with the starred densities and leads to

$$\begin{aligned} E^* &= E_i + kT \ln e_p/c_n n_i \\ &= E_i + kT \ln c_p n_i/e_n. \end{aligned} \tag{36}$$

The equivalence of (36) to (26) also follows from the detailed balance condition on the constants.

It is convenient to introduce a set of voltages. One of these is defined as the electrostatic potential in the semiconductor and is given by

$$E_i = qV_i, \qquad V_i = -E_i/q \tag{37}$$

where $-q$ is the charge on the electron. We may interpret this equation as follows. Suppose the zero of voltage is represented by a copper bar sunk in the earth. Let us imagine that we remove an electron from this copper bar and place it in the semiconductor. Suppose that the electron was initially in an energy state at the Fermi level in the copper bar and when placing it in the semiconductor we place it at an energy level equal to E_i. (For this purpose, we may have to imagine that there is an impurity or trap in the semiconductor which happens to have its energy level at E_i so that there will be a place into which we can put this electron; this restriction is not really necessary, but the argument to show that it is not would be out of place here.) By definition, the energy required to make this move is E_i; that is, we define E_i in respect to a zero, which is a copper bar sunk into the earth. If the semiconductor consists of a small sphere supported by an insulator, then the value of E_i will depend upon the net charge on the sphere and will vary in accordance with the conventional formula relating voltage to capacity to ground. Voltage difference is energy difference divided by the quantity of charge moved, thus leading to (37).

We may similarly define three quasi-voltages for electrons, holes, and traps:

$$F_n = -qV_n, \quad F_p = -qV_p, \quad F_t = -qV_t. \tag{38}$$

Under conditions of thermal equilibrium, all of the quasivoltages except V_i assume the same value V, given by

$$F = -qV. \tag{38a}$$

This is the voltage which would be read on a voltmeter connected between the semiconductor and earthed copper post, provided the impedance of the voltmeter were high enough so that it did not disturb the situation. We also introduce an effective thermal voltage V_θ

$$kT \equiv qV_\theta. \tag{39}$$

The energy levels E_t and E^* may also be expressed in terms of voltage:

$$E_t - E_i = -qU_t, \quad E^* - E_t = -qU^*. \tag{40}$$

The values of these quantities are readily found to be:

$$U_t = (V_\theta/2) \ln e_p c_n / e_n c_p \tag{41}$$

$$U^* = (V_\theta/2) \ln e_n c_n / e_p c_p. \tag{42}$$

In terms of the quantities just introduced, we may rewrite the equations for n, p, and f as follows:

$$n = n_i \exp (V_i - V_n)/V_\theta \tag{43}$$

$$p = n_i \exp (V_p - V_i)/V_\theta \tag{44}$$

$$f = [1 + \exp -(V_i + U_t - V_t)/V_\theta]^{-1} \tag{45}$$

$$f/f_p = \exp (V_i + U_t - V_t)/V_\theta. \tag{46}$$

The signs in these equations are easy to remember. Take, for example, the equation for n. It states that if the electrons are in equilibrium with the distribution at voltage V_n, then the electron density in the semiconductor will be intrinsic if the electrostatic potential V_i of the semiconductor is equal to V_n. On the other hand, if V_i is more positive than V_n, the semiconductor is relatively more attractive, and the electron density is raised by a factor of e above the intrinsic value each time the voltage V_i increases by thermal voltage above V_n. Eq. (44) has a similar interpretation except that the positive charge of holes tends to make a more positive charge on the semiconductor repel them. In (45) U_t is a measure of how much more attractive one of the traps is than the intrinsic level. Consequently, if U_t is positive, the traps may be half filled, even though the electrons in the traps are in equilibrium with a voltage V_t which is more positive than V_i.

One of the principal reasons for introducing (43) to (46) is to enable us to use the variations in quasi-voltages in an equivalent circuit for time-constant effects involving traps, in a subsequent section. For this purpose we note that if all of the quasi-voltages deviate by small amounts, denoted by small letters, from their equilibrium values, the following changes in densities and in f arise:

$$\delta n = n(v_i - v_n)/V_\theta \tag{47}$$

$$\delta p = p(v_p - v_i)/V_\theta \tag{48}$$

$$\delta(\ln f/f_p) = \delta f/f f_p = (v_i - v_t)V_\theta \tag{49}$$

$$\delta f = f f_p (v_i - v_t)/V_\theta. \tag{50}$$

Eqs. (47)–(49) are readily obtained by differentiating the logarithms of (43), (44), and (46), and (50) follows directly from (49).

It is convenient to express the coefficients in these equations in terms of pseudo capacities. For this purpose, we note that the change in potentials results in a net change in charge per unit volume given by

$$-q\delta n + p\delta p - qN_t\delta f = C_n(v_n - v_i) + C_p(v_p - v_i) + C_t(v_t - v_i) \tag{51}$$

where the coefficients C_n, C_p, and C_t are given by

$$C_n \equiv qn/V_\theta, \quad C_p \equiv qp/V_\theta, \quad C_t \equiv qN_t f f_p/V_\theta. \tag{52}$$

These quantities have the dimensions of capacitance per unit volume or farads per cubic centimeter or per cubic meter, depending upon the system of units used.

The form of the capacitance C_t is of interest. The total possible change in charge on traps is one electronic charge per trap, or qN_t per unit volume. This change occurs most rapidly just as the Fermi level passes the trap level. When the Fermi level lies on the trap level, both f and f_p are equal to one half. Eq. (52) shows that the maximum value obtained for this capacitance is as if 25 per cent of the total possible change were produced by a voltage swing of thermal voltage just as the Fermi level passes through the trap level. If the Fermi level lies a substantial distance on either side of the trap level, then either f or f_p will be very small, and the capacitance will accordingly be small. The integral of C_t integrated over voltage is just equal to qN_t as may readily be derived using relationships like those of (49).

Except in space-charge regions such as occur at p-n junctions or near the surfaces of semiconductors, charge neutrality must be preserved. This restricts variation in the quasi-voltages to such values that (51) is equal to zero in the body of the semiconductor.

The theory developed for traps may also be applied to surface traps. In this case, C_t may be defined as a surface density with N_t being, in this case, the number of traps per unit area of the surface. In this form, C_t may be conveniently used in deriving equivalent circuits for surface effects involving traps.

V. Small Signal Theory and Equivalent Circuit

In this section we consider small disturbances from the equilibrium situation and derive an equivalent circuit which can be applied to both steady-state and transient situations. We make use of the quasi-voltages and capacitances discussed in the previous section and relate these to the electronic processes involving the traps.

We deal with the problem on a unit-volume basis and derive equations for the rate of change of electron density, hole density, and average state of charge of the traps. The equation for the rate of change of electron density is

$$dn/dt = e_n f - c_n n f_p + g_n, \tag{53}$$

where the first two terms represent the normal processes involving electron emission from full traps and electron capture into empty traps. The term g_n represents additional electron generation not produced by the normal processes involving the traps. It may, for example, represent generation of electrons by electron-hole pair formation by light absorption. In this case, an equal generation term occurs for holes. The term g_n may also represent a divergence of electron flow as discussed in Section III. Still another possibility is that it may represent electrons entering the semiconductor by bombardment of the semiconductor with electrons from a Van de

Graaff generator. We need not specify the origin of g_n at this point. The reader may find it convenient to imagine that it is produced by photon absorption when trying to give physical interpretation to the equations.

The corresponding equation for rate of change of hole density is

$$dp/dt = e_p f_p - c_p p f + g_p. \tag{54}$$

The equation for rate of change of fraction of traps that are full is given by

$$N_t df/dt = (c_n n + e_p) f_p - (c_p p + e_n) f + g_t. \tag{55}$$

The first term after the equal sign represents the rate at which traps change from the empty to the full state, due to electron capture and hole emission, and the negative term represents changes in the reverse direction. If these two terms alone were present, it is evident that the excess of the first over the second represents the net rate of increase of full traps, and this is, by definition, equal to the trap density N_t times the rate of increase of the fraction f of these which are full. The term g_t represents changes due to other causes such as, for example, the absorption of photons with just sufficient energy to excite holes from the traps; for such a process g_t would have a positive value. The energy of the photons in such a case will be approximately $E_t - E_v$. It is evident that if such a process occurs, it will make equal contributions to the g_t term in (55) and the g_p term in (54). A negative contribution to g_t and an equal positive contribution to g_n would arise from optical absorption which excited electrons from the traps into the conduction band.

Eqs. (53) to (55) are valid for arbitrary values of n, p, and f, subject only to the conditions discussed in Sections I and II that the processes are such that constant coefficients may be used to describe the elementary capture and emission processes. In the remainder of this section, however, we restrict our considerations to small disturbances from equilibrium.

We denote thermal equilibrium values of the quantities concerned by symbols n, p, and f, and the disturbances by δn, δp, and δf. Under conditions of thermal equilibrium, detailed balance requires—as discussed previously—that electron capture and electron emission exactly balance. Let the rate at which each of these processes proceeds per unit volume under conditions of thermal equilibrium be denoted by X_n. In terms of X_n, we introduce R_n, an effective resistance on a unit volume basis (R_n has dimensions of ohm cm^3), and a set of corresponding quantities for holes, as follows:

$$e_n f = c_n n f_p \equiv X_n \equiv V_\theta / q R_n \tag{56}$$

$$e_p f_p = c_p p f \equiv X_p \equiv V_\theta / q R_p. \tag{57}$$

The behavior of the two resistances as a function of n or p is very simple. For $n > n_1$ the traps are filled, X_n has its maximum value e_n, and R_n has its minimum value $V_\theta / q e_n$. For $n < n_1$, the value of R_n is increased by a factor of $n_1/n = p/p_1$. Similarly, R_p is constant for $p > p_1$ and increases linearly with n for $n > n_1$.

The rate of change of electron density, given by (53) can be expressed in simplified form in terms of X_n by considering the fractional disturbances in the variable quantities of (53). This leads to

$$\begin{aligned} d\delta n/dt &= X_n \delta f/f + X_n \delta f/f_p - X_n \delta n/n + \delta g_n \\ &= X_n \delta f/f f_p - X_n \delta n/n + \delta g_n \\ &= X_n (v_i = v_t)/V_\theta - X_n (v_i - v_n)/V_\theta + \delta g_n. \end{aligned} \tag{58}$$

In this equation we have represented the disturbance due to additional electron generation as δg_n; it is, of course, understood that g_n itself is zero, corresponding to thermal equilibrium for the undistrubed condition. The expressions in terms of the δv quantities follow from (47)–(50). It should be noted in (58) that the relationship between f and f_p and the equality of the emission and capture terms leads to the coefficient of δf having the product ff_p in the denominator; this permits its expression readily in terms of a difference in the pseudo voltages.

If (58) is multiplied by $-q$, then the left side becomes the rate of change of charge per unit volume due to changing electron density. As discussed in Section IV, this can be expressed in terms of the electron capacity C_n. Eq. (58) can then be rewritten in the form

$$C_n d(v_n - v_i)/dt = (v_t - v_n)/R_n + i_n \tag{59}$$

in which we have introduced an effective electron generation current

$$i_n = - q \delta g_n \tag{60}$$

which represents algebraically the accumulation of positive charge per unit volume due to electron generation.

Proceeding similarly with (54) and (55) we obtain

$$C_p d(v_p - v_i)/dt = (v_t - v_p)/R_p + i_p \tag{61}$$

$$C_t d(v_t - v_i)/dt = (v_n - v_t)/R_n + (v_p - v_t)/R_p + i_t. \tag{62}$$

For convenient reference, we repeat the expressions for the capacitances derived in Section IV.

$$C_t = q N_t f f_p / V_\theta, \quad C_p = q p / V_\theta, \quad C_n = q n / V_\theta. \tag{63}$$

The expressions for the currents per unit volume are

$$i_t = - q \delta g_t, \quad i_p = + q \delta g_p, \quad i_n = - q \delta g_n. \tag{64}$$

The equivalent circuit corresponding to these equations is shown in Fig. 3.

The Case of Small Trap Density

We investigate the consequences of this equivalent circuit in more generality in subsequent sections. In order to illustrate its significance at this point, let us deal with the simplified case in which the trap density is made very small. By very small, we mean that the capacitance C_t is much smaller than either C_n or C_p. Furthermore, we assume that we are dealing with an n-type semiconductor so that C_n is very much larger than C_p. (The results for a p-type situation can be obtained by exchanging n and p.)

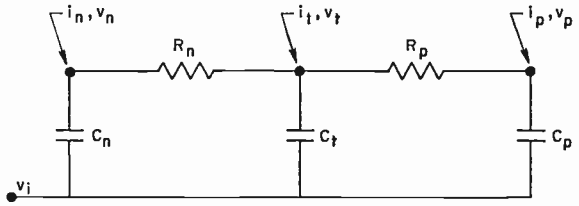


Fig. 3—The equivalent circuit for small disturbances from equilibrium for holes, electrons, and traps.

In order to illustrate the significance of small C_t, let us suppose that the system is disturbed by a flash of light. If the energy of the photons is greater than the energy gap, the net result will be to generate hole-electron pairs, and this is equivalent to adding a positive charge to C_p and an equal negative charge to C_n. The voltage developed across C_n will thus be negligible compared to the voltage v_p developed across C_p.

This disturbance will set up two transients corresponding to the two normal modes of relaxation for the circuit. If the trap density is very small so that C_t is relatively a very small capacitance, the shortest relaxation time will involve charging C_t to the potential corresponding to the ratio of division of the voltage v_p between R_n and R_p. Subsequent to this transient, C_p and C_n in series will discharge through the two R's also in series. It is this latter process which represents the normal relaxation.

This relaxation evidently satisfies

$$(C_p^{-1} + C_n^{-1})d(v_p - v_n)/dt = (v_n - v_p)/(R_p + R_n) + i_p \quad (65)$$

where the first coefficient is the series capacity of the two capacitors. This leads to a relaxation time for the circuit of

$$\tau \text{ (relaxation)} = C_pC_n(R_p + R_n)/(C_p + C_n). \quad (66)$$

This expression is valid for n type or p type and is the same for both holes and electrons. It is limited to small trap densities and small signals. The condition of electrical neutrality requires that δn and δp be equal and thus that

$$C_pv_p + C_nv_n = 0. \quad (67)$$

From this it follows that for a steady-state condition the current δi_p produces the voltage drop of $(R_p+R_n)i_p$ across the resistors and a charge of $q\delta p$ equal to this voltage times the capacitance of the two capacitors in series. Thus the charges on the two capacitors are equal and opposite and

$$q\delta p = C_pC_n(R_p + R_n)i_p/(C_p + C_n). \quad (68)$$

The steady-state lifetime, defined in Section II as the time required for the hole generation i_p/q to produce the accumulation δp of holes, is thus

$$\tau \text{ (steady state)} = \delta p/(i_p/q) = C_pC_n(R_p + R_n)/(C_p + C_n) \equiv \tau_0. \quad (69)$$

Therefore, the two definitions of lifetime lead to the same expression. We denote this lifetime by τ_0 corresponding to the situation $N_t \rightarrow 0$.

The reason for the simplification in (69) is that for $N_t \rightarrow 0$, we are dealing with a situation with one degree of freedom. For values of N_t so large that C_t cannot be neglected, there are a number of different possible values for τ depending on the definition. These we consider in later sections.

For nonintrinsic specimens, the smaller of the C's dominates the series capacity. For an n type specimen

$$C_n/C_p = n/p = n^2/n_i^2 \gg 1 \quad (70)$$

and the lifetime τ_0 can be written as

$$\tau_p \doteq C_p(R_p + R_n) \quad (71)$$

where the subscript p implies holes in n type.

Relationships of Equivalent Circuit to Effectiveness Term of Section II

We next show that (71) reduces to the form derived in Section II, and then show how the effectiveness term discussed in Section II actually arises. In terms of the definitions of the resistance terms given in (56) and (57), the expressions for τ_p may be reduced to the form of Section II as follows:

$$\tau_p = C_pR_p\left(\frac{R_p + R_n}{R_p}\right) = C_pR_p\frac{X_n + X_p}{X_n} = \frac{qp}{V_\theta}\frac{V_\theta}{qc_pnf}\left[1 + (e_pf_p/c_nnf_p)\right] = \tau_{po}\left(1 + \frac{n_1}{n}\right)\left(1 + \frac{n^*}{n}\right) \quad (72)$$

where τ_{po} replaces C_p^{-1} in the last line in keeping with (18). In (69) it is seen that the product of the first two factors in the last line represents the relaxation time or lifetime, C_pR_p, in the equivalent circuit of Fig. 3. Thus, they correspond to the situation in which R_n is zero. In the discussion of Section II this corresponds to the recombination process neglecting the disturbance of f due to δp. In this Section, neglecting δf is equivalent to setting the voltage v_t equal to zero.

The last factor $1+(n^*/n)$ of (72) is seen to be the ratio of the total resistance R_n+R_p to the resistance R_p. This factor is the reciprocal of the effectiveness of capture factor given in (12).

We are now in a position to see why the effectiveness of capture factor (12) was correctly derived by considering only hole emission and electron capture terms in its derivation. We show that the argument given in Section II is actually a superposition operation when viewed in the light of the equivalent circuit of Fig. 3. The process of considering the hole capture rate to be $c_pf\delta p$, neglecting the change δf, is equivalent to applying a voltage $v_p = q\delta p/C_p$ while grounding v_n and v_t. We shall refer to

the resulting distribution of voltage and current as *Distribution Prime* in the tabulation given below. In this distribution the current i_p, which we denote by $I(\delta p)$, is evidently

$$\delta i_p' = I(\delta p) = v_p/R_p = q\delta p/R_pC_p. \tag{73}$$

The subsequent considerations of Section II regarding the effectiveness of capture factor correspond to a *Distribution Double Prime* in which v_n and v_p are grounded and the current $I(\delta p)$ is introduced at the v_t point. This current divides with a fraction $X_n/(X_n+X_p)$ flowing to C_n and the remainder flowing back to C_p and reducing the net rate of hole capture.

Superimposing these two current voltage distributions (which is possible because we are dealing with a small signal theory) yields the situation represented in (69). In the following tabulation, some quantities not needed for the argument are omitted.

Distribution Prime:

Impose:

$$i_p' = I(\delta p), \qquad v_n' = v_t' = 0$$

Consequence:

$$i_t' = -I(\delta p), \qquad i_n = 0, \qquad v_p' = I(\delta p)/R_p.$$

Distribution Double Prime:

Impose:

$$i_t'' = I(\delta p), \qquad v_p'' = 0 = v_I$$

Consequence:

$$\begin{aligned} i_p'' &= -I(\delta p)R_n/(R_n + R_p) \\ &= -I(\delta p)R_p^{-1}/(R_p^{-1} + R_n^{-1}) \\ &= -I(\delta p)X_p/(X_p + X_p), \end{aligned}$$

similarly

$$\begin{aligned} i_n'' &= -I(\delta p)X_n/(X_n + X_p) \\ v_t'' &= \text{(not needed for argument).} \end{aligned}$$

Superimposed Distribution, Triple Prime:

$$\begin{aligned} v_p''' &= v_p', \qquad v_n''' = 0, \qquad v_t''' = ? \\ i_n''' &= -I(\delta p)X_n/(X_n + X_p) \\ i_p''' &= i_p' + i_p'' = I(\delta p)X_n/(X_n + X_p). \end{aligned}$$

The net rate of recombination is thus

$$\begin{aligned} S_L &= i_p'''/q - q^{-1}I(\delta p)X_n/(X_n + X_p) \\ &= \delta p(R_pC_p)^{-1}X_n/(X_n + X_p) \end{aligned}$$

and this is readily seen to be equivalent to (72) as it must, since the superimposed distribution is equivalent to the current flow that leads to (69).

Thus the $X_n/(X_n+X_p)$ factor, which arises from *Double Prime*, is seen to represent the fraction of current flow produced by disturbing v_t which proceeds to C_n. This fraction is physically equivalent to the fraction of trapped holes producing a disturbance δf which are recombined by electron capture.

The reader may well ask: What is the point of establishing this same result over and over again? The reason is that a rather subtle but significant point is involved. In Section II when we calculated the effectiveness of capture factor, we considered only c_nn and e_p. We neglected the effect of δf upon e_nf and c_ppf. Actually, these terms are also important. However, the *ratio* of electron processes to hole processes due to disturbance δf is

$$\frac{X_n}{X_p} = \frac{c_nn}{e_p} = \frac{e_nf}{c_ppf} \tag{74}$$

so that the neglected terms are in the same ratio as the considered terms. This equality is a consequence of detailed balance as given in (56) and (57).

There are four ways of writing X_n/X_p using terms in (56) and (57):

$$\frac{X_n}{X_p} = \frac{n}{n^*} = \frac{p^*}{p} = \frac{e_nf}{e_pf_p} = \frac{c_nnf_p}{c_ppf} = \frac{R_p}{R_n}. \tag{75}$$

The form n/n^* permits writing (72) for τ_p in terms of n only. The other forms are useful for other purposes.

VI. The Complete Small Signal Equations

As a starting point for several possible extensions of the theory, let us write down for a homogeneous body the equations governing the electron density, the hole density, the state of charge of the traps, and the electrostatic potential. The equations apply for small signals so that the capacitors and resistances of Section V are constant. Also, the electric fields and frequencies are low enough so that the mobilities and diffusion constants are constant.

On the basis of these assumptions, the equations governing small signal disturbances are

$$C_pd(v_p - v_i)/dt = (v_t - v_p)/R_p - \nabla\cdot\sigma_p\nabla v_p + i_p \tag{76}$$

$$C_nd(v_n - v_i)/dt = (v_t - v_n)/R_n\nabla\cdot\sigma_n\nabla v_n + i_n \tag{77}$$

$$C_td(v_t - v_i)/dt = (v_p - v_t)/R_p + (v_n - v_t)/R_n + i_t \tag{78}$$

$$\begin{aligned} \kappa\rho\epsilon_0\nabla^2 v_i = {} & (C_p + C_n + C_t)v_i - C_pv_p \\ & - C_nv_n - C_tv_t. \end{aligned} \tag{79}$$

The new term in (76) is the divergence of the hole current produced by diffusion and drift. This current is $-\sigma_p\nabla v_p$ and vanishes when the imref for holes is constant. The σ quantities are the conductivities due to holes and electrons

$$\sigma_p = q\mu_pp \quad \text{and} \quad \sigma_n = q\mu_nn. \tag{80}$$

Derivations of the current expressions are given in various references.[8–10] In brief, we have

[8] W. Shockley, "Electrons and Holes in Semiconductors," *op. cit.*, p. 302.

[9] W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

[10] R. D. Middlebrook, "An Introduction to Junction Transistor Theory," John Wiley and Sons, Inc., New York, N Y.; 1957.

current density due to holes

$$= q(-D\nabla_p + \mu_p p E) = -q\mu_p p \nabla v_p \qquad (81)$$

the last expression following from the definition of p in terms of n_i, V_i, and V_p, and the electric field E in terms of V_i as follows:

$$D_p \nabla_p = (D_p p/V_\theta)\nabla(V_p - V_i) = \mu_p p \nabla(V_p - V_i) \qquad (82)$$

$$-\mu_p p E = -\mu_p p \nabla V_i. \qquad (83)$$

The terms i_p, i_n, and i_t have the definitions of Section IV and are currents per unit volume generated by light absorption, bombardment, and the like—in fact, everything except particle flow and normal processes involving the traps.

Eq. (79) is simply Poisson's equation.

Eqs. (76)–(79) constitute a set of four partial differential equations in the four unknowns: v_p, v_n, v_t, and v_i. In principle, they can be solved subject to certain boundary conditions. In the next two sections we deal with homogeneous cases in which none of the disturbances depend on position. For these cases the charge density must vanish and the terms involving σ_n and σ_p vanish.

In Fig. 4 we represent the distributed line corresponding to (76)–(79). We do not attempt to discuss them fully. The following remarks may be helpful to the interested reader.

Four cases are shown:

a) The situation represented corresponds to a rod of unit area. If the length of each unit is dx, then the top capacitor and the horizontal resistors are

$$\kappa\rho\epsilon_0/dx, \qquad dx/\sigma_p, \qquad dx/\sigma_n$$

and the recombination resistors and capacitors are

$$R_p/dx, \qquad R_n/dx, \qquad C_p dx, \qquad C_t dx, \qquad C_n dx.$$

b) The situation in an n type specimen, σ_n, is assumed very large, and so is C_n. This represents the situation treated in Section IX.

c) This represents a dc situation in which no currents flow. The model requires that there be no net charge on the condensers surrounding a node. A disturbance in v_i is attenuated down the line of capacitors. The attenuation length is the Debye length.

d) This represents the situation treated in Sections VII and VIII. The disturbance does not depend on position, and the sum of the three currents vanishes.

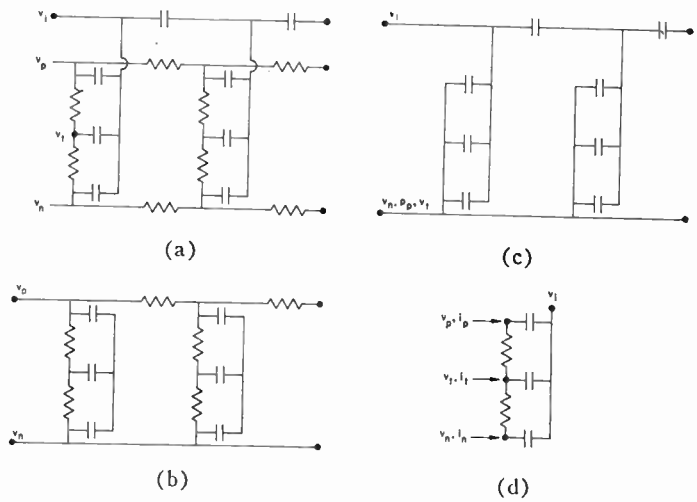


Fig. 4—Some equivalent circuits representing a semiconductor as a distributed line. (a) The complete small signal circuit. (b) The circuit suitable for minority carrier flow in n-type material. (c) The thermal equilibrium case which leads to the Debye length. (d) The homogeneous case.

VII. Small Signal, Large Trap Density, Steady State

In Section V, the equivalent circuit was treated for the case of small trap density so that the charge on C_t was negligible compared to that on C_p, the capacitor for minority carriers. In that case the system is essentially one with a single degree of freedom, namely the relaxation of C_p through R_p and R_n; consequently, the transient and steady-state solutions lead to the same lifetime. For large trap density, there are two relaxation times, which may be comparable; in general, neither of these is equal to the steady-state lifetime. In this section we derive the steady-state lifetime and compare it with the transient lifetime in the next section.

Since we are dealing with a homogeneous case, v_i is independent of position, and the charge density is zero. The actual value of v_i is unimportant, since it simply represents potential of the specimen in respect to ground. Accordingly, we eliminate v_i by introducing u_p, u_n, and u_t, defined as follows:

$$u_p = v_p - v_i, \quad u_n = v_n - v_i, \quad u_t = v_t - v_i. \qquad (84)$$

The differential equations (76)–(78) then become

$$C_p du_p/dt = (u_t - u_p)/R_p + i_p \qquad (85)$$

$$C_n du_n/dt = (u_t - u_n)/R_n + i_n \qquad (86)$$

$$C_t du_t/dt = (u_p - u_t)/R_p + (u_n - u_t)/R_n + i_t, \qquad (87)$$

and the condition of neutrality is

$$C_p u_p + C_n u_n + C_t u_t = 0. \qquad (88)$$

We now apply these to the case of steady-state generation by photon generation of hole-electron pairs. This gives

$$i_p = q\delta g_p = -i_n, \qquad i_t = 0. \qquad (89)$$

Inserting these quantities in (85) and (86), we obtain

$$u_p - u_t = R_p i_p \qquad (90)$$

$$u_t - u_n = R_n i_p \qquad (91)$$

which state that the generation current produces voltage differences equal to the ohmic drops across the recombination resistances. The actual voltages must give electrical neutrality in accordance with (88). Solving (88), (90), and (91) simultaneously leads readily to

$$u_p = i_p[(R_p + R_n)C_n + R_p C_t]/(C_p + C_n + C_t) \qquad (92)$$

$$u_n = -i_p[(R_p + R_n)C_p + R_n C_t]/(C_p + C_n + C_t) \qquad (93)$$

$$u_t = i_p(R_n C_n - R_p C_p)/(C_p + C_n + C_t). \qquad (94)$$

In terms of these relationships, we may define a steady-state hole lifetime $\tau_p(\text{ss})$ as for (10)

$$\tau_p(\text{ss}) = \tau_p(\text{steady state}) = C_p u_p/i_p$$
$$= C_p(R_p + R_n)\frac{1 + R_pC_t/C_n(R_p + R_n)}{1 + (C_t/C_n) + (C_p/C_n)}. \quad (95)$$

This expression reduces to that of Section V if either 1) $C_t \ll C_n$, or 2) $R_n \ll R_p$. In case 1, the condition of Section V applies, and the presence of C_t cannot cause an appreciable voltage to appear across C_n; consequently, the full voltage drop of $(R_p+R_n)i_p$ appears across C_p. In case 2, the voltage drop across R_n is negligible and, again, the full voltage drop appears across C_p.

By writing the quantities in (95) in terms of the trap constants, it may be re-expressed as follows:

$$\tau_p \text{ (steady state)} = \tau_p(\text{ss})$$
$$\frac{\tau_{po}[n + n_1 + N_t(1 + n/n_1)^{-1}] + \tau_{no}(p + p_1)}{n + p + N_t(1 + n/n_1)^{-1}(1 + n_1/n)^{-1}} \quad (96)$$

which is the form given in the original Shockley-Read treatment, and the symbol $\tau_p(\text{ss})$ is introduced for use in subsequent equations.

Several remarks should be made about the expression for τ_p (steady state) of (95). From its definition in terms of the ratio of minority carrier density to recombination current per unit volume, it is evident that (96) is the appropriate lifetime to use in calculating minority carrier diffusion lengths. The reasoning is the same as that presented in Section III.

On the other hand, τ_p (steady state) is not adequate for describing the change in conductivity of a specimen in which a fixed and known generation of hole electron pairs is being produced. For the situation of Section V, δp and δn are equal to each other and to $\tau_p(\text{ss})$ times the generation current per unit volume. The photoconductivity is thus

$$\delta\sigma = (\mu_p + \mu_n)\tau_p(\text{ss})i_p. \quad (97)$$

Unless C_t is small compared to both C_p and C_n, however, δ_p and δ_n will not be equal. The change in conductivity can be conveniently expressed in terms of two other steady-state lifetimes, defined from (93) and (94) just as (95) is defined in terms of (92). Thus, we write

$$\tau_n(\text{ss}) = -C_n u_n/i_p \quad (98)$$
$$\tau_t(\text{ss}) = C_t u_t/i_p. \quad (99)$$

The positive charge densities per unit volume due to i_p are thus i_p times the corresponding τ's for holes and traps, and times minus $\tau_n(\text{ss})$ for electrons. The change in conductivity is thus:

$$\delta\sigma = \mu_p\tau_p(\text{ss})i_p + \mu_n\tau_n(\text{ss})i_p$$
$$= (\mu_p + \mu_n)\tau_p(\text{ss})i_p + \mu_n\tau_t(\text{ss})i_p. \quad (100)$$

The last form results from the effect of the electrical neutrality condition (88) upon the definitions of the τ's.

The effect of τ_t may be quite large in (100). For example, if the Fermi level lies below E^*, $R_n > R_p$, and if C_t is larger than C_p and less than C_n, which may well occur for strongly n type material, then $\tau_p(\text{ss})$ will be approximately C_pR_n and $\tau_t(\text{ss})$ will be C_tR_n. Thus, holes will be in effect trapped for a long time in the traps, and there will be many more mobile electrons produced than mobile holes. (These extra electrons do *not* diffuse with a diffusion length given by D_n and τ_n (steady state); see Section II and references.)

A systematic treatment of the dependence of the τ's upon trap density and Fermi level can be presented in terms of E_t, E^*, and the Fermi level. In order to understand the relationships, we rewrite the three τ's as follows:

$$\tau_p(\text{ss}) = \tau_0\frac{1 + \dfrac{C_tR_p}{C_n(R_p + R_n)}}{1 + \dfrac{C_t}{C_p + C_n}} \equiv \tau_0K_p \quad (101)$$

$$\tau_n(\text{ss}) = \tau_0\frac{1 + \dfrac{C_tR_n}{C_p(R_p + R_n)}}{1 + \dfrac{C_t}{C_p + C_n}} \equiv \tau_0K_n \quad (102)$$

$$\tau_0 = C_pC_n(R_p + R_n)/(C_p + C_n) \quad (103)$$

$$\tau_t(\text{ss}) = \tau_n(\text{ss}) - \tau_p(\text{ss}). \quad (104)$$

The expression for τ_0 is that of (69) and corresponds to neglecting the charge on C_t. The correction factors K_p and K_n are introduced for brevity.

Let us consider first the effect of K_p upon minority carrier lifetime. The following conclusion can be reached:

K_p reduces $\tau_p(\text{ss})$ below τ_0 in N-type material if the Fermi level lies between E_i and E^*.

This conclusion follows from the fact that in n-type material $C_p \ll C_n$, so that K_p may be reduced to

$$K_p = [1 + (C_t/C_n)(1 + n^*/n)^{-1}]/[1 + (C_t/C_n)]. \quad (105)$$

Evidently the numerator is significantly less than the denominator only $n < n^*$, which requires that the Fermi level lies below E^*. The reduction will be significant only if C_t/C_n is comparable to unity or larger.

The effect of C_t in K_p and K_n can be used to separate E^* from E_t experimentally. If a set of specimens is made having increasing trap concentration, then the minority carrier lifetime will vary inversely as N_t and be given by τ_0 up to certain value of N_t. Above this value the lifetime will drop more rapidly in specimens having the Fermi level and E^* both on the same side of E_i, and the effect will be most pronounced for near intrinsic specimens.

The dependencies of the τ's upon Fermi level, N_t, E_t, and E^* can be seen in general terms by considering the limiting for large N_t. For these we find

$$\tau_p(\text{ss}) = C_pR_p \quad (106)$$
$$\tau_n(\text{ss}) = C_nR_n. \quad (107)$$

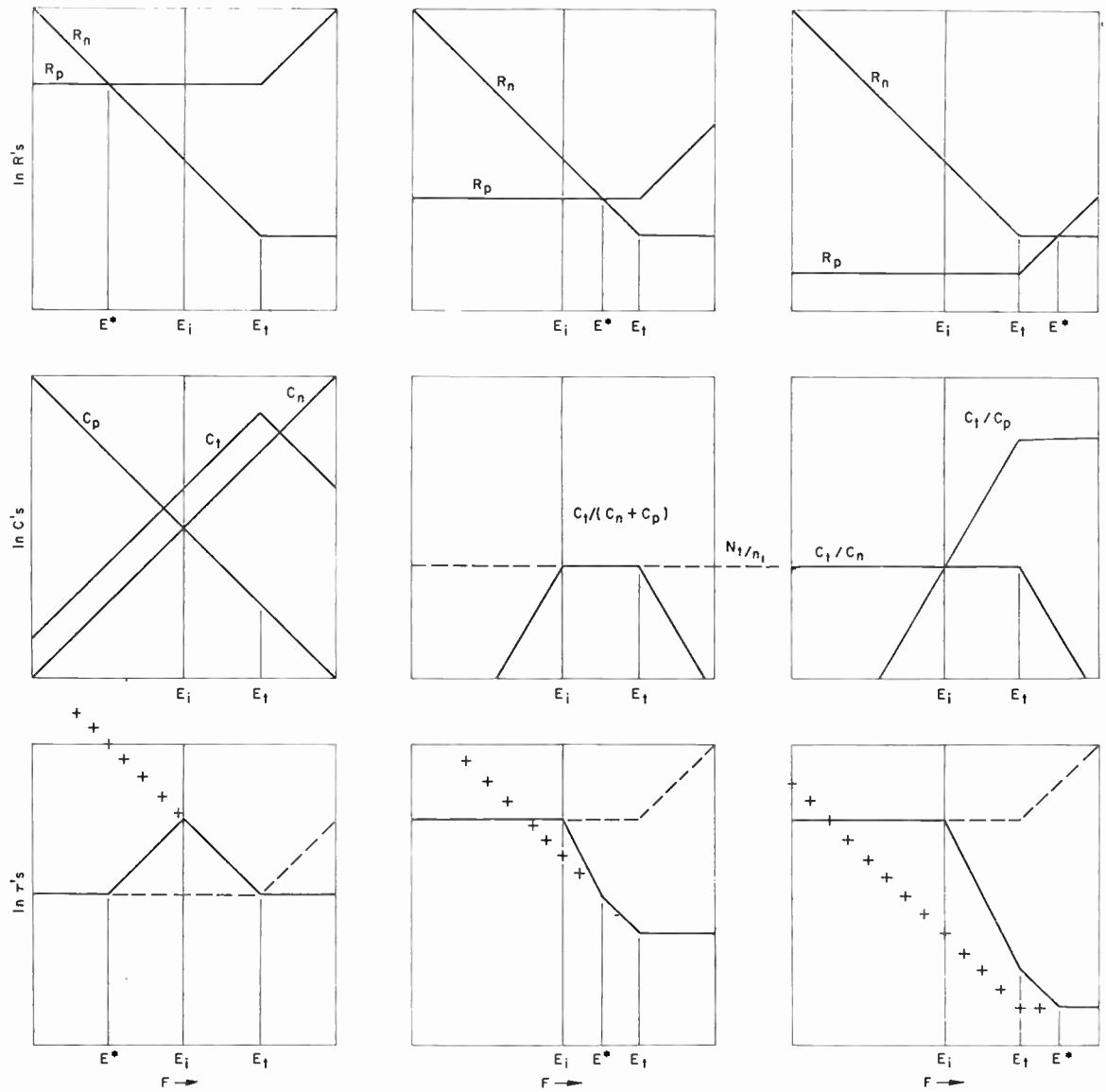


Fig. 5—The effect of the charge on the traps upon lifetime. The first row shows the dependence of the resistors of Fig. 3 upon Fermi level for three different values of E^*, all for the same value of E_t. The second row represents the capacities C_p, C_n, and C_t and certain ratios of them as functions of Fermi level, all for the same value of E_t. The bottom row corresponds to the top row and shows how the corrections due to charge on C_t modify τ_0 (solid line) by the factors K_p and K_n to give rise to τ_p(ss) (line of + signs) and τ_n(ss) (line of − signs); where the curves coincide only τ_0 is shown.

On Fig. 5 in the top row the variation of R_p and R_n is shown for three different relationships of E^* and E_t. Throughout the figure the rounding of corners over a range of about $2kT$ is not represented. The second row represents the capacitor values, and their ratios. These depend only on N_t and E_t. The expression for C_t suitable for visualizing these curves is obtained as follows:

$$C_t = qN_t f f_p/V_\sigma = qN_t/(1 + n/n_1)(1 + n_1/n)V_\sigma. \quad (108)$$

This is seen to reduce to

$$C_t = qnN_t/n_1V_\sigma = (N_t/n_1)C_n \quad (109)$$

for $n < n_1$, and to fall off as $1/n > n_1$. The bottom row shows τ_0 as the solid line and the limiting forms of τ_p(ss), and τ_n(ss) as the lines of + and − marks respectively.

It is seen that the K_p and K_n factors always reduce minority carrier lifetime and may either raise or lower majority carrier lifetime. In each case, there is one value of the Fermi level for which both K_p and K_n are equal to unity. This is the case for which no charge accumulates in the traps, and τ_t(ss) is zero. This is seen to correspond to $C_pR_p = C_nR_n$, a result consistent with (99) and (94).

The density of traps required to produce a significant deviation from the τ_0 curve depends upon the values of n_1 and n^*. For an n-type specimen [as for (105)] and Fermi level between E_i and E_t, we find

$$K_p = [1 + (N_t/n_1)(1 + n^*/n_1)^{-1}]/[1 + (N_t/n_1)]. \quad (110)$$

For this case, the effect will be largest if $n^* > n$, *i.e.*, for the Fermi level between E_i and E^*, and will show an onset for $N_t = n_1$.

A more sensitive situation corresponds to τ_n(ss) for the lower right-hand corner of Fig. 5. For the case of $n^* > n > n_1$, K_n becomes

$$K_n = [1 + (N_t/p_1)(1 + n/n^*)^{-1}]/[1 + (N_t n_1/n^2)].$$

For this case an appreciable effect occurs for $N_t \doteq p_1$. The meaning of this relationship is that, for this case, the holes get into equilibrium with the traps more easily than the electrons do (*i.e.*, $n<n^*$ so that $R_p<R_n$), and the ratio of free holes to trapped holes is approximately p_1/N_t. This situation with $N_t \gg p_1$ corresponds to very effective traps with most of the photoconductivity due to majority carriers.

VIII. Small Signal, Large Trap Density, Transient

In this section we assume that the hole and electron densities vary as functions of time of the form exp $(-\nu t)$. The results apply to the exponential decay of a disturbance.

Eqs. (84)–(87) can be rewritten in a simpler form suitable to this case expressing them in terms of the charges on the condensers. We let

$$C_p u_p = q_p, \quad C_n u_n = q_n, \quad C_t u_t = q_t \tag{111}$$

and obtain from (84) the new equation

$$\begin{aligned} C_p du_p/dt = dq_p/dt &= (u_t - u_p)/R_p + i_p \\ &= + C_t u_t/C_t R_p - C_p u_p/C_p R_p + i_p. \end{aligned} \tag{112}$$

We introduce four relaxation constants, which we express in atomic as well as circuit terms, as follows:

$$\nu_p = 1/C_p R_p = c_p f \tag{113}$$

$$\gamma_p = 1/C_t R_p = (c_p/N_t)(p + p_1) \tag{114}$$

$$\nu_n = 1/C_n R_n = c_n f_p \tag{115}$$

$$\gamma_n = 1/C_t R_n = (c_n/N_t)(n + n_1) \tag{116}$$

and use the neutrality condition

$$q_p + q_n + q_t = 0 \tag{117}$$

to eliminate q_t from (112) and the companion equation for q_n:

$$dq_p/dt = -(\nu_p + \gamma_p)q_p - \gamma_p q_n + i_p \tag{118}$$

$$dq_n/dt = -\gamma_n q_p - (\nu_n + \gamma_n)q_n + i_n. \tag{119}$$

These equations are equivalent to those published by Sandiford[6] for i_p and $i_n=0$. They can be transformed by straightforward substitutions for the relaxation constants into

$$\begin{aligned} d\delta p/dt = &-(c_p/N_t)(N_t f + p + p_1)\delta p \\ &+ (c_p/N_t)(p + p_1)\delta n \end{aligned} \tag{120}$$

$$\begin{aligned} d\delta n/dt = &-(c_n/N_t)(N_t f_p + n + n_1)\delta n \\ &+ (c_p/N_t)(n + n_1)\delta p \end{aligned} \tag{121}$$

which are identical with his form.

If we assume that q_p and q_n vary as exp $(-\nu t)$ and let i_p and i_n vanish, (118) and (119) reduce to

$$(-\nu + \nu_p + \gamma_p)q_p + \gamma_p q_n = 0 \tag{122}$$

$$\gamma_n q_p + (-\nu + \nu_n + \gamma_n)q_n = 0. \tag{123}$$

The solution of these equations corresponds to the situation following a flash of light, for example. (In the following section, the case of periodic effects will be considered. For these $i_p = -i_n$ to keep charge neutrality.)

Eqs. (118) and (119) have solutions only if the determinant of the coefficients of q_p and q_n vanish. This leads to a quadratic in ν having two real roots. For each root, (118) and (119) give the same ratio for q_p/q_n, this ratio being

$$\frac{q_p}{q_n} = \frac{\gamma_p}{\nu - \nu_p - \gamma_p} = \frac{\nu - \nu_n - \gamma_n}{\gamma_n}. \tag{124}$$

It is evident that the larger root for ν is larger than the larger of $\nu_p+\gamma_p$ and $\nu_n+\gamma_n$ and gives a positive ratio for q_p/q_n; it corresponds to a disturbance in which C_t is charged opposite to both C_p and C_n. For low trap density, it represents the quick relaxation of the traps. The smaller root has ν less than the smaller of $\nu_p+\gamma_p$ and $\nu_n+\gamma_n$. This case has opposite signs for q_p and q_n and for small trap density it represents the decay of photoconductivity.

It can also be concluded from (124) that if $\nu_p+\gamma_p$ is several times larger than $\nu_n+\gamma_n$, then for the larger root for ν, q_p is larger than q_n; *i.e.*, for the faster relaxation process more charge is involved on the quick relaxing side. This conclusion follows from the fact that the larger root is larger than $\nu_p+\gamma_p$ so that

$$q_p/q_n > (\nu_p + \gamma_p - \nu_n = \gamma_n)/\gamma_n \tag{125}$$

and the right side is larger than unity.

The quadratic equation for ν may be obtained by clearing (124) of fractions and is

$$\nu^2 - \nu(\nu_p + \gamma_p + \nu_n + \gamma_n) + \nu_p\nu_n + \nu_p\gamma_n + \nu_n\gamma_p = 0. \tag{126}$$

This is the standard form of quadratic equation

$$ax^2 + bx + c = 0. \tag{127}$$

In this equation if $b^2/4ac$ is large compared to unity, the numerically larger root is approximately $-b/a$, and the smaller is approximately $-c/b$. If one of the relaxation constants is much larger than the others, then b^2, because it contains this constant squared, will be much larger than $4ac$, which contains only cross terms. If $b^2/4ac$ is much greater than unity, the values of ν are

$$\nu = \nu_p + \gamma_p + \nu_n + \gamma_n \tag{128}$$

$$\nu = \frac{\nu_p\nu_n + \nu_p\gamma_n + \nu_n\gamma_p}{\nu_p + \gamma_p + \nu_n + \gamma_n}. \tag{129}$$

These are equivalent to Sandiford's solutions.[6] The general case can be written as

$$\nu = (\nu_p + \gamma_p + \nu_n + \gamma_n)/2 \pm B/2 \tag{130}$$

where

$$B = [(\nu_p - \nu_n + \gamma_p - \gamma_n)^2 + 4\gamma_p\gamma_n]^{1/2}. \tag{131}$$

Let us consider several limiting cases.

Small N_t, γ's larger than ν's

If N_t is small so that C_t is much less than either C_p or C_n, then the γ's are larger than the ν's. The 4 ac term of (127) does not contain a product of γ's and thus the approximations of (128) and (129) are valid leading to

$$\nu = \gamma_p + \gamma_n \tag{132}$$

for one root and for the other to

$$\begin{aligned}\nu &= (\nu_p\gamma_n + \nu_n\gamma_p)/(\gamma_p + \gamma_n)\\ &= (C_p + C_n)/(R_p + R_n)C_pC_n = 1/\tau_0\end{aligned} \tag{133}$$

where τ_0 is defined in (69). The solution (132) leads in (124) to

$$q_p/q_n = \gamma_p/\gamma_n = R_n/R_p. \tag{134}$$

This corresponds to the charge on C_t leaking off into C_p and C_n at rates proportional to $1/R_p$ and $1/R_n$. The other solution has ν much smaller than the γ's and leads to $q_p/q_n = -1$ corresponding to the situation discussed for (69).

Large N_t; γ's smaller than ν's

For this case we take C_t larger than either C_p or C_n. The two roots are approximately ν_p and ν_n. The faster decay corresponds to a positive ratio for q_p/q_n and most of the charge on the condenser having the relaxation time nearest to the root. One root thus corresponds to electrons relaxing into the traps and the other to holes relaxing into the traps.

There do not appear to be any very simple generalizations to make for more general cases. A disturbance which starts initially with equal and opposite values of q_p and q_n will decay at two rates obtained by superimposing exponential decays with the two roots. Depending upon the relationship between the Fermi level, E_t, E^* and the ratio N_t/n_1, the resulting situations may be quite diverse.

For example, if $\gamma_p < \nu_p$ and $\gamma_p < \gamma_n < \nu_n$, a situation corresponding to hole traps, there will be a relatively quick relaxation with decay constant of ν_p of the holes into the traps and a slower decay of about γ_n with the electrons combining with these holes in the traps. After a flash of light the quick decay will thus eliminate the holes with a fractional drop of photoconductivity of $(1+b)^{-1}$ where b is the ratio electron to hole mobility. The subsequent slower decay will eliminate the electrons.

IX. Small Signal, Periodic in Time, Diffusion Currents

In this section we derive equations applicable to minority carrier diffusion due to a disturbance of the form $\exp(i\omega t + ax)$ in minority carrier density. Such a solution can be used to calculate the contribution of diffusion to the admittance of a p-n junction or the transmission through the base layer of a transistor.

For the assumed condition of minority carrier diffusion, the electrostatic potential and imref for electrons will be substantially constant (see Section III and references quoted there). The divergence of hole diffusion current thus plays the role of the current source for holes. Thus, we may write

$$C_p\partial u_p/\partial t = (u_t - u_p)/R_p + D_p\partial^2 C_p u_p/\partial x^2 \tag{135}$$

the last term being simply the familiar $qD_p\partial^2\delta p/\partial x^2$; it can also be derived from (76) using the definitions of the imrefs.

The condition of substantially complete charge neutrality requires that the divergence of electron current be equal and opposite to the hole current so that the sum of the three charge densities

$$C_pu_p = q_p, \qquad C_nu_n = q_n, \qquad C_tu_t = q_t \tag{136}$$

remains zero or at least small compared to q_p. Proceeding as for Section VIII, we can rewrite the equations in the form

$$(i\omega + \nu_p + \gamma_p)q_p + \gamma_pq_n = i_p = D_p\partial^2 q_p/\partial x^2 \tag{137}$$

$$\gamma_nq_p + (i\omega + \nu_n + \gamma_n)q_n = -i_p = -D_p\partial^2 q_p/\partial x^2 \tag{138}$$

where it is assumed that

$$q_p \text{ and } q_n \text{ are functions of } x \text{ times } \exp(i\omega t).$$

The q_n term can be eliminated, and this leads to a diffusion equation in q_p:

$$D_p\partial^2 q_p/\partial x^2 = \nu_p(\omega)q_p \tag{139}$$

$$\begin{aligned}\nu_p(\omega) \equiv\ &[(i\omega)^2 + i\omega(\nu_p + \nu_n + \gamma_n)\\ &+ (\nu_p + \gamma_p)(\nu_n + \gamma_n) - \gamma_p\gamma_n]\\ &+ (i\omega + \nu_n + \gamma_n + \gamma_p).\end{aligned} \tag{140}$$

The attenuator factor a for a disturbance varying as

$$q_p = \text{const} \exp(i\omega t + ax) \tag{141}$$

is

$$a = [\nu_p(\omega)/D_p]^{1/2}. \tag{142}$$

This form reduces to the familiar

$$a = [(i\omega\tau_p + 1)/D_p\tau_p]^{1/2} \tag{143}$$

for the case of small trap density treated in Section V. [Let the γ's be very large in (140) and see (133).] For large trap densities, it gives effects of trapping on diffusion, as we shall show below.

The low frequency effects may be obtained by expanding $\nu_p(\omega)$ in powers of ω:

$$\nu_p(\omega) = \nu_0 + i\omega\nu_1. \tag{144}$$

The constant term is:

$$\begin{aligned}\nu_0 &= (\nu_p\nu_n + \nu_p\gamma_n + \nu_n\gamma_p)/(\nu_n + \gamma_n + \gamma_p)\\ &= (C_t + C_n + C_p)/(C_tC_pR_p + C_pC_nR_p + C_pC_nR_n)\\ &= 1/\tau_p(\text{ss})\end{aligned} \tag{145}$$

where the second line results from multiplying numerator and denominator by $(C_pC_nC_tR_nR_p)$ and $\tau_p(\text{ss})$ is given by (95).

The coefficient of $i\omega$ is found to be

$$\frac{\nu_p(\nu_p+\gamma_p+\gamma_n)+(\nu_n+\gamma_n)(\nu_n+\gamma_p+\gamma_n)}{(\nu_n+\gamma_n+\gamma_p)^2} \quad (146)$$

and the expansion of $\nu_p(\omega)$ up to two terms is valid for

$$\omega < \nu_n+\gamma_n+\gamma_p. \quad (147)$$

If ν_n can be neglected because of the large value of C_n compared to C_t, this reduces to

$$\begin{aligned}\nu_1 &= 1+\nu_p\gamma_p(\gamma_p+\gamma_n)^{-2}\\ &= 1+C_t/C_p[1+(R_p/R_n)]^2. \end{aligned} \quad (148)$$

In terms of ν_0 and ν_1 we may rewrite the diffusion equation as

$$(D_p/\nu_1)\partial^2 q_p/\partial x^2 = [i\omega+(\nu_0/\nu_1)]q_p. \quad (149)$$

This form is a diffusion equation corresponding to a diffusion constant smaller than D_p by a factor ν_1 and a lifetime larger than τ_p (steady state) by the same factor.[11] Thus, ν_1 can be thought of as a trapping factor; the fraction of time a hole is free to diffuse is $1/\nu_1$ so that the diffusion constant is reduced by that ratio, and since the total number of holes, both trapped and free, is ν_1 times the free holes, the effective lifetime is ν_1 times larger than $\tau_p(\text{ss})$.

Thus we are led to introduce still another definition of lifetime

$$\begin{aligned}\tau_p\ (\text{diffusion}) &= \nu_1/\nu_0\\ &= \tau_0\{1+C_t/C_p[1+(R_p/R_n)]^2\}, \end{aligned} \quad (150)$$

the last form being valid for ν_n negligible compared to the other ν's and γ's corresponding to C_n much larger than C_t as in strongly n type specimens.

If the holes interchange easily with the traps so that $R_n/R_p=n^*/n$ is much greater than unity, ν_1 becomes $(C_p+C_t)/C_p$ which is simply interpreted as the ratio of holes on traps and free to free holes. On the other hand, if n/n^* is much greater than unity, ν_1 becomes unity; the interpretation in this case is that the hole does not return to the mobile condition once it has been trapped, *i.e.*, the effectiveness of capture of Section II is unity.

We consider next how the traps may affect the alpha cutoff frequency of a transistor. If the base layer is thin compared to

$$L(0) = (D_p\tau_p(\text{ss}))^{1/2}$$

then the low-frequency transmission factor

$$\text{sech}\ (W/L(0)) \doteq 1-W^2/L^2(0) = 1-\epsilon \quad (151)$$

will approach unity. One might then expect on alpha cutoff frequency of the order of

$$\begin{aligned}f_\alpha^{co} &= D_p/\pi W^2 = D_p/\pi L^2(0)\epsilon\\ &= 1/\pi\tau_p(\text{ss})\epsilon. \end{aligned} \quad (152)$$

Thus an estimate of alpha cutoff frequency based on the dc diffusion length lifetime and ϵ as deduced from dc alpha may be high by a factor ν_1 because D_p/ν_1 should be used in place of D_p in the expression for alpha-cutoff frequency.

The series expansion of $\nu_p(\omega)$ is valid for small values of ω compared to $\nu_n+\gamma_n+\gamma_p$. For large values of ω the approximation is

$$\nu_p(\omega) = i\omega+\nu_p. \quad (153)$$

The interpretation of this equation is that the only effect which has time to happen is hole capture by the traps. The traps do not have time to change charge. For this approximation to hold

$$\omega > \nu_n+\gamma_n+\gamma_p. \quad (154)$$

There are evidently a variety of other situations to which the equations of this section may be applied.

X. Recombination for Steady-State, Large Disturbances

In this section we derive some results valid for large disturbances in carrier density for steady-state conditions. Although limitations of space and time prevent a full discussion, the method of approach outlined here should be helpful. Letting

$$g_p = g_n = g \quad (155)$$

represent the total rate of recombination of holes and electrons through traps, we have

$$g = g_n = c_n n f_p - e_n f \quad (156)$$

$$g = g_p = c_p p f - e_p f_p. \quad (157)$$

Eliminating the g's from these equations leads to

$$\begin{aligned}f/f_p &= (c_n n+e_p)/(c_p p+e_n)\\ &= (c_n/c_p)(n+n^*)/(p+p^*). \end{aligned} \quad (158)$$

From this we see that the distribution of charge in the traps is controlled by electron and hole densities as follows:[12]

denuded:	$n<n^*$,	$p<p^*$;	$f/f_p = e_p/e_n$ (159)
p dominated:	$n<n^*$,	$p>p^*$;	$f/f_p = p_1/p$ (160)
n dominated:	$n>n^*$,	$p<p^*$;	$f/f_p = n/n_1$ (161)
flooded:	$n>n^*$,	$p>p^*$;	$f/f_p = c_n n/c_p p$. (162)

Flooded and denuded conditions cannot correspond to thermal equilibrium. On the other hand, p dominated and n dominated can and in fact they do give the same results as (6) for equilibrium.

Solving (158) for f and f_p and inserting in (156) leads to

$$g = c_n c_p(np-n_i^2)/(c_n n+c_p p+e_n+e_p). \quad (163)$$

This result shows simply why it is difficult to distinguish between E_t and E^*; interchanging values between

[11] This interpretation is similar to that given by the author in footnote 9.

[12] This classification was introduced in Sah and Shockley. See reference 4.

e_n and e_p does not affect g and does interchange E_t and E^* without upsetting the detailed balance relationship.

The functional dependence of g upon n and p and the four trap constants can be rewritten in terms of the densities n_1p_1 and n^*p^* and a pair of geometric mean densities

$$n_m = (n_1n^*)^{1/2}, \qquad p_m = (p_1p^*)^{1/2} \tag{164}$$

corresponding to a mean level

$$E_m = (E_t + E^*)/2. \tag{165}$$

The result is

$$g = \frac{(e_ne_p)^{1/2}[(np/n_i^2) - 1]}{\dfrac{n}{n_m} + \dfrac{p}{p_m} + \dfrac{n_1}{n_m} + \dfrac{p_1}{p_m}}. \tag{166}$$

The last two terms in the denominator are reciprocals of each other and also may be written as

$$\exp \pm (E_t - E^*)/2kT = \exp \pm (E_t - E_m)/kT = (e_n/e_p)^{\pm 1/2} = (n_m/n^*)^{\pm 1} \text{ etc.} \tag{167}$$

This shows that the denominator is substantially constant and equal to

$$\exp |E_t - E^*|/2kT \tag{168}$$

so long as $n <$ (the larger of n_1 or n^*) and $p <$ (the larger of p_1 and p^*). The bracket in the numerator is simply

$$[\exp (V_n - V_p)/V_\theta] - 1. \tag{169}$$

Eq. (166) shows in simple form how the flat region of surface recombination velocity discussed by various authors[13] arises in terms of E_t and E^*. It also is helpful in visualizing how the current dependence in p-n junctions arises.

Two special cases for g should be mentioned. In the denuded condition, as in space-charge region, we have

$$g = e_ne_p/(e_n + e_p) \tag{170}$$

and in the flooded condition with $n = p$ corresponding to higher injection levels

$$g = nc_nc_p/(c_n + c_p) \tag{171}$$

corresponding to a lifetime given by

$$n/g = (c_n + c_p)/c_nc_p = \tau_{po} + \tau_{no}. \tag{172}$$

[13] A. Many and D. Gerlich, "Distribution and cross sections of fast states on germanium surfaces in different gaseous ambients," *Phys. Rev.*, vol. 107, pp. 404–411; July 15, 1957. (References.)

Recombination in Semiconductors*

G. BEMSKI†

Summary—Excess carriers in semiconductors recombine either by direct recombination of electrons and holes, or through the intermediacy of recombination centers. The latter process is the one observed in silicon and germanium.

Various impurity atoms, dislocations, vacancies, and interstitials are known to act as recombination centers. The capture rates associated with these imperfections vary over a wide range depending on their state of charge.

Recombination at the surfaces is described in terms of a similar model in which surface states replace the recombination centers present in the bulk. The surface recombination velocity measures the density and capture properties of these states.

A given center can act as a recombination center or a trap depending on the relative magnitude of the capture cross sections for electrons and holes. This paper reviews the analytical treatments of the different processes as well as specific experimental results.

* Original manuscript received by the IRE, March 21, 1958; revised manuscript received, April 11, 1958.

† Bell Telephone Labs., Inc., Murray Hill, N. J.

I. Introduction

THE lifetime of free carriers in the phenomena of photoconductivity and of transport problems in semiconductors is a concept of basic importance. A very large number of papers have appeared in recent years concerning both topics. We shall limit ourselves mainly to the review of recombination processes in semiconductors, principally in germanium and silicon. In the next section we examine briefly the theoretical aspect of recombination processes in semiconductors. Section III reviews the basic methods employed in measurements of lifetime. This is followed by a discussion of experimental results in connection with specific mechanisms for carrier recombination. Finally in Section V we outline briefly some of the practical aspects concerning the lifetime as encountered in silicon and

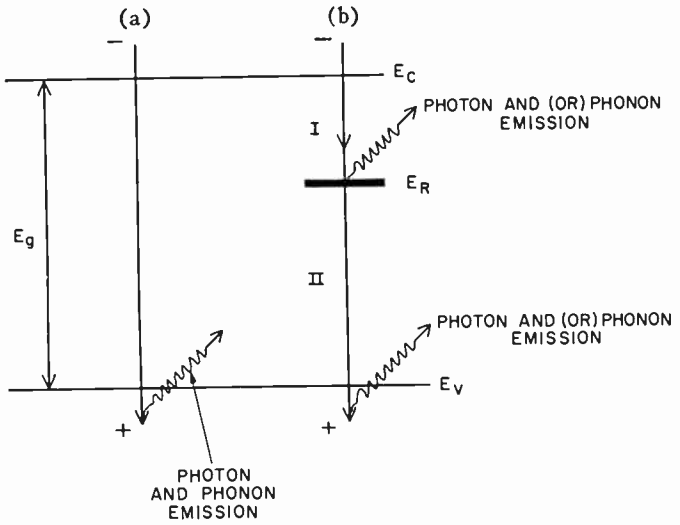


Fig. 1—Recombination in semiconductors: (a) direct, radiative recombination, (b) indirect, two-step recombination in presence of the recombination center at E_R.

germanium technology. Readers interested specifically in the role of the lifetime in photoconductivity may be referred to the excellent papers on the subject given at the Photoconductivity Conference [1] and to other review papers [2], [3].

II. Theory

A. Kinetics of Recombination

Excess carriers can be introduced into a semiconductor by means of photo-excitation (this includes X rays and gamma rays, as well as light of longer wavelength), particle irradiation (*i.e.*, high-energy electrons) or by electrical injection at the contacts to the semiconductor. In all cases these excess carriers consist of electron-hole pairs produced above those normally present at equilibrium. The excess carriers gradually recombine and the lattice is restored to its thermal equilibrium condition. The lifetimes describe the mean times spent by the excess electrons and holes in the conduction and valence bands, respectively, and are defined as the times during which the added carrier concentrations are reduced to $1/e$ of their original values.

Two principal mechanisms are believed to exist controlling the recombination processes observed in semiconductors. The first is a direct recombination of electrons and holes accompanied by photon and phonon emission [4]. This is the intrinsic, radiative recombination process. The second mechanism for recombination requires the presence of localized energy states in the forbidden gap of the semiconductor [5], [6]. The presence of such states in the crystal increases the rates of recombination of electrons and holes. Impurities, dislocations, vacancies, and interstitials are frequently known to introduce these kinds of states, with the net effect of decreasing the lifetime of excess carriers. The two possibilities are shown schematically in Fig. 1. There exist several ways in which the energy of a carrier can be dissipated in the process of capture.

TABLE I
Table of Symbols

τ_i—Lifetime of free carriers in intrinsic material (seconds).
τ_n—Lifetime of electrons (seconds).
τ_p—Lifetime of holes (seconds).
n_i—Density of intrinsic carriers (cm^{-3}).
n_0—Density of electrons in the conduction band (extrinsic) (cm^{-3}).
p_0—Density of holes in the valence band (extrinsic) (cm^{-3}).
R—Rate of radiative recombination ($cm^3\ sec^{-1}$).
v_n, v_p—Thermal velocity of excess electrons and holes, respectively ($cm\ sec^{-1}$).
σ_n, σ_p—Capture cross section for electrons and holes, respectively (cm^2).
N—Density of recombination centers (cm^{-3}).
$p_1 = N_v \exp{(E_v - E_R)/kT}(cm^{-3})$.
$n_1 = N_c \exp{(E_R - E_c)/kT}(cm^{-3})$.
$\tau_{n0} = 1/(Nv_n\sigma_n)$ (seconds).
$\tau_{p0} = 1(Nv_p\sigma_p)$ (seconds).
$\delta n, \delta p$—Density of excess carriers (cm^{-3}).
N^0—Density of empty recombination centers (cm^{-3}).
N^-—Density of recombination centers occupied by electrons (cm^{-3}).
τ_σ—Mean time spent by an electron in a trap (seconds).
U—Rate of recombination ($cm^{-3}\ sec^{-1}$).
y—Fraction of occupied traps.
s—Surface recombination velocity ($cm\ sec^{-1}$).
N_t—Number of surface states per cm^2.
c_n, c_p—Capture probabilities per second for electrons and holes, respectively $= v\sigma$ ($cm^3\ sec^{-1}$).
n_s, p_s—Free carrier densities at the surface (cm^{-3}).
n_{s1}, p_{s1}—Surface-carrier densities when $E_F = E_S$ (cm^{-3}).
E_F—Position of the Fermi level in the energy gap (electron volts).
E_S—Position of the surface states in the gap (ev).
E_i—Value of E_F in an intrinsic semiconductor (ev).
E_g—Width of the energy gap (ev).
E_c—Lower edge of the conduction band (ev).
E_v—Upper edge of the valence band (ev).
$\phi = (E_F - E_i)/q$ (volts).
I_{sc} = Short circuit current.
μ_n, μ_p—Hall mobilities for electrons and holes ($cm^2\ volts^{-1}\ sec^{-1}$).
$\theta = \theta_n + \theta_p$ sum of Hall angles for electrons and holes.
D—Diffusion constant ($cm^2\ sec^{-1}$).
ΔG—Conductance increase (ohm^{-1}).
L—Diffusion length (cm).

1) *Intrinsic Recombination:* In the case of direct recombination the energy is carried away by photons of a wavelength close, but not necessarily identical to that which corresponds to the width of the energy gap, E_g. The difference may be taken up by phonons and dissipated as heat in the lattice.

The lifetime for direct, radiative recombination in the intrinsic semiconductor, in the case of small disturbance in carrier concentration, can be expressed as:

$$\tau_i = \frac{n_i}{2R} \tag{1}$$

(See Table I for the meaning of symbols.) In p-type material it becomes:

$$\tau_p = \frac{p_0}{R} = 2\left(\frac{n_i}{n_0}\right)\tau_i \tag{2}$$

and in n-type:

$$\tau_n = \frac{n_0}{R} = 2\left(\frac{n_i}{n_0}\right)\tau_i. \tag{3}$$

The recombination rate can be expressed in terms of the capture cross section, σ, such that:

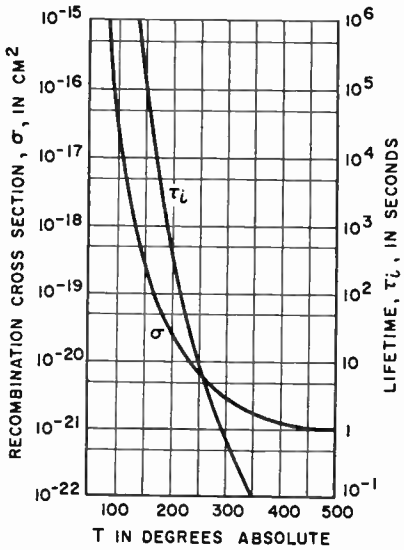


Fig. 2—Calculated recombination cross section σ and lifetime τ_i as a function of temperature for germanium [4].

$$\tau_i = \frac{1}{2n_i v\sigma} \tag{4}$$

where v is the mean thermal velocity of the carriers.

σ therefore relates to n_i and R as:

$$\sigma = \frac{R}{n_0 p_0 v} = \frac{R}{n_i^2 v}. \tag{5}$$

Fig. 2 gives a plot of σ and τ_i as a function of temperature for intrinsic germanium [4].

2) *Recombination in the Presence of Recombination Center:* If the recombination process is dominated by the presence of recombination centers, the energy of the recombining carriers can be taken up by photons or phonons in one of the following ways:

a) Radiative recombination in which photons of longer wavelengths than in the intrinsic case are emitted.

b) Auger effect mechanism in which a collision of two carriers occurs, one of them recombines at a recombination center while the other one carries away the energy lost by the first one.

c) Exciton formation in which an electron-hole pair (exciton) travels until one of the members recombines while the other one continues its travel with increased energy.

d) Phonon emission (or emission and absorption) in which net emission of phonons accompanies the act of capture. These possibilities have been discussed by several authors [7–11]. Lax [7] has arrived at the conclusion that a net multiphonon emission is the most probable one in a variety of experimental circumstances in semiconductors.

In Table II it is shown that lifetime calculated on the basis of a direct radiative process yields values considerably higher than those observed in silicon and germanium, but not far different from the experimental ones for In Sb. Also the lifetimes measured in similar samples of silicon or germanium differ frequently, indicating that the radiative process is not the limiting one. From the experimental results one can conclude that the recombination centers limit the observed recombination lifetime in silicon and germanium, but the direct recombination process becomes a competitive, and possibly a dominant one in semiconductors with narrower energy gaps [12].

TABLE II

	Si[1]	Ge[2]	In Sb [13]
E_g (ev)	1.12	0.75	0.17
R (cm^{-3} sec^{-1})	$2–10^{9}$ [3]	$3.7–10^{13(e)}$	$2–10^{21}$
τ_{rad} (seconds)	3.5	0.30	$1.5–10^{-6}$
τ_{obs} (seconds)	$<10^{-3}$	$<10^{-3}$	$1.5–10^{-7}$

Energy gap, E_g; radiative rate of recombination, R; calculated radiative lifetime, τ_{rad}; and observed lifetime, τ_{obs} for Si and Ge at 300°K and In Sb at 250°K.

[1] F. J. Morin and J. P. Maita, *Phys. Rev.*, vol. 96, p. 28; July, 1954.
[2] F. J. Morin and J. P. Maita, *Phys. Rev.*, vol. 94, p. 1525; June, 1954.
[3] R. L. Petritz, p. 63; see [1].

We shall henceforth restrict ourselves mainly to the recombination process in the presence of recombination centers. The recombination process limited by the center is a two-step one in which the electron capture (Fig. 1) by the empty center E_R is followed by a hole capture at the same center, after which it is ready for further electron capture. The kinetics of this process have been discussed in the literature [5], [6] and are reviewed by W. Shockley. It is apparent that in the direct process there is little doubt as to the interpretation of the experimental observation. The lifetime of an electron is here identical with a lifetime of a hole. One can therefore refer to it as the lifetime of a pair. This is not necessarily the case if the recombination occurs via a recombination center. It is then possible to divide the recombination process into a steady state and a transient one. Shockley and Read [5], Hall [6], and Adirovich and Guro [14] treat the steady-state case while Sandiford [15], Wertheim [16], and Clarke [17] discuss the transient case.

The net result of all these equations is that the recombination process cannot be described simply in a closed form applying to all the situations. The change in the excess carrier concentration, δp, occurs as:

$$\frac{d(\delta p)}{dt} = -\frac{\delta p}{\tau}$$

where τ is the lifetime. The lifetime for electrons and holes can be defined in terms of the net rate of capture, U, as:

$$\tau_n = \frac{\delta n}{U}$$

and

$$\tau_p = \frac{\delta p}{U}.$$

In the steady-state case the net rate of capture of electrons equals that of holes. In the most general case $\delta n \neq \delta p$, and Shockley obtains expressions for $\tau_n \neq \tau_p$ for low injection case ($\delta p \to 0$). They are:

$$\tau_p = \left[\frac{\tau_{n0}(p_0+p_1)+\tau_{p0}\left[n_0+n_1+N\left(1+\frac{n_0}{n_1}\right)^{-1}\right]}{n_0+p_0+N\left(1+\frac{n_0}{n_1}\right)^{-1}\left(1+\frac{n_1}{n_0}\right)^{-1}} \right]_{\delta n \to 0} \tag{6}$$

and

$$\tau_n = \left[\frac{\tau_{p0}(n_0+n_1)+\tau_{n0}\left[p_0+p_1+N\left(1+\frac{p_0}{p_1}\right)^{-1}\right]}{n_0+p_0+N\left(1+\frac{p_0}{p_1}\right)^{-1}\left(1+\frac{p_1}{p_0}\right)^{-1}} \right]_{\delta \to 0} \tag{7}$$

where n_1 and p_1 are defined as:

$$n_1 = N_c \exp\left(\frac{E_R - E_c}{kT}\right),$$

$$p_1 = N_v \exp\left(\frac{E_v - E_R}{kT}\right).$$

Only if N, the density of recombination centers, is low do both lifetimes reduce to a *single* lifetime:

$$\tau = \frac{\tau_{p0}(n_0 + n_1) + \tau_{n0}(p_0 + p_1)}{n_0 + p_0}, \tag{8}$$

$$\tau = \tau_0 \frac{1 + \left[\frac{\delta n(\tau_{n0} + \tau_{p0})}{\tau_{p0}(n_0 + n_1) + \tau_{n0}(p_0 + p_1)}\right]}{1 + \frac{\delta n}{(n_0 + p_0)}}, \tag{9}$$

where τ_0 is expressed by (8). This allows for a variation of the lifetime between $\tau = \tau_0$ (for $\delta n = 0$) and $\tau = \tau_\infty = \tau_{n0} + \tau_{p0}$ (for $\delta n \to \infty$).

Several authors [18–20] have considered in detail the recombination and generation of carriers occurring in the space-charge layer of a p-n junction which differs from the rest of a semiconductor in that it contains a strong electric field. This field forces the carriers out of this region very rapidly. Shockley and Read [5] have shown that the net rate of generation may be greatly increased, exceeding the rate of generation further away from the junction. As a result, the lifetime in the space charge region is smaller than elsewhere in the semiconductor. These effects are particularly important in explaining the observed deviations in the current-voltage characteristics of silicon and germanium diodes [18–20].

The steady-state lifetimes discussed heretofore are those measured, for example, in the diffusion-length type of experiments. However for the transient case, (in the photoconductive decay type of measurements), Sandiford shows that the decay of the excess carrier concentration δp is of the form [15]:

$$\delta p = Ae^{-t/\tau_1} + Be^{-t/\tau_2} \tag{10}$$

where τ_1 is the readjustment time of charges on recombination centers to the condition of equal capture rate of electrons and holes, and τ_2 is the principal lifetime term always identical for electrons and holes.

For low densities of δn and δp, τ_1 and τ_2 are given by:

$$\tau_1 = v_p\sigma_p\left[p_0 + p_1 + N\left(1 + \frac{p_0}{p_1}\right)^{-1}\right] + v_n\sigma_n\left[n_0 + n_1 + N\left(1 + \frac{n_0}{n_1}\right)^{-1}\right]. \tag{11}$$

$$\tau_2 = \frac{\tau_{n0}\left[p_0 + p_1 + N\left(1 + \frac{p_0}{p_1}\right)^{-1}\right] + \tau_{p0}\left[n_0 + n_1 + N\left(1 + \frac{n_0}{n_1}\right)^{-1}\right]}{n_0 + p_0 + N\left(1 + \frac{n_0}{n_1}\right)^{-1}\left(1 + \frac{n_1}{n_0}\right)^{-1}}. \tag{12}$$

where

$$\tau_{n0} = \frac{1}{Nv_n\sigma_n} \qquad \tau_{p0} = \frac{1}{Nv_p\sigma_p}.$$

It is obvious from above that the lifetime is a function of 1) the density of recombination levels (N), 2) the equilibrium carrier density of the semiconductor (n_0, p_0), and 3) the temperature (both n_0, p_0 and n_1, p_1 are functions of temperature, and σ's may also vary with T). The lifetime depends on the density of injected carriers in the way expressed by [5]:

It is apparent that for typical values of N, σ_n and σ_p, τ_2 is orders of magnitude higher than τ_1. It also should be noticed that 1) under transient conditions, with large density of centers N, τ_2 is not identical with the steady-state lifetime, (6) and (7); 2) for small N, (8) and (12) are identical; and 3) the lifetime of electrons and holes is always identical, under transient conditions. If the injection density, δn, satisfies the relations

$$\delta n \leq N^0 + n_0 + n_1$$

$$\delta p \leq N^- + p_0 + p_1$$

and

$$\delta N \leq \delta n + \delta p \tag{13}$$

then it has been shown by Wertheim [16] that the resulting lifetime is identical with the steady-state, large δn case [5]. If, however, one exceeds the conditions of (13) other solutions must apply.

3) *Many Level Case:* Several authors have been interested in the case of recombination rates resulting from the simultaneous presence of more than one type of recombination centers [16], [21–24]. The belief that the reciprocal of the resulting lifetime can be justifiably expressed as a sum of the reciprocal time constants due to the several separate recombination levels is true only for the case of small density of centers with the resulting equation becoming considerably more complex in a more general case [16], [22].[1]

4) *Recombination and Trapping:* If the net rate of reemission of the carriers from an energy level in the forbidden gap is more probable than the final step of recombination, then an imperfection giving rise to this level is called a trap. This phenomenon of trapping, as distinct from recombination, has also been studied in semiconductors [23], [25]. In Fig. 3, the electron from the conduction band has been trapped in the level E_t, after which it has a "choice" of return to the conduction band 2, or recombination with a hole 7. If reemission 2 occurs the electron continues traveling and may be trapped 3 several times more. Eventually it recombines via the same level E_t, or through a different level E_R 5, 6.

It is apparent that the process of trapping differs from recombination only in the relative values of the capture cross sections. In particular, in Fig. 3, the re-emission 2 of the electron will occur rather than recombination if the hole-capture cross section is very low. This could be the case for instance, if the level E_t represented a doubly ionized donor (with two positive charges). Such a level would be singly positively ionized after step 1 and would have a very small attraction for a hole (low-capture cross section for step 7). It is important to emphasize that the time spent by the carrier in the conduction band is always the lifetime (mean life) of the electron and trapping affects only the time spent in the traps which can exceed by orders of magnitude the lifetime of the carrier. This effectively reduces the mobility of the carrier and gives an apparent long decay constant.

Hornbeck and Haynes [25] have studied trapping in silicon crystals. The observed time constants of the decay of carriers, τ, in the transient case is shown to be

$$\tau = \tau_g + \tau_r \tau_g N \sigma v (1 - y) \tag{14}$$

where τ_r is the lifetime of the electrons in the conduction band, and τ_g is the mean time spent by the electron in the trap. N is the density of traps, and $(1-y)$ is the fraction of empty traps present. The time τ_g is strongly temperature dependent, decreasing exponentially with increasing temperature. It is shown that different kinds of traps exist in silicon and these are observed at room temperature in contrast to germanium where trapping is produced only below 200°K. The time τ_g spent in traps can be as long as 0.3 second (p-type silicon). Trapping in copper doped and bombarded germanium has also been studied [26]. More recently there have been indications that trapping effects are greatly reduced in silicon which contains a reduced amount of oxygen as compared to the amount which is present in the majority of the pulled silicon crystals.

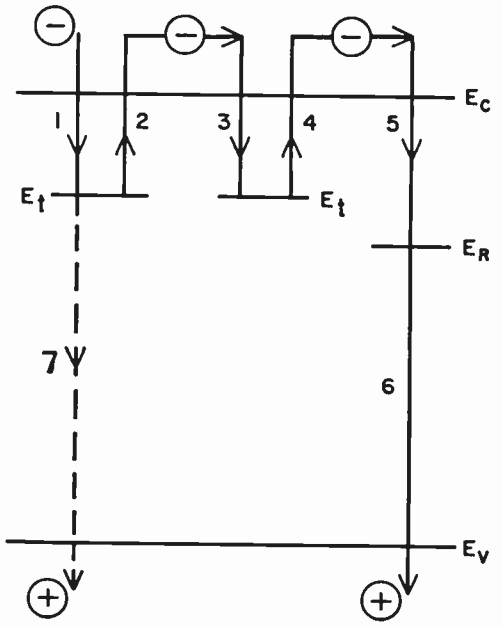


Fig. 3—Trapping effects. The electron can be trapped at the levels E_t and released into the conduction band several times before recombining at E_R. The mean time spent in the conduction band is defined as the lifetime of the electron.

5) *Surface Recombination:* It is necessary to exercise considerable care in the interpretation of experiments describing changes in the concentration of free carriers in terms of the processes of recombination in the bulk of the material.

The phenomena of surface recombination as well as of trapping should be carefully distinguished from the recombination in the volume of the material. It has been found that if the free carriers are produced close to the surface of the samples, they may diffuse towards them and consequently recombine at the surface rather than in the volume. This will be influenced by several factors such as the geometry of the sample, the degree of penetration of photons (when photons are used to produce the carriers) as well as the effectiveness with which the surface reflects the carriers which approach it from the volume.

In a filament in the presence of a field in the x direction, the current has components perpendicular to the

[1] This problem is also discussed in considerable detail by C. T. Sah and W. Shockley, "Electron-hole recombination statistics in semiconductors through flaws with many charge conditions," *Phys. Rev.*, vol. 109, pp. 1103–1115; February, 1958.

x direction of the form [27]:

$$\vec{I} = -qD\cdot[\text{grad}\ (\delta p)]_n.$$

These components represent a diffusion current. The solution of the continuity equation shows that the current toward the surfaces satisfies

$$I_y = I_z = \pm qs\delta p. \tag{15}$$

The quantity s has the dimension of velocity and is called the surface recombination velocity. The rate of recombination is as if a current of minority carriers (of density δp) were drifting with an average velocity s into the surface and being removed.

Small magnitude of s implies that a large fraction of carriers approaching the surface is reflected, while s large means that most carriers approaching are captured at the surface. Any measurement of lifetime will provide a time constant which is due to the recombination at the surfaces as well as in the bulk of the material. This time constant, τ_{obs}, is of the form [27]

$$\frac{1}{\tau_{obs}} = \frac{1}{\tau} + \frac{1}{\tau_s} \tag{16}$$

τ_s is the lifetime due to surface recombination and τ is the bulk lifetime. The two contributory constants can be separated by changing the cross sectional dimensions of the sample. The relation between τ_s and the dimension of the sample in shape of a filament with rectangular cross section is

$$\frac{1}{\tau_s} = D\left(\frac{\eta^2}{A^2} + \frac{\xi^2}{B^2}\right) \tag{17}$$

where

$$\eta \tan \eta = s\frac{A}{D}$$

$$\xi \tan \xi = s\frac{B}{D}$$

provide infinite number of roots η and ξ for a known cross sectional dimension $2A \times 2B$ of the filament and the known diffusion constant D. The smallest pair of roots, η_0 and ξ_0 correspond to the longest lifetime and are of principal importance. Knowledge of τ enables us to determine the surface-recombination velocities.

Tamm [28] and Shockley [29] have postulated the existence of surface states with energy levels in the forbidden gap of the semiconductor. Fig. 4 shows an energy level diagram of n-type germanium with surface states at the position E_s (as given by Stevenson and Keyes [30]). Conservation of charge requires the bending of the energy levels and results in a space charge of about 10^{-5} cm thickness. E_i represents the value of the Fermi level in the intrinsic case; $q\phi_B$ represents the distance between E_F and E_i in the bulk of the material; and $q\phi_s$ is a similar distance at the surface.

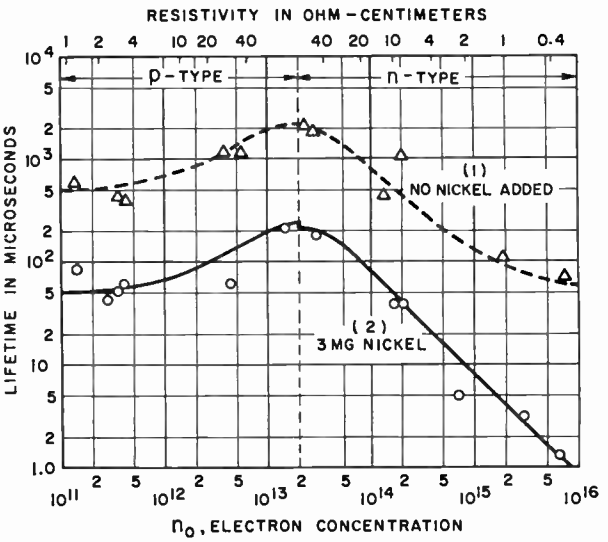


Fig. 4—Lifetime in nickel doped germanium as a function of resistivity [60].

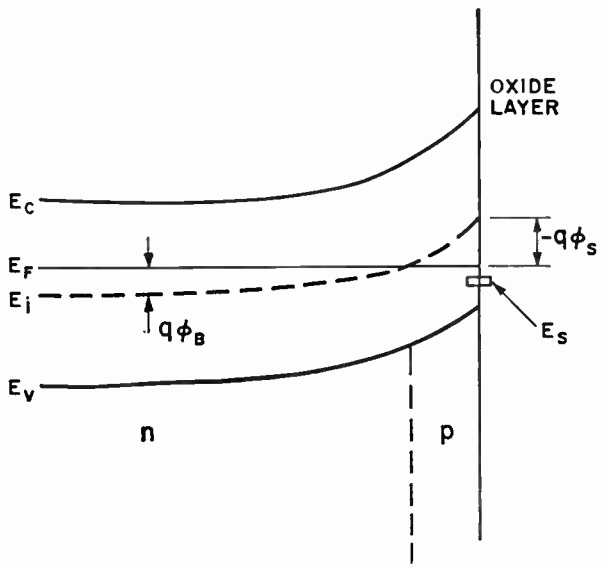


Fig. 5—Energy-level diagram of germanium surface [30].

An analysis, analogous to the Shockley-Read-Hall analysis for the case of the bulk of a semiconductor has been carried out by Brattain and Bardeen for surface recombination [31]. Stevenson and Keyes obtained, on the same basis, the following expression for the surface recombination velocity s [30].

$$s = C_pC_nN_t(p_0 + n_0)[C_n(n_s + n_{s1}) + C_p(p_s + p_{s1})]^{-1}. \tag{18}$$

This leads to a plot of s as a function of ϕ_s for a single level of surface states at E_S (Fig. 5). s represents therefore a condition of the surface and is closely related to the density of surface states responsible for recombination. It is thus apparent that the recombination mechanisms at the surface and in the body of germanium or silicon are very similar ones, the main difference being a purely dimensional one. The recombination rate, U, describes the volume recombination in three dimensions,

while the surface recombination velocity s describes a two-dimensional process.

III. Measurements

There exists at the present time a large number of methods used for measurement of the bulk lifetime τ, and the surface-recombination velocity s. We shall limit ourselves to a cursory description of a few basic methods. Many other very useful ones have been devised which in most cases are variants of the ones to be described.

In all cases injection of carriers has to be accomplished. This injection can take place either by optical means or by electrical means at an emitter in contact with the semiconductor.

Optical injection generates electron-hole pairs due to the absorbed radiation. Electrons and holes are generated above their thermal equilibrium densities in equal concentrations and the condition of electrical neutrality is preserved. After injection the recombination of carriers is measured by one of the various methods. The light source has to be sufficiently intense to produce a detectable modulation of the original conductivity of the sample whose lower limit of resistivity is partially determined by the intensity of the penetrating radiation.

In the case of measurements of bulk lifetime, it is necessary to use sufficiently penetrating radiation in order to generate the carriers in the volume of the sample. Otherwise the decay of the carriers produced near the surface contributes significantly to the observed recombination process. A simple method for obtaining a penetrating radiation is the use of a filter of the same material as the sample placed between the source and the sample.

The most frequently used sources are a spark gap, a flash tube, constant light (*i.e.*, carbon arc) in connection with a rotating mirror, and a Kerr cell together with a constant light source.

Injection of carriers can also be accomplished at a point contact to the semiconductor or at a p-n junction. However, due to over-all electrical neutrality, an equal concentration of majority carriers (of opposite sign) must appear at any point in the semiconductor. All the methods of lifetime measurements are concerned with the measurement of change in concentration of excess carriers. They can be divided into transient methods and steady-state methods.

A. Transient Methods

1) *Filamentary Transistor:* The first method of measuring lifetime was used by Haynes and Shockley [32]. It is often called the "filamentary transistor" method. A pulse of carriers is injected at a point contact biased as emitter and is swept along the filament by an externally applied electric field. The pulse is then detected at a second point contact biased as collector and the shape of the pulse is examined as a function of the time of travel. The field and the spacing between the emitter and collector can be changed. In general one expects that the total area under the pulse will decrease as exp $(-t/\tau)$ with the time (t) of travel. However, the surface recombination often prevents an observation of a single exponential decay and the observed time constant τ_{obs} is actually composed of both surface and volume components (16). This method of lifetime measurement has one noticeable advantage. It is easy to perform checks concerning the mobility of the carriers constituting the traveling pulse. This in turn suggests whether trapping was important in the process.

2) *Photoconductivity Decay* [32], [33]: Short pulses of light are used to introduce excess carriers, resulting in increase of the conductivity of the sample. This increase can be measured by passing a small dc current through the sample and through a resistor. The decay of the voltage across the resistor after termination of the pulse is always observed. The lifetime is the time necessary for this voltage to decay to $1/e$ of its original amplitude. The decay constant can be affected by surface recombination and by trapping of carriers. In silicon the former usually results in a sharp initial decay and can be minimized by the use of penetrating radiation. Trapping in silicon has been studied by the use of this method [25]. It usually affects the decay curve by contributing a long-time constant tail, which however, can often be eliminated or reduced by dc illumination. This illumination produces carriers which fill the traps forcing the carriers produced by the pulsed light to recombine rather than be trapped.

The photoconductivity-decay method permits a resolution of lifetime limited by the sharpness of the cutoff of the light pulses. In practice this limit is about 0.1 μsec. Of the light sources previously mentioned, the spark gap most often reaches this limit. A Kerr cell can also be used for pulsing the light to resolve lifetimes of this order of magnitude. Pulses produced by rotating mirrors usually put a somewhat higher limit on the resolution of lifetimes. Photoconductivity decay also permits the investigations of surface recombination. It can also furnish information about lifetime in a localized portion of the sample if such a portion alone is exposed to light and provided the electric field E is small. There are several variants of this method, depending on the source of injection [34–36].

Navon *et al.* [34] have injected carriers electrically and have measured the decay of voltage during the time of travel of the carriers between two contacts. If the lifetime is shorter than the transit time the analysis is similar to the photoconductivity case.

Wertheim [36] has adapted the method for measurements of lifetime of the order of 10^{-8} second by the use of a pulsed electron beam from a van de Graaff accelerator. Electrons of about 700 kev energy, producing a current of 15 μa/cm^2 are pulsed with a duty cycle of about 10^{-7}. This results in a generation of electron-hole pairs in silicon at a rate of 10^{20} cm^{-3} sec^{-1}. The radiation dam-

age produced in germanium or silicon during this measurement has been shown to be negligible on account of the low-duty cycle.

Many [37] observed the time constant of the voltage decay across a sample with the sample constituting one arm of a balanced bridge. The lifetime can be read directly as a product of RC with the voltage decay across the filament simulated by exponential decay of current charging a condenser through a resistance. This method is capable of measuring minority carrier mobility as well as injection rates. It is also used for measurements of surface recombination velocities. The balancing of the bridge presents a problem if the decay is not exponential, *i.e.*, in the presence of traps.

B. Steady-State Methods

1) *Diffusion Length Measurements:* In these measurements the diffusion length L is usually measured and the result converted to the lifetime. The simplest case is one in which the carriers injected optically are collected at a p-n junction [38, 39]. The distance between the collecting junction and the injecting light can be varied. The excess minority carrier density in n-type material can be expressed as:

$$\frac{d(\delta p)}{dt} = -\frac{\delta p}{\tau} \tag{19}$$

and the observed response in the simplest case is proportional to

$$\exp\left(-\frac{x}{\sqrt{D_p\tau_p}}\right). \tag{20}$$

Chopped light is usually used for reasons of sensitivity and the rms voltage across a resistor in series with the detector is measured. The motion of the carriers proceeds by diffusion as long as the applied electric field is small. Adam [40] improved this method by allowing for a simultaneous measurement of the diffusion constant. This is done by scanning a light spot across the junction.

Different geometries for the injecting light have been considered [41]. It is usually possible to determine τ and s in these experiments.

The advantage of this method consists of its freedom from trapping effects, as the diffusion length rather than time is measured. However, several authors [42] report difficulties with measurements on silicon due to surface conductance which upsets the exponential response. The experimental resolution in the diffusion length method is limited by the size of the light spot. The experimental arrangement requires sophisticated instrumentation if lifetimes shorter than 1μsec are to be resolved [43]. The final resolution is of course dependent on the mobility of minority carriers, the higher the mobility the lower the resolved lifetime. Use of X-rays for the production of carriers is reported by Malkovska [44].

2) *Photomagnetoelectric Effect:* This is an effect similar to the Hall effect except that it is associated with the diffusion of optically injected carriers [45], [46]. The direction of light and of the magnetic field are mutually perpendicular. The carriers produced by the light diffuse across the slab, but are deflected by the magnetic field in opposite directions. The electron and hole currents add and give a total PME short circuit current. In the open-circuit condition this current is cancelled by a drift current in the opposite direction. The total current will be composed in varying degrees of the PME current and the drift current, depending upon the volume recombination. If the short circuit PME current is measured as well as the relative conductance increase without the field, then it is possible to calculate the lifetime of carriers without the necessity of measuring the light intensity or the recombination velocity of the surface [47].

In particular in the case of a sample at least several diffusion lengths thick, and for injection which is not exceedingly high, one has:

$$I_{sc} = -\theta(\mu_n + \mu_p)^{-1}\left(\frac{D}{\tau}\right)^{1/2}\Delta G. \tag{21}$$

Eq. (21) implies that under identical experimental conditions the product of lifetime and mobility is constant. The PME method is therefore able to resolve very low lifetimes in semiconductors with highly mobile carriers. Determinations of lifetime as low as 10^{-9}–10^{-10} seconds have been reported, and several semiconducting materials have been examined [48]–[54]. Buck and McKim [48] have investigated the PME effect in germanium finding a good agreement with the theory. The lifetimes obtained agreed well with those determined by photoconductivity decay methods. The method is also very versatile in its ability to measure the recombination velocities of both the illuminated and the dark surfaces.

C. Other Methods

Measurements of lifetime based on the combined effect of electric and magnetic fields on injected carriers (Suhl effect) have also been reported [55]. The electric field deflects the carriers toward one side. For strong electric fields the minority carriers flow in a thin layer close to the surface and the surface recombination velocity can be studied.

Another method consisting of electrical injection of a pulse of carriers across a p-n junction followed by reversal of the polarity of the junction and subsequent collection of the carriers by the same junction has been used to determine the lifetime in the neighborhood of the junction [56–58]. Harrick [59] has developed a very elegant method which allows measurement of lifetime without contacts to the sample. It consists of optical creation of carriers and simultaneous observation of infrared absorption in the sample due to these carriers.

IV. Recombination Due to Specific Mechanisms

A. Impurities

Impurities are very frequently encountered lifetime limiting imperfections. Their recombination properties have been studied in several cases when sufficient information concerning their other electrical properties has become available. In all cases, these experiments have dealt with recombination at impurities which introduce energy levels fairly deep in the energy gap ($\geqq$0.15 ev from either of the bands). With the exception of indium in germanium there are no data concerning lifetimes due to "shallow" column III and V elements. It is, in general, believed that such impurities would have low-capture cross sections for carriers of similar charge, but high probability of escape exists for carriers of sign opposite to that of the impurities. The presence of these impurities would be fairly ineffective as recombination centers except possibly at low temperatures.

Burton, *et al.* [60] have measured lifetime in copper and nickel doped germanium as a function of donor and acceptor concentrations. They have established that these two impurities lead to an increase in the recombination rates in agreement with the Shockley-Read-Hall single level theory. Their results for copper are plotted in Fig. 6. The room-temperature capture cross sections are given in Table III (opposite).

Baum and Battey [61–62] have measured the lifetime in copper and nickel doped germanium as a function of temperature, and interpret the results in terms of capture cross sections increasing exponentially with increasing temperature. Okada [24] interprets similar data by assuming that in nickel doped germanium both nickel acceptor levels contribute to the recombination process; this latter explanation does not result in an exponential dependence of the capture cross section on temperature.

The variation of capture cross section with temperature is also reported by Fan, *et al.* [63] in their studies of recombination in germanium without specifically added recombination centers. They observe that the capture cross section decreases slowly as temperature increases.

A similar observation is made by Wertheim on plastically deformed germanium. His data are consistent with the assumption that the capture radius around dislocations varies as T^3 [64].

A decrease of the capture cross section with increasing temperature has been theoretically postulated by Lax [7], and should be more pronounced in the case of capture by charged centers than by neutral centers.

In general, the experimental results imply that a very careful analysis of the data concerning changes of lifetime with temperature is necessary in order to separate the different temperature dependent terms and arrive at the correct interpretation of the process.

The capture cross sections should be the highest in the case of a coulombic attraction between the impurity and the carrier. Weaker interaction is expected between uncharged impurity (*i.e.*, unionized donor or acceptor level) and the charged carrier. Still weaker interaction should be present if the carrier and the center have charges of the same sign. A very good proof of these assertions has been given by Tyler and Woodbury [65]. Iron, cobalt, manganese, nickel, and gold introduce at least two acceptor levels in germanium [66]. It is observed that under conditions such that one of these levels has a double negative charge, it attracts a hole. However, the center plus the hole still have a single negative charge with a very low-capture cross section for electrons. The hole may then have a higher chance of escaping back to the valence band, than of recombining with an electron. This results in a "hole-trap" behavior of these centers. If however the Fermi level is located below the second acceptor level, no doubly negative center will exist. The holes can then be captured only by the negative singly charged acceptors and have a much higher-capture cross section for electrons. The process becomes one of recombination rather than trapping. Typical values of the hole-capture cross section by doubly charged acceptors are: 10^{-13}–10^{-15} cm^2, while for electrons they are as low as 10^{-22} cm^2.

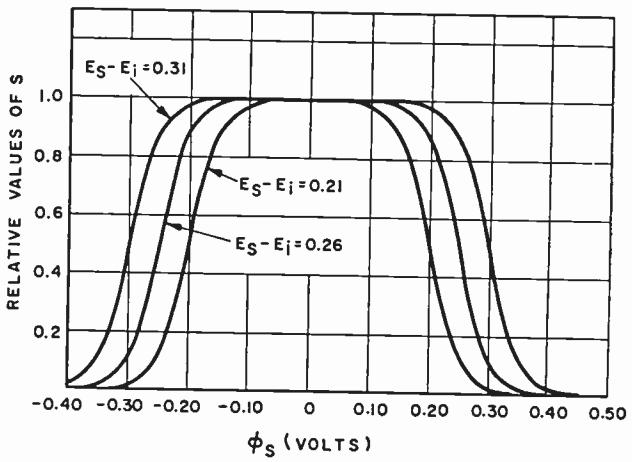


Fig. 6—Surface-recombination velocity s_1 vs ϕ_s for different positions of the surface-recombination centers, E_S [30].

Other authors [63], [67–70] report results of recombination studies in germanium crystals with no purposeful addition of impurities. It appears that in such cases the presence of one or more levels due to unknown impurities is often responsible for the observed effects. The accuracy of measurements and the complications in the interpretations of the experiments are, however, such that it is at present difficult or almost impossible to use the method of recombination studies as a sort of fine spectrometer in a reliable detection of unknown impurities in semiconductors. It is probable that in the future a fairly complete list of capture cross sections of a large number of impurities in semiconductors will permit such an analysis.

There are less data available concerning recombination due to the known impurities in silicon. Several of

TABLE III

CAPTURE CROSS SECTIONS OF DIFFERENT IMPERFECTIONS IN SILICON AND GERMANIUM

SILICON

	Impurities					Irradiation
	Fe [74]	Au [72]	Au[1] [73]	In [16]	Dislocations [93]	Electrons [84]
P type $\sigma_p(cm^2)$			4.10^{-15}	$1.7\cdot10^{-14}$		$1.8\cdot10^{-15}$
$\sigma_n(cm^2)$	$>1.5\cdot10^{-15}$	$3.5\cdot10^{-15}$	1.10^{-13}	$3.7\cdot10^{-17}$	$\tau=\frac{60^2}{N}$	$1.9\cdot10^{-15}$
N type $\sigma_p(cm^2)$	3.10^{-16}	1.10^{-15}	3.10^{-13}			8.10^{-13}
$\sigma_n(cm^2)$		5.10^{-16}	4.10^{-15}			$9.5\cdot10^{-15}$

GERMANIUM

	Impurities							Irradiation		
	Cu [60]	Ni [60]	Mn [129][3]	Co [129][3]	Fe [129][3]	Au [130][1]	Dislocations [94]	Electrons [86]	γ-rays [85]	Neutrons [85]
P type $\sigma_p(cm^2)$		$\sim10^{-13}$ [131][1]				2.10^{-16}				
$\sigma_n(cm^2)$	1.10^{-17}	8.10^{-17}	10^{-16}–10^{-15}	10^{-16}–10^{-15}	10^{-16}–10^{-15}	2.10^{-14}	$\tau=\frac{1.22^5}{N}$			
N type $\sigma_p(cm^2)$	1.10^{-16}	$> 4.10^{-15}$					$\tau=\frac{4.14}{N}$	$1.10^{-16\,6}$	5.10^{-16}	4.10^{-15}
$\sigma_n(cm^2)$			$<10^{-22\,4}$			$\sim10^{-16}$		$5.10^{-17\,7}$		

[1] At 77°K.
[2] In high-resistivity silicon.
[3] At low temperatures, capture due to neutral impurity.
[4] At singly charged, negative Mn.
[5] N = dislocations/cm².
[6] 750 kev electrons.
[7] 550 kev electrons.
Note: Also, *N*-type germanium—$\sigma_p = 3.10^{-15}$ cm² at Fe^-
$\sigma_p = 1.10^{-14}$ cm² at Fe^{--}
(K. D. Glinchunk, E. G. Miseluk, and N. N. Fortunatova, *Zh. Tekh. Fiz.*, vol. 27, pp. 2451–2457; November, 1957.)

the elements outside of columns III and V, known to introduce donor or acceptor levels in silicon, also possess high-capture cross sections. This pertains to gold [71–73], iron [74], copper [74], [75], and manganese [76]. The known capture cross sections are in the 10^{-16}–10^{-13} cm² range. Wertheim has also studied the recombination of the column III acceptor Indium [16]. Its capture cross section for holes is of the order of 10^{-14} cm² at room temperature. Due to the competing trapping effects in silicon the experimental difficulties in transient type experiments are often pronounced.

Several authors report lifetime observations on pure silicon [77–82] (no recombination centers added purposefully). There is a general agreement that the lifetime is found to increase with injection density, particularly in *p*-type material. Several authors have observed behavior in silicon which implies the existence of more than one level [78], [80], [81]. It is probable that different raw material used and different crystal growing techniques may be responsible for the presence of varying amounts of different residual imperfections in such cases.

B. Bombardment Damage

Resistivity measurements on semiconductors subjected to irradiation with high energy particles (neutrons, deuterons, electrons, heavy ions, alpha particles), or with gamma rays, indicate that the lattice gets disrupted in the process. It is believed that in the simplest case the damage consists of introduction of isolated interstitials and vacancies. In germanium and silicon donor and acceptor levels in the forbidden energy gap after irradiation have been identified [83], [84].

These imperfections also have a pronounced effect on the lifetime of excess carriers. In recent years the recombination processes in irradiated germanium and silicon have been studied [84–86] and have been used to investigate the damage introduced [87], [88].

Russian investigators [86] report that, in *n*-type germanium, the hole-capture cross section increases from

5.10^{-17} cm² to 1.10^{-16} cm² as the energy of the bombarding electrons is increased from 550 kev to 750 kev. This may indicate that at higher energy a more complex type of damage is introduced beyond the simple, isolated, interstitial-vacancy pairs believed to be produced at low energies.

Curtis, *et al.* [85], after irradiation of *n*-type germanium, have reported results which point in this same direction. They found that the capture cross section for holes is larger after neutron irradiation than after gamma irradiation, probably for the same reasons mentioned above. Hole trapping in germanium has been reported by Cranford [89] and studied by Shulman [26] who has observed densities of traps of the same order of magnitude as the densities of acceptor introduced on electron bombardment.

Loferski and Rappaport [87] have studied the threshold of introduction of damage as a function of the electron energy in silicon and germanium by means of lifetime measurements. Wertheim has studied recombination processes in silicon irradiated with 0.7 mev electrons [84]. The capture cross sections are given in Table III. As few as 10^{14} electrons/cm² are found to be sufficient to produce a noticeable change in the lifetime. The kinetics of the annealing cycle in silicon have also been studied with the use of lifetime measurements [88]. The lifetime effects of the electron damage proved to be reversible on heating in the temperature range of 200°–400°C.

Damage in irradiated semiconductors may well be studied by means of lifetime measurements, a method particularly attractive in relatively weak-bombarded specimens in which small changes in carrier concentrations are introduced. This may prove advantageous in the studies of relatively isolated defects introduced on irradiation.

C. Dislocations

Several papers [90–98] deal with recombination rates associated with edge type dislocations in semiconductors. McKelvey [94] reports that the recombination cross sections in germanium correspond to a cylindrical area about a dislocation of diameter 1.15Å for holes in *n*-type Ge, and 2.8Å for electrons in *p*-type germanium. It is believed from this data that the frequently encountered lifetime in germanium, of the order of milliseconds, is limited by the dislocations rather than by impurities. Most authors agree quite closely on the values of recombination cross sections. The problem of knowing the exact dislocation density is particularly serious in silicon, where different methods of determination of the densities often provide different numerical answers.

It is reported that the lifetime in pbs in the 0.1 μsec–10 μsec range is limited by the presence of dislocations and obeys the Shockley-Read relations [95].

In very recent diffusion length measurements on silicon and germanium it is claimed that the diffusion length is anisotropic, being higher in the direction parallel than in the direction perpendicular to dislocations [98]. The authors interpret this as an anisotropy in the lifetime rather than in the diffusion constant. However, Logan's data on the anisotropy of mobility at dislocations in germanium imply that the anisotropy in D needs to be considered [99].

D. Radiative Recombination

Several authors have observed recombination radiation in semiconductors. These results are of great interest for the theoretical interpretation of the band structure. They also can provide direct data concerning the energy levels of recombination centers.

Recombination with the resulting emission of photons has been observed in: germanium [100], [102], [105], [106], silicon [101], [107], silicon carbide [103], [104], indium antimonide [108], gallium antimonide [109] gallium arsenide [109], indium phosphide [109], and germanium silicon alloys [109].

In most cases the wavelength of the radiation corresponded to intrinsic recombination. In silicon Haynes also observed recombination of the electrons at the acceptors [101]. Similarly in germanium Newman [102] has detected radiation arising from recombination at dislocations and at copper atoms. Photon emission in SiC is believed [104] to be due to recombination at some imperfections in the energy gap. In all these cases the observed radiation represents probably only a small fraction of the total energy lost in the process.

In InSb, Wertheim [13] observed that the direct recombination may be the limiting process near room temperature, while at lower temperatures a recombination center mechanism dominates. Similarly Redfield [110] reports that in Tellurium the direct recombination dominates at 100°K, even though no recombination radiation has been detected. DeCarvalho [54] believes that the Auger mechanism is operative in Tellurium.

E. Surface Recombination

The theoretical picture of surface states [111] has been verified on many occasions. A large number of experiments has been performed whose results are described in terms of surface states [112–114].

It is generally believed that there exist two kinds of states on the surface of germanium, the semiconductor whose surfaces have been by far the most widely studied. The first type, the "fast" states are located at the interface of the semiconductor and the almost always present oxide layer. The density of these states is about 10^{11} cm^{-2} and depends on the initial treatment of the surface. These states are believed to be responsible for surface recombination as well as effects observed in field effect and photo voltage experiments. It has been observed that the density of these states is increased if the surface is mechanically damaged, heated in vacuo bombarded with ions, etched in HF [115], [116], *etc.* This implies an increase in the values of *s*.

An exact correlation of the surface states with a definite type of imperfection has not as yet been achieved. By specifying their position in energy, their state of charge (donor, acceptor), and their capture cross sections in the Shockley-Read sense, the recombination properties of the states can be described similarly to the recombination centers present in the volume of the semiconductors.

Many of the experimental techniques measure the properties of the surfaces, such as the surface conductance, inversion layers, contact potential, and surface-recombination velocities, as a function of ϕ_s. This can then be compared with the plot of Fig. 5 [30]. Several authors believe at the present time that, while basically correct, the actual description of recombination at the surfaces is more complex and involves two or more levels, or even a continuum of levels [113]. Garrett and Brattain obtained the values for the capture cross sections for holes and electrons [117], (see Table III) which indicate that the fast states near the center of the gap are acceptor like. The magnitudes of these cross sections are of the same order as for the case of the volume levels. Statz finds levels in germanium at 0.14 ev below the middle of the gap and in silicon at 0.42–0.48 above the middle, also at 0.45 ev below the middle of the gap [113].

There exists another group of states, in densities exceeding 10^{13} cm^{-2}, so called "slow" states, which exhibit long decay times [113]. These states are probably due to imperfections at the surface of the oxide or in the oxide layer. Their capture cross sections are orders of magnitude lower than of the "fast" states. These states are responsible for the variation of ϕ_s with the ambient gases, but not for the observed recombination velocities.

Considerable effort has gone into production of a perfectly clean germanium surface [113]. Handler has observed under "clean" conditions, a density of acceptor like surface states in excess of 10^{13}/cm^2.

To make the analogy with the bulk properties complete, one should add that trapping effects have also been observed on germanium surfaces, indicating that for some centers the capture probability is much higher for one type of carrier than for the other [118]. It is justifiable to say that the basic understanding of surface phenomena is fairly complete and that the difficulties lie in the fact, at least in germanium, that the experimental situation is quite complex. The simplest case which can be achieved in the bulk of the semiconductors, consisting of a single recombination level due to known imperfections, is not as easily achieved on the surfaces.

V. Practical Aspects

The importance of lifetime in transistor technology is too well-known to be repeated. It is interesting, however, to examine some cases of more practical interest, in which the lifetime has been affected in certain controlled ways.

Reduction of storage time in *p-n* junction silicon diodes has been accomplished by neutron [119] or electron irradiation [120]. On the opposite side of the picture, attempts are frequently made to conserve "high" lifetime after heat treatment of silicon or germanium at elevated temperatures. The complexity of this problem is considerable because of the very minute concentration of impurities which can have a pronounced effect in lowering the lifetime. A typical impurity (see Table III) with a capture cross section of 10^{-15} cm^2 present in concentration of 10^{14} cm^{-3} may lower the lifetime to about 1 μsec. Fuller and Logan [121] have shown that lifetimes as high as 100 μsec or more can be maintained in germanium heated to 875°C if extreme cleanliness is used. It is also necessary to avoid introduction of strains or plastic flow. It is believed that the introduction of copper, in particular, has to be minimized if high lifetimes are desired. Logan [122] has also shown that gettering of copper from germanium is possible in the presence of a third component such as liquid lead or gold on the surface of germanium. Restoration of the original lifetime and resistivity is then possible.

Several papers deal with the similar problem in silicon [123–127]. In view of the higher temperatures involved in the fabrication of silicon devices, the problem is even more acute here. It has been previously mentioned that gold, copper, iron, manganese, and dislocations reduce the lifetime in silicon. It is, however, possible to employ gettering techniques by the use of nickel or some other metallic films on the surfaces of silicon to remove the recombination centers, or to prevent them from diffusing into silicon [72], [125]. Gold, for example, can be effectively removed from the bulk by the use of a third component on the surface (*i.e.*, nickel). The thermodynamics of these processes have been discussed by Thurmond [128].

It has been satisfactorily demonstrated that under ideally clean conditions floating zone silicon can be heated without appreciable loss of lifetime [124]. This requires at the present time the absence of any handling of the crystal between the time of growth and the heat treatment. The implication of this is that impurities are responsible for the decrease of lifetime observed normally in heat treatment. At least one of the impurities listed above, gold, has been identified on the surfaces of most crystals which have been chemically etched [72]. Even though the surface concentration of gold corresponded only to 10^{-3} of a monolayer it supplied the explanation of some of the electrical changes which were subsequently observed on heat treatment.

VI. Conclusions

The extremely high degree of purity achieved in germanium and silicon single crystals in the last few years has provided scientists with excellent opportunities for research concerning recombination phenomena under better controlled conditions than were previously

possible. In particular, the ideal experimental situation, in which the process is studied in the presence of minute concentrations of imperfections which have been added in a controlled fashion, makes it possible to test the existing recombination theories. It appears that the recombination at localized levels in the energy gap is the prevalent condition in silicon and germanium. The dependence of semiconductor lifetime on temperature and concentration of excess carriers can, theoretically, be controlled. This dependence is important in device operation. Clearly, this is an excellent example of the extent of practical potentialities offered by basic understanding of physical phenomena.

VII. Acknowledgment

The author wishes to thank several of his colleagues for illuminating discussions during the preparation of this paper. He also wishes to thank the Lincoln Laboratory for the permission to reproduce Figs. 5 and 6 in this paper.

VIII. Bibliography

[1] Breckenridge, R. G., Russell, B. R., and Hahn, E. E., eds. *Photoconductivity Conference (Atlantic City, 1954)*. New York: John Wiley and Sons, Inc., 1956.
[2] Schultz, M. L. and Morton, G. A. "Photoconduction in Germanium and Silicon," Proceedings of the IRE, Vol. 43 (December, 1955), pp. 1819–1882.
[3] Fifth Conference on Luminescence, *Izvestia Akademiya Nauk SSSR (Moscow), Seriia Fizicheskaia*, Vol. 21 (May, 1957), pp. 643–775.
[4] Van Roosbroeck, W. and Shockley, W. "Photon-Radiative Recombination of Electrons and Holes in Germanium," *Physical Review*, Vol. 94 (June, 1954), pp. 1558–1560.
[5] Shockley, W. and Read, W. T., Jr. "Statistics of the Recombinations of Holes and Electrons," *Physical Review*, Vol. 87 (September, 1952), pp. 835–842.
[6] Hall, R. N. "Electron-Hole Recombination in Germanium," *Physical Review*, Vol. 87 (July, 1952), p. 387.
[7] Lax, M. Second Symposium on the Physics of Semiconductors, Washington, D. C., 1956.
[8] Pincherle, L. "Auger Effect in Semiconductors," *Proceedings of the Physical Society*, Vol. 68B (May, 1955), pp. 319–320.
[9] Bowlden, H. J. "Radiative Transitions in Semiconductors," *Physical Review*, Vol. 106 (May, 1957), pp. 427–431.
[10] Bess, L. "Possible Mechanism for Radiationless Recombination in Semiconductors," *Physical Review*, Vol. 105 (March, 1957), pp. 1469–1475.
[11] Dumke, W. P. "Indirect Transitions at the Center of the Brillouin Zone with Application to InSb," *Physical Review*, Vol. 108 (December, 1957), pp. 1419–1425.
[12] MacKintosh, I. M. "Photon Radiative Recombination in PbSe, PbTe and PbS," *Proceedings of the Physical Society*, Vol. 69B (January, 1956), p. 115.
[13] Wertheim, G. K. "Carrier Lifetime in Indium Antimonide," *Physical Review*, Vol. 104 (November, 1956), pp. 662–664.
[14] Adirovich, E. I. and Guro, G. M. "Characteristic Times of Electronic Processes in Semiconductors," *Soviet Physics Doklady*, Vol. 108 (May, 1956), p. 417.
[15] Sandiford, D. J. "Carrier Lifetime in Semiconductors for Transient Conditions," *Physical Review*, Vol. 105 (January, 1957), p. 524.
[16] Wertheim, G. K. "Transient Recombination of Excess Carriers in Semiconductor," *Physical Review*, Vol. 109 (February, 1958), pp. 1086–1091.
[17] Clarke, D. H. "Semiconductor Lifetime as a Function of Recombination State Density," *Journal of Electronics and Control*, Vol. 3 (October, 1957), pp. 375–386.
[18] Pell, E. M. "Reverse Current and Carrier Lifetime as a Function of Temperature in Germanium Diodes," *Journal of Applied Physics*, Vol. 26 (June, 1955), pp. 658–665.
[19] Bernard, M. "Mesures en Fonction de la Temperature du Courant dans les Jonctions de Germanium *n-p*," *Journal of Electronics*, Vol. 2 (May, 1957), pp. 579–596.
[20] Sah, C., Noyce, R. N., and Shockley, W., "Carrier Generation and Recombination in *p-n* Junctions and *p-n* Junction Characteristics," Proceedings of the IRE, Vol. 45 (September, 1957), pp. 1228–1243.
[21] Landsberg, P. T. "Defects with Several Trapping Levels in Semiconductors," *Proceedings of the Physical Society*, Vol. 69B (October, 1956), pp. 1056–1058.
[22] Kalashnikov, S. G. "Recombination of Electrons and Holes in the Presence of Various Types of Traps," *Zhurnal Tekhnicheskoĭ Fiziki*, Vol. 26 (February, 1956), pp. 241–250.
[23] Rose, A. "Recombination Process in Insulators and Semiconductors," *Physical Review*, Vol. 97 (January, 1955), pp. 322–333.
[24] Okada, J. "Recombination of Excess Carriers in Semiconductors," *Journal of the Physical Society, Japan*, Vol. 12 (December, 1957), pp. 1338–1343.
[25] Haynes, J. R. and Hornbeck, J. "Trapping of Minority Carriers in Silicon," *Physical Review*, Vol. 97 (January, 1955), pp. 311–321, and *Physical Review*, Vol. 100 (October, 1955), p. 606.
[26] Shulman, R. G. "Hole Trapping in Germanium Bombarded by High Energy Electrons," *Physical Review*, Vol. 102 (June, 1956), pp. 1451–1455.
[27] Shockley, W. "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., p. 321, 1950.
[28] Tamm, I. "Über eine mögliche art der elektronen-bindung an Kristalloberflächen," *Physik Zeitschrift Sowjetunion*, Vol. 1, No. 6 (1932), pp. 733–746.
[29] Shockley, W. "On the Surface States Associated with a Periodic Potential," *Physical Review*, Vol. 56 (August, 1939), pp. 317–323.
[30] Stevenson, D. T. and Keyes, R. J. "Measurement of the Recombination Velocity at Germanium Surfaces," *Physica*, Vol. 20 (August, 1954), pp. 1041–1046.
[31] Brattain, W. H. and Bardeen, J. "Surface Properties of Germanium," *Bell System Technical Journal*, Vol. 32 (January, 1953), pp. 1–41.
[32] Haynes, J. R. and Shockley, W. "The Mobility and Life of Injected Holes and Electrons in Germanium," *Physical Review*, Vol. 81 (March, 1951), pp. 835–843.
[33] Stevenson, D. T. and Keys, R. J. "Measurement of Carrier Lifetimes in Germanium and Silicon," *Journal of Applied Physics*, Vol. 26 (February, 1955), pp. 190–195.
[34] Navon, D., Bray, R., and Fan, H. Y. "Lifetime of Injected Carriers in Germanium," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1342–1347.
[35] Avery, D. G. and Gunn, J. B. "The Use of a Modulated Light Spot in Semiconductor Measurements," *Proceedings of the Physical Society*, Vol. 68B (November, 1955), pp. 918–921.
[36] Wertheim, G. K. and Augustyniak, W. M. "Measurement of Short Carrier Lifetimes," *Review of Scientific Instruments*, Vol. 27 (December, 1956), pp. 1062–1064.
[37] Many, A. "Measurement of Minority Carrier Lifetime and Contact Injection Ratio on Transistor Material," *Proceedings of the Physical Society*, Vol. 67B (January, 1954), pp. 9–17.
[38] Goucher, F. S. "Measurement of Hole Diffusion in *n*-type Germanium," *Physical Review*, Vol. 81 (February, 1951), p. 475.
[39] Valdes, L. "Measurement of Minority Carrier Lifetime in Germanium," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1420–1423.
[40] Adam, G. "A Flying Light Spot Method for Simultaneous Determination of Lifetime and Mobility of Injected Current Carriers," *Physics*, Vol. 20 (August, 1954), pp. 1037–1041.
[41] See, *i.e.*, Van Roosbroeck, W. and Buck, T. M. *Transistor Technology*, New York: D. Van Nostrand Co., Inc., Vol. 3, ch. 9, 1958.
[42] Arthur, J. B., Bardsley, W., Gibson, A. F., and Hogarth, C. A. "On the Measurement of Minority Carrier Lifetime in *n*-type Silicon," *Proceedings of the Physical Society*, Vol. 68B (March, 1955), p. 121.
[43] Gudmundsen, R. A., Waters, W. P., Wannlund, A. L., and Wright, W. V. "Recent Developments in Silicon Fusion Transistors," IRE Transactions on Electronic Devices, Vol. ED-2 (January, 1955), pp. 74–81.
[44] Malkovska, M. "The Diffusion Distances of Electrons and Holes in Germanium Measures by X-rays," *Czechoslovak Journal of Physics*, Vol. 5 (December, 1955), pp. 545–546.
[45] Kikoin, I. K. and Noskov, M. M. "A New Photoelectric Effect in Cuprous Oxide," *Physik Zeitschrift Sowjetunion*, Vol. 5, No. 5 (1934), pp. 586–596.
[46] Garreta, O. and Grosvalet, J. "Progress in Semiconductors," (review paper), Vol. 1, Heywood and Co., London, p. 165, 1956.
[47] Van Roosbroeck, W. "Theory of the Photomagnetoelectric Effect in Semiconductors," *Physical Review*, Vol. 101 (March, 1956), pp. 1713–1725.
[48] Buck, T. M. and McKim, F. S. "Experiments on the Photomagnetoelectric Effect in Germanium," *Physical Review*, Vol 106 (June, 1957), pp. 904–909.

[49] Moss, T. S., Pincherle, L., and Woodward, A. M. "Photoelectromagnetic and Photodiffusion Effects in Germanium," *Proceedings of the Physical Society*, Vol. 66B (September, 1954), pp. 743–752.
[50] Moss, T. S. "Photoelectromagnetic and Photoconductive Effects in Lead Sulphide Single Crystals," *Proceedings of the Physical Society*, Vol. 66B (December, 1953), pp. 993–1002.
[51] Aigrain, P. "Mesures de durée de vie des Porteurs Minoritaires dans les Semi-Conducteurs," *Annales de Radioelectricite Compagnies, Gen. TSF*, Vol. 9 (July, 1954), pp. 219–226.
[52] Oberly, J. J. "Photoelectric Hall Effect in Germanium Single Crystals," *Physical Review*, Vol. 93 (February, 1954), p. 911.
[53] Kurnick, S. W., Strauss, A. J., and Zitter, R. N. "Photo-Conductivity and Photoelectromagnetic Effect in InSb," *Physical Review*, Vol. 94 (June, 1954), p. 1791.
[54] Pires de Carvalho, A. "Etudes du Processus de Recombinaison dans le Tellure," *Comptes Rendus*, Vol. 244 (January, 1957), pp. 461–462.
[55] Suhl, H. and Shockley, W. "Concentrating Holes and Electrons by Magnetic Field," *Physical Review*, Vol. 75 (May, 1949), pp. 1617–1618.
[56] Pell, E. M. "Recombination Rate in Germanium by Observation of Pulsed Reverse Characteristic," *Physical Review*, Vol. 90 (April, 1953), pp. 278–279.
[57] Kingston, R. H. "Switching Time in Junction Diodes and Junction Transistors," PROCEEDINGS OF THE IRE, Vol. 42 (May, 1954), pp. 829–834.
[58] Lax, B. and Neustadter, S. "Transient Response of a *p-n* Junction," *Journal of Applied Physics*, Vol. 25 (September, 1954), pp. 1148–1154.
[59] Harrick, N. J. "Lifetime Measurements of Excess Carriers in Semiconductors," *Journal of Applied Physics*, Vol. 27 (December, 1956), pp. 1439–1442.
[60] Burton, J. A., Hull, G. W., Morin, F. J., and Severiens, J. C. "Effect of Nickel and Copper Impurities on the Recombination of Holes and Electrons in Germanium," *Journal of Physical Chemistry*, Vol. 57 (November, 1953), pp. 853–859.
[61] Baum, R. M. and Battey, J. F. "Electron Capture Probability of the Upper Copper Acceptor Level in Germanium," *Physical Review*, Vol. 98 (May, 1955), pp. 923–925.
[62] Battey, J. F. and Baum, R. M. "Carrier Capture Probabilities in Nickel Doped Germanium," *Physical Review*, Vol. 100 (December, 1955), pp. 1634–1637.
[63] Fan, H. Y., Navon, D., and Gebbie, H. "Recombination and Trapping of Carriers in Germanium," *Physica*, Vol. 20 (August, 1954), pp. 855–872.
[64] Wertheim, G. K. and Pearson, G. L. "Recombination in Plastically Deformed Germanium," *Physical Review*, Vol. 107 (August, 1957), pp. 694–698.
[65] Tyler, W. W. and Woodbury, H. H. "Scattering of Carriers from Doubly Charged Impurity Sites in Germanium," *Physical Review*, Vol. 102 (May, 1956), pp. 647–655.
[66] See Conwell, E., this issue, p. 1300.
[67] Rzhanov, A. V. "Influence of Impurities on Lifetime of Excess Charge Carriers in Germanium," *Zhurnal Tekhnicheskoi Fiziki*, Vol. 26 (July, 1956), pp. 1389–1393.
[68] Paramanova, A. and Rzhanov, A. V. "Dependence of Lifetime on Concentration of Equilibrium Carriers of Charge," *Zhurnal Tekhnicheskoi Fiziki*, Vol. 25 (July, 1955), pp. 1342–1344.
[69] Ransom, P. and Rose, F. W. G. "Trap Activation Energies in *n*-type Germanium," *Journal of Electronics*, Vol. 1 (May, 1956), pp. 625–628.
[70] Ostroborodova, V. V. and Kalashnikov, S. G. "Recombination of Excess Charge Carriers at Thermal Acceptors in Germanium," *Zhurnal Tekhnicheskoi Fiziki*, Vol. 25 (July, 1955), p. 1168.
[71] Collins, C. B., Carlson, R. O., and Gallagher, C. J. "Properties of Gold Doped Silicon," *Physical Review*, Vol. 105 (February, 1957), pp. 1168–1173.
[72] Bemski, G. and Struthers, J. "Gold in Silicon," presented at the Meeting of the Electrochemical Society, Buffalo; 1957.
[73] Davis, W. D. Private communication.
[74] Collins, C. B. and Carlson, R. O. "Properties of Silicon Doped with Iron or Copper," *Physical Review*, Vol. 108 (December, 1957), pp. 1409–1414.
[75] Bemski, G. and Bridgers, H., Jr. "Copper in Silicon," presented at the Meeting of the Electrochemical Society, Cleveland; 1956.
[76] Carlson, R. O. "Properties of Silicon Doped with Manganese," *Physical Review*, Vol. 104 (November, 1956), pp. 937–941.
[77] Bittmann, C. A. and Bemski, G. "Lifetime in Pulled Silicon Crystals," *Journal of Applied Physics*, Vol. 28 (December, 1957), pp. 1423–1426.
[78] Blakemore, J. S. "Lifetime of Minority Carriers in Single Silicon Crystals," *Bulletin of the American Physical Society*, Vol. 2 (March, 1957), p. 153.
[79] Bemski, G. "Lifetime of Electrons in *p*-type Silicon," *Physical Review*, Vol. 100 (October, 1955), pp. 523–524.
[80] Ridout, M. S. "The Temperature Dependence of Minority Carrier Lifetime in *p*-type Germanium and Silicon," Report of the Meeting on Semiconductors, Rugby, England; 1956.
[81] Ross, B. and Madigan, J. R. "Thermal Generation of Recombination Centers in Silicon," *Physical Review*, Vol. 108 (December, 1957), pp. 1428–1433.
[82] Pell, E. M. and Roe, G. M. "Reverse Current and Carrier Lifetime as a Function of Temperature in Silicon Junction Diodes," *Journal of Applied Physics*, Vol. 27 (July, 1956), pp. 768–772.
[83] Fan, H. Y. and Horowitz, K. L. "Irradiation of Semiconductors," presented at the International Conference on Semiconductors and Phosphors, Garmisch Partenkirchen; 1956.
[84] Wertheim, G. K. "Energy Levels in Electron-Bombarded Silicon," *Physical Review*, Vol. 105 (March, 1957), pp. 1730–1735.
[85] Curtis, O. L., Jr., Cleland, J. W., Crawford, J. H., Jr., and Pigg, J. C. "Effect of Irradiation on the Hole Lifetime of *n*-type Germanium," *Journal of Applied Physics*, Vol. 28 (October, 1957), pp. 1161–1165.
[86] Smirnov, L. S. and Vavilov, V. S. "On the Probability of Recombination Capture of Charge Carriers by Frenkel Defects," *Zhurnal Tekhnicheskoi Fiziki*, Vol. 27 (February, 1957), p. 427.
[87] Loferski, J. J. and Rappaport, P. "Electron Voltaic Study of Electron Bombardment Damage and its Thresholds in Germanium and Silicon," *Physical Review*, Vol. 98 (June, 1955), pp. 1861–1863.
[88] Bemski, G. and Augustyniak, W. M. "Annealing of Electron Bombardment Damage in Silicon Crystals," *Physical Review*, Vol. 108 (November, 1957), pp. 645–648.
[89] Crawford, J. H., Jr., Cleland, J. W., Horowitz, K. L., Pigg, J. C., and Young, F. W. "Evidence of Hole Traps in Germanium Produced by Fast Neutron Bombardment," *Physical Review*, Vol. 85 (February, 1952), p. 730.
[90] Vogel, F. L., Read, W. T., Jr., and Lovell, L. C. "Recombination of Holes and Electrons at Lineage Boundaries in Germanium," *Physical Review*, Vol. 94 (June, 1954), p. 1791.
[91] McKelvey, J. P. and Longini, R. L. "Recombination of Injected Carriers at Dislocation Edges in Semiconductors," *Physical Review*, Vol. 99 (August, 1955), pp. 1227–1232.
[92] Longini, R. L. "Temperature Dependent Factor in Carrier Lifetime," *Physical Review*, Vol. 102 (April, 1956), p. 584.
[93] Kurtz, A. D., Kulin, S. A., and Averbach, B. L. "Effect of Dislocations on the Minority Carrier Lifetime in Semiconductors," *Physical Review*, Vol. 101 (February, 1956), pp. 1285–1291.
[94] McKelvey, J. P. "Experimental Determination of Injected Carrier Recombination Rates at Dislocations in Semiconductors," *Physical Review*, Vol. 106 (June, 1957), pp. 910–917.
[95] Scanlon, W. W. "Lifetime of Carriers in Lead Sulfide Crystals," *Physical Review*, Vol. 106 (May, 1957), pp. 718–720.
[96] Morrison, S. R. "Recombination of Electrons and Holes at Dislocations," *Physical Review*, Vol. 104 (November, 1956), pp. 619–623.
[97] Okada, J. "Effects of the Dislocations on Minority Carrier Lifetime in Germanium," *Journal of the Physical Society, Japan*, Vol. 10 (December, 1955), pp. 1110–1111.
[98] Bell, R. L. and Hogarth, C. A. "Anisotropic Diffusion Length in Germanium and Silicon Crystals Containing Parallel Arrays of Edge Dislocations," *Journal of Electronics and Control*, Vol. 3 (November, 1957), pp. 455–470.
[99] Logan R., Private communication.
[100] Haynes, J. R. "New Radiation Resulting from Recombination of Holes and Electrons in Germanium," *Physical Review*, Vol. 98 (June, 1956), pp. 1866–1868.
[101] Haynes, J. R. and Westphal, W. C. "Radiation Resulting from Recombination of Holes and Electrons in Silicon," *Physical Review*, Vol. 102 (June, 1956), pp. 1676–1678.
[102] Newman, R. "Recombination Radiation from Deformed and Alloyed Germanium *p-n* Junction at 80°K," *Physical Review*, Vol. 105 (March, 1957), pp. 1715–1720.
[103] Lossey, O. V. "Luminous Carborundum Detector and Detection Effect and Oscillations with Crystals," *Philosophical Magazine*, Vol. 6 (November, 1928), pp. 1024–1044.
[104] Lehavec, K., Accardo, C. A., and Jamhogian, E. "Injected Light Emission of Silicon Carbide Crystals," *Physical Review*, Vol. 83 (August, 1951), pp. 603–607.
[105] Benoit a la Guillaume, C. "Recombination Light Emission Via Traps in Germanium," *Comptes Rendus*, Vol. 243 (August, 1956), pp. 704–707.
[106] Aigrain, P. "Light Emission from Injecting Contacts on Germanium in the 2M to 6M Band," *Physica*, Vol. 20 (August, 1954), pp. 1010–1013.
[107] Haynes, J. R., Lax, M., and Flood, W. "Fine Structure of Intrinsic Recombination Radiation in Silicon," *Bulletin of the American Physical Society*, Vol. 3 (January, 1958), p. 30.
[108] Moss, T. S. and Hawkins, T. H. "Recombination Radiation

from InSb," *Physical Review*, Vol. 101 (March, 1956), pp. 1609–1610.
[109] Braunstein, R. "Radiative Transitions in Semiconductors," *Physical Review*, Vol. 99 (September, 1955), pp. 1892–1893.
[110] Redfield, D. "Recombination Processes in Tellurium," *Physical Review*, Vol. 100 (November, 1955), pp. 1094–1100.
[111] Bardeen, J. "Surface States and Rectification at a Metal Semiconductor Contact," *Physical Review*, Vol. 71 (May, 1947), pp. 717–727.
[112] Kingston, R. H. "Review of Germanium Surface Phenomena," *Journal of Applied Physics*, Vol. 27 (February, 1956), pp. 101–114.
[113] Kingston, R. H., ed. *Semiconductor Surface Physics*. University of Pennsylvania Press, 1957.
[114] Brattain, W. H. and Bardeen, J. "Surface Properties of Germanium," *Bell System Technical Journal*, Vol. 32 (January, 1953), pp. 1–41.
[115] Buck, T. M. and McKim, F. S. "Depth of Surface Damage Due to Abrasion on Germanium," *Journal of the Electrochemical Society*, Vol. 103 (October, 1956), pp. 593–597.
[116] Holonyak, N., Jr. and Letaw, H., Jr. "Sparked Hydrogen Treatment of Germanium Surfaces," *Journal of Applied Physics*, Vol. 26 (March, 1955), p. 355.
[117] Garrett, C. G. B. and Brattain, W. H. "Distribution and Cross-Sections of Fast States on Germanium Surfaces," *Bell System Technical Journal*, Vol. 35 (September, 1956), pp. 1041–1058.
[118] Lawrance, R. "The Temperature Dependence of the Drift Mobility of Injected Holes in Germanium," *Proceedings of the Physical Society*, Vol. 67B (January, 1954), pp. 18–27.
[119] Gorton, R. "Effects of Irradiation upon Diodes of the Silicon Junction Type," *Nature*, Vol. 179 (May, 1957), p. 864.
[120] Miller, W., Bemig, K., and Salzberg, B. "Note on the Reduction of Carrier Lifetime in *p-n* Junction Diodes by Electron Bombardment," *Journal of Applied Physics*, Vol. 27 (December, 1956), pp. 1524–1527.
[121] Logan, R. A. and Schwartz, M. "Thermal Effects on Lifetime of Minority Carriers in Germanium," *Physical Review*, Vol. 96 (October, 1954), pp. 46–47.
[122] Logan, R. A. and Schwartz, M. "Restoration of Resistivity and Lifetime in Heat Treated Germanium," *Journal of Applied Physics*, Vol. 26 (December, 1955), pp. 1287–1289.
[123] Bemski, G. "Quenched-in Recombination Centers in Silicon," *Physical Review*, Vol. 103 (August, 1956), pp. 567–569.
[124] Theuerer, H. C., Whalen, J. M., Bridgers, H. E., Jr., and Buehler, E. "Heat Treatment of Silicon Using Zone Heating Techniques," *Journal of the Electrochemical Society*, Vol. 104 (December, 1957), pp. 721–723.
[125] Silverman, S. J. and Singleton, J. Presented at the meeting of the Electrochemical Society, Buffalo, N. Y., 1957.
[126] Nijland, L. M. and van de Pauw, L. J. "The Effect of Heat Treatment on the Bulk Lifetime of Excess Charge Carriers in Silicon," *Journal of Electronics and Control*, Vol. 3 (October, 1957), pp. 391–395.
[127] Sandiford, D. J. and Shields, J. *Reverse Currents and Carrier Lifetimes in p-n Junctions*. Ashorne Hill: Report on Semiconductor Conference, 1956.
[128] Thurmond, C. D. and Logan, R. A. "Copper Distribution between Germanium Ternary Melts Saturated with Germanium," *Journal of Physical Chemistry*, Vol. 60 (May, 1956), pp. 591–594.
[129] Tyler, W. W. and Woodbury, H. H. "Transient Photo-Conductivity Measurements in Germanium Containing Double Acceptor Impurities," *Bulletin of the American Physical Society*, Vol. 1 (March, 1956), p. 127.
[130] Johnson, L. F. "Photoconductivity in Gold Doped Germanium," *Bulletin of the American Physical Society*, Vol. 1 (March, 1956), p. 127.
[131] Newman R. Private communication.

Noise in Semiconductors and Photoconductors*

K. M. VAN VLIET†

Summary—A survey is given of theory and experiments on noise in bulk semiconductors and photoconductors. This paper is divided into four parts, including generation-recombination (gr) noise in semiconductors, gr noise in photoconductors, 1/f noise in single crystals, and modulation noise in granular materials. In the first part an account is given of the appropriate analyses and the results are applied to extrinsic as well as intrinsic fluctuations, generated either in the bulk or at the surface. In the part about photoconductors the limiting sensitivity caused by photon noise is calculated and present infrared detectors are discussed. Next, a survey is given about present understanding of 1/f noise, and of its relation to the field effect as proposed by McWhorter and others. Finally, some remarks are made about 1/f noise in granular material and the proposed theories are briefly reviewed.

I. Introduction

In 1932 Williams and Thatcher [107] observed that current carrying carbon resistors generated a large amount of noise. When the current was absent, on the contrary, the noise satisfied Nyquist's formula, according to which the noise of any resistive device in thermal equilibrium equals $4kTR$ in unit bandwidth. The first extensive investigation of noise in current-carrying nonmetallic resistors was made by Bernamont in 1934 [5]. He found that the spectral density of the noise was approximately inversely proportional to the frequency in the audio-frequency range. In that respect the spectrum looked similar to the flicker effect in vacuum tubes discovered nine years before by Johnson [44]. Surdin [88] extended the frequency range of Bernamont's data and observed this noise up to a few mc beyond which thermal noise usually predominated. Only rarely a change from a 1/f spectrum into a $1/f^2$ spectrum was observed.

The first theory for noise in carbon and evaporated metal layer resistors was also developed by Bernamont. He predicted a spectrum of the form constant/$(1+\omega^2\tau^2)$ where τ is the time constant of the fluctuations. In order to fit this result to the experimental data, Bernamont [5] and Surdin [87] suggested that there should be a distribution of time constants, ranging from $\sim$1 sec up to $\sim$1 μsec. However, as is known today, Bernamont's theory does not apply to carbon resistors but to other noise processes (see Section II-A).

The cause of the noise in carbon and similar resistors

* Original manuscript received by the IRE, March 3, 1958.
† Dept. Elec. Eng., University of Minnesota, Minneapolis, Minn.

became clearer from the experiments of Christenson and Pearson [27] in 1936. They could locate the source of the noise in the contacts between the grains in these materials and introduced the name contact noise. Their observations were verified by many other research workers, for a variety of materials. Shortly after World War II, when single crystals of semiconducting materials became available, the hope was raised that this noise would be absent. It was found, however, that again noise of a 1/f nature was present; a detailed investigation on germanium single crystals was reported by Montgomery [60], [81]. Montgomery could show that the noise was associated with the surface of the crystals. This noise apparently was no contact noise and was generally named excess noise.

Deviations from the 1/f dependence in germanium were first reported by Herzog and van der Ziel [39]. The spectrum showed a characteristic time constant of $\sim$1 μsec which later on was identified with the minority carrier lifetime. The noise was thus attributed to the random excitation and capture of free carriers, which during their stay in the conduction band (or valence band) give rise to a current pulse in the output circuit. Because of the resemblance to the random emission of electrons in vacuum tubes this noise was called shot noise. The effect had been predicted before by Gisolf [36] in 1949. His theory, being in error, was modified by van der Ziel [93], [95], who also showed that Bernamont's original theory was equivalent to the modified Gisolf theory [92], [95]. There are two reasons for such a late discovery of this noise. First, as mentioned above, 1/f noise depends on the surface conditions of the crystals. Therefore, 1/f noise masked all other effects until considerable progress had been made in various laboratories in improving the surface properties. Secondly, it is known that 1/f noise is roughly independent of temperature in contrast to the spontaneous fluctuations in excitation and capture of carriers which give appreciable noise in certain temperature regions only.

In photoconductors similar noise effects have been reported. Here the facts are more complicated because of the fluctuations in the incident radiation field (photon noise). Completely contradictory opinions have been stated about the effect of photon noise on the performance of photoconductors. Buttler [24], [25] assumes that the effect is negligible. Shulman [84], on the contrary, states that under ideal conditions all the noise results from photon fluctuations. The author holds the opinion that photon fluctuations can account for up to 50 per cent of the observed noise in an ideal trap free photoconductor (see Section III-B).

In the above survey, several noises have been introduced in a more or less historical order. Before proceeding, it seems appropriate to introduce a terminology which better agrees with present understanding of the various processes. Noise above thermal noise will generally be called *current noise*. This will be divided into two main areas.

A. Generation-Recombination Noise

This is henceforth denoted as gr noise. This noise is caused by spontaneous fluctuations in the generation rates, recombination rates, trapping rates, etc., thus producing fluctuations in the free carrier densities. The term gr noise seems more adequate than shot noise since these carrier fluctuations exist even in equilibrium when there is no applied field. There is some difference in treatment of gr noise in semiconductors and in photoconductors. In semiconductors the noise is of thermal origin and can therefore be calculated with the generalized Nyquist formula (Section II-C). In photoconductors only statistical arguments can be applied.

B. Modulation Noise

This term introduced by Petritz [70] will refer to noise which is not *directly* caused by fluctuations of the carrier transition rates, but instead is due to carrier density fluctuations or current fluctuations caused by some modulating effect. Let us give some examples. According to some investigators 1/f noise in germanium filaments is primarily caused by fluctuations in the occupancy of the "slow surface states" (Section IV-B). When a carrier is trapped in these states it produces a change in the number of carriers and as such is gr noise. However, this is not the full story. Owing to the change in surface charge the surface recombination velocity will change. The bulk conductivity in turn is affected more severely by this effect than just by the gain or loss of one carrier. Thus, the "slow surface states" cause conductivity modulation of the bulk. As another example we turn to some mechanisms proposed for contact noise (Section V-A). Suppose that the current between two grains is affected by molecules diffusing over the contact area. This will produce current modulation. Most of these modulation effects found in practice are of the $1/f^{\alpha}$ type (with $\alpha \approx 1$) and most of the theories fail to give a 1/f dependence over a long frequency range. Some modulation noises have been observed which do not have a 1/f spectrum (*e.g.*, heat conductivity induced fluctuations) [99], [4]. Although more effects of such a nature may be found in the future, at the present time the name modulation noise may be considered a synonym with "1/f noise."

As illustrated by the two previous examples it may well be that 1/f noise in single crystals and in granular materials is of quite different origin. Nothing definite is known about this today, however.

In accordance with the above classification and comments this paper will be divided into four parts. In Section II we discuss theory and experiments of gr noise in semiconductors. In Section III we consider gr noise in photoconductors and the necessarily related topic of photon noise. In Section IV, 1/f noise in single crystals is discussed and in Section V we mention very briefly some modulation effects in granular material. The order of the topics indicates, apart from the author's preference, the order of decreasing understanding and increasing need for future research.

II. Generation-Recombination Noise in Semiconductors

A. Older Theories

Gisolf [36] considered the current pulses caused by the individual carriers during their lifetime. For simplicity it was assumed that only one type of carrier participated in the current. The current pulse caused by a carrier with lifetime τ_p is

$$\left.\begin{aligned} F_p(t) &= e\mu EL^{-1} \qquad t_0 \leq t \leq t_0 + \tau_p \\ F_p(t) &= 0 \qquad \text{elsewhere,} \end{aligned}\right\} \tag{1}$$

where E is the electric field, L the electrode distance, and μ the mobility. The total current can be written as

$$i(t) = \Sigma_p \Sigma_r F_p(t - t_r).$$

Assuming that the elementary pulses are independent and occur at random instants t_r with a rate $\overline{N}_p$, we can apply Campbell's theorem [34] for the fluctuations $\Delta i(t) = i(t) - \langle (i) \rangle$

$$\langle \Delta i(t)^2 \rangle = \Sigma_p \overline{N}_p \int_{-\infty}^{\infty} | F_p(\xi)^2 | \, d\xi. \tag{2}$$

The next step is to make a Fourier integral analysis of $F_p(\xi)$ and apply Parseval's theorem. This leads to the generalized Carson's theorem:

$$S_i(f) = 2\Sigma_p \overline{N}_p | A_p(f)^2 | \tag{3}$$

where $A_p(f)$ is the Fourier coefficient of $F_p(\xi)$ and $S_i(f)$ is the spectral intensity of the fluctuations defined by

$$\langle \Delta i(t)^2 \rangle = \int_0^{\infty} S_i(f) df. \tag{4}$$

Assuming a statistical weight factor for τ_p [95], [10] and carrying out the above procedure one is led to the simple result

$$S_i(f) = 4I^2\tau / n_0(1 + \omega^2\tau^2) \tag{5}$$

which was first found by Bernamont [5]. Here n_0 is the total steady-state number of carriers in the sample.

The above theory, interesting for historical reasons and for its simplicity, is incorrect since the elementary pulses are not independent. The possible transitions are limited by the Pauli exclusion principle. A better result was obtained by van der Ziel [93] and by Machlup [52], who suggested that the current carriers were subject to binomial statistics. Let λ be the probability that a carrier is free and $1-\lambda$ the probability that a carrier is bound. Then the noise is found to be

$$S_i(f) = 4(1 - \lambda) I^2\tau / n_0(1 + \omega^2\tau^2). \tag{6}$$

Obviously, for $\lambda \approx 1$ the noise is negligible. This is the case for extrinsic germanium and silicon at room temperature when all the donors are ionized or all the acceptors filled. Although (6) gives a good approximation, it was first shown by Burgess [18], [21] that an improvement could be obtained.

B. The Langevin Equations [100]

The relation between the instantaneous current and the instantaneous number of carriers is

$$i(t) = e\mu_n n(t) EL^{-1} + e\mu_p p(t) EL^{-1}. \tag{7}$$

Consequently, denoting by S_{nn} the spectrum of $\langle \Delta n^2 \rangle$, by S_{pp} the spectrum of $\langle \Delta p^2 \rangle$ and by S_{np} the spectrum of $\langle \Delta n \Delta p \rangle$, which quantities are defined analogous to (4), we have

$$S_i = [I/(bn_0 + p_0)]^2 (b^2 S_{nn} + 2bS_{np} + S_{pp}). \tag{8}$$

Note that n_0 and p_0 are not densities but mean *total numbers; I* is the dc current and $b = \mu_n/\mu_p$. In the particular case that $p_0 = 0$ we have

$$S_i = (I/n_0)^2 S_{nn} \tag{8a}$$

and in another practical case where $\Delta n(t) = \Delta p(t)$ we have

$$S_i = I^2[(b + 1)/(bn_0 + p_0)]^2 S_{nn}. \tag{8b}$$

In accordance with what has been said in the introduction, it will suffice to find S_{nn}, S_{np}, and S_{pp}. The current noise spectrum S_i follows from the above relations.

First we will consider "single step processes," *i.e.*, transitions occur only between the conduction band (containing n electrons) and localized levels, or between the conduction band and the valence band. Noise in a p-type semiconductor involving transitions between the valence band and localized levels can be found *mutatis mutandis*. Let $g(n)$ be the generation rate of electrons and $r(n)$ the recombination rate with either holes in the localized levels or with holes in the valence band. It will be assumed that the mass action law applies to the rates $g(n)$ and $r(n)$. The effect of different spins will be neglected since the variances and covariances are not affected by this refinement in statistics. Hence, the kinetic equation is

$$dn/dt = g(n) - r(n) + f(t) \tag{9}$$

where $f(t)$ is a stochastic source function for the fluctuations. Expanding $g(n)$ and $r(n)$ up to first-order terms in $n - n_0 = \Delta n$, we obtain the Langevin equation

$$d(\Delta n)/dt = -\Delta n/\tau + f(t) \tag{10}$$

where τ is the lifetime of added current carriers

$$\tau = \left\{ \left(\frac{dr}{dn} \right)_0 - \left(\frac{dg}{dn} \right)_0 \right\}^{-1}; \tag{11}$$

the subscript zero refers to the values of the derivatives for $n = n_0 = \langle n \rangle$. The solution of (10) yields:

$$\langle \Delta n(t + s) \Delta n(t) \rangle = \langle \Delta n(t)^2 \rangle e^{-s/\tau}. \tag{12}$$

With the Wiener-Khintchine theorem we find

$$\begin{aligned} S_{nn}(f) &= 4 \int_0^{\infty} \langle \Delta n(t + s) \Delta n(t) \rangle \cos \omega s ds \\ &= 4 \langle (\Delta n)^2 \rangle \tau / (1 + \omega^2\tau^2). \end{aligned} \tag{13}$$

The method does not give a complete answer unless $\langle(\Delta n)^2\rangle$ is specified. Eq. (9) suggests the following relation between the statistical properties of $g(n)$ and $r(n)$

$$W(n)r(n) - W(n-1)g(n-1) = 0, \qquad (14)$$

where $W(n)$ is the probability distribution of n. Since (14) is a recurrent relation, $W(n)$ and $\langle(\Delta n)^2\rangle$ can be found as was shown by Burgess [18], [21]; the result is the gr theorem:

$$\langle(\Delta n)^2\rangle = g_0\left\{\left(\frac{dr}{dn}\right)_0 - \left(\frac{dg}{dn}\right)_0\right\}^{-1} = g_0\tau. \qquad (15)$$

Hence, the spectrum (13) can be written as

$$S_{nn}(f) = 4g_0\tau^2/(1+\omega^2\tau^2). \qquad (16)$$

The Langevin method can easily be extended to multistep processes [100]. Let there be s energy levels (including the conduction band and the valence band) between which transitions occur. The electron concentrations in these levels will be denoted by $n_1 \cdots n_s$. Since $\Sigma_j n_j$ is constant one variable can be eliminated. Let the independent variables then be $n_1 \cdots n_{s-1}$. If p_{ij} is the transition rate per second from i to j then

$$dn_i/dt = \sum_{j=1}^{s}{}' p_{ji} - \sum_{j=1}^{s}{}' p_{ij} + f_i(t). \qquad (17)$$

The Langevin equations follow from expansion of p_{ji} up to first-order terms in the fluctuations. They can be written as

$$d(\Delta n_i)/dt = \sum_{j=1}^{s-1} a_{ij}\Delta n_j + f_i(t) \qquad (18)$$

where

$$a_{ij} = \sum_{k=1}^{s}{}'\left\{\left(\frac{\partial p_{ki}}{\partial n_j}\right)_0 - \left(\frac{\partial p_{ik}}{\partial n_j}\right)\right\}. \qquad (19)$$

Eq. (18) can be formally solved, given the values of $\Delta n_i(t)$ at some time t_0. It is then found that the correlation matrix has the following form:

$$\langle\Delta n_i(t+s)\Delta n_j(t)\rangle = \sum_{k=1}^{s-1} c_k{}^{ij}e^{-t/\tau_k} \qquad (20)$$

where the quantities $-1/\tau_k$ are the eigenvalues of the matrix $[a_{ij}]$. The spectrum follows from the Wiener-Khintchine theorem:

$$S_{ij}(f) = 4\sum_{k=1}^{s-1} c_k{}^{ij}\tau_k/(1+\omega^2\tau_k{}^2). \qquad (21)$$

The general form is depicted in Fig. 1. The quantities $c_k{}^{ij}$ depend on the variances $\langle\Delta n_i{}^2\rangle$ and the covariances $\langle\Delta n_i\Delta n_j\rangle$ which have again to be found by other means. Van Vliet and Blok [100] calculated the variances and the covariances from the Fokker Planck equation and generalized the gr theorem (15). The spectrum (21) is therefore completely known once the transition rates p_{ij} between the various energy levels can be specified according to the mass action law. A detailed example employing this method was given for noise in cadmium sulphide [101]. The calculations are cumbersome, however, and comparison with experiment is not readily achieved. In semiconductors the situation is slightly better than in photoconductors since an alternate method based on thermodynamics can be applied (see below).

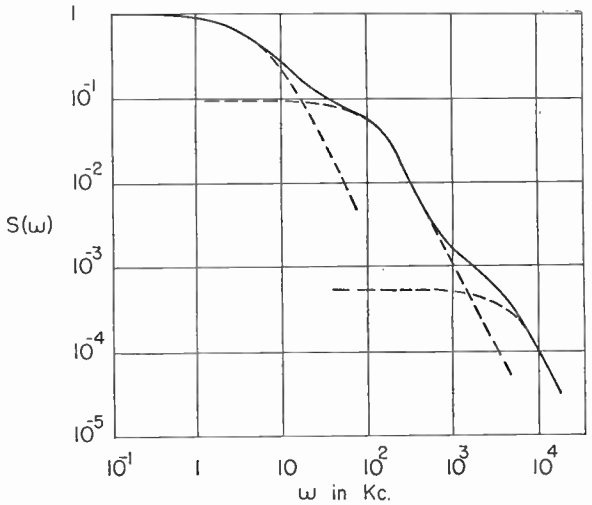


Fig. 1—Illustration of possible spectra in the case of multistep processes or several simultaneous single-step processes. The following values are chosen: $\tau_1 = 2\times10^{-4}$, $\tau_2 = 10^{-6}$; $\tau_3 = 2\times10^{-7}$ sec and $c_1 = 5\times10^3$, $c_2 = 10^4$, $c_3 = 3\times10^3$; these numbers are relative values.

C. Application of the Generalized Nyquist Formula

It has been emphasized already that it is sufficient to calculate the spectral densities S_{nn} under the conditions that there is no electric field. For moderate electric fields S_{nn}, etc., will not change and (8) can be applied to find the current noise. Consequently for the calculation of S_{nn}, etc., we may assume that the crystal is in thermal equilibrium. Obviously, we restrict ourselves here to semiconductors.

As is well known, the steady-state electron distribution in semiconductors can be obtained thermodynamically (minimization of the electronic free energy) and from mass action law considerations. In accordance with this, two methods which are entirely different in principle can be employed to find the noise. This was shown by Burgess [19] for the variances $\langle\Delta n^2\rangle$, etc. Recently, the author [104] proved that the noise spectra S_{nn}, etc., can also be found from thermodynamic arguments; the results are in complete agreement with the previous section in which the mass action law was used. The method is based on the generalized Nyquist formula, established first by Callen and Welton [26] in 1951. According to this relation the spectral density of a single fluctuating extensive thermodynamic variable can be expressed as

$$S(f) = 4kT\omega^{-2}\,\mathrm{Re}\,(Y). \qquad (22)$$

The admittance Y relates a sinusoidal variation of the

external variable to a sinusoidal perturbation by a generalized force acting on the system. In semiconductors such a perturbation can be accomplished if the Fermi level is changed into two or more quasi-Fermi levels for the various carrier groups. The response of the transition rates to such a perturbation defines the admittance Y. If only two carrier groups are considered, *e.g.*, electrons and bound or free holes, the admittance is found to be [104]

$$Y = \frac{g_0}{kT} \cdot \frac{j\omega\tau}{1 + j\omega\tau} . \tag{23}$$

This equation and (22) immediately yield the result (16).

The method has also been applied to more variable cases.

D. Extrinsic Semiconductors

With the framework developed in the previous sections the noise in extrinsic semiconductors is readily found. We will assume that the semiconductor is n-type and has N_d donors all at a single level E_d below the conduction band. It will first be assumed that no intrinsic transitions occur. The generation rate and recombination rate can then be expressed as follows:

$$g(n) = \gamma(N_d - n); \qquad r(n) = \delta n^2 \tag{24}$$

and in equilibrium we have

$$g_0 = \gamma(N_d - n_0) = \delta n_0^2 = r_0. \tag{25}$$

The quantities γ and δ are transition constants which include cross sections, etc. The constant δ depends only weakly on the temperature; γ, on the contrary, depends on T according to a Boltzmann factor $\gamma = \gamma_0 \exp(-E_d/kT)$. Substituting (24) and (25) into (11) and (15) yields

$$\langle \Delta n^2 \rangle = \frac{n_0(N_d - n_0)}{2N_d - n_0} = N_d \frac{\lambda(1 - \lambda)}{2 - \lambda} \tag{26}$$

$$\tau = \frac{1}{\delta n_0} \cdot \frac{N_d - n_0}{2N_d - n_0} = \frac{1}{\delta N_d} \cdot \frac{1 - \lambda}{\lambda(2 - \lambda)} \tag{27}$$

where $\lambda = n_0/N_d$ is the fractional ionization of the impurities and has the same meaning as in Section II-A. The noise spectrum is found from (8a), (16), (26), and (27) which give after some arrangement:

$$S_i(f) = 4\left(\frac{I^2}{n_0}\right)\left(\frac{1 - \lambda}{2 - \lambda}\right)\left(\frac{\tau}{1 + \omega^2\tau^2}\right). \tag{28}$$

The conclusions drawn in Section II-A that the noise occurs only if the fraction λ of ionized donors is low is still valid. This explains why hardly any noise measurements on strongly extrinsic material have been reported. A characteristic feature of this noise is the strong temperature dependence of τ which is the majority carrier lifetime. According to (27) τ decreases rapidly with increasing temperature since λ approaches 1. Noise of this nature was reported by Gebbie [35]. However, Gebbie attributed the noise to trapping effects.

In the above analysis it was assumed that the donor levels are filled at absolute zero. If the donor levels contain already m_0 holes at $T = 0$, the properties of the semiconductor are markedly different. The Fermi level, instead of lying approximately midway between the conduction band and the donor levels, is now clamped at the position of the donor levels for sufficiently low temperatures such that $n_0 \ll m_0$ [66]. The noise properties are also different, since the transition rates, instead of (24), are now given as

$$\left.\begin{aligned} g(n) &= \gamma(N_d - m_0 - n_0) \approx \gamma(N_d - m_0) \\ r(n) &= \delta n_0(m_0 + n_0) \approx \delta n_0 m_0 \end{aligned}\right\}. \tag{29}$$

The reaction mechanism is now approximately monomolecular. One finds from (29)

$$\langle \Delta n^2 \rangle \approx n_0; \qquad \tau \approx 1/\delta m_0. \tag{30}$$

Instead of (28) the noise is now given by Bernamont's formula (5) as is easily shown from (30) and (16). Moreover, the lifetime τ is approximately constant. For p-type material similar arguments apply. Measurements were recently performed on manganese-doped germanium in the range 77°K–125°K by Fassett [31a] at the University of Minnesota. It was estimated that 50 per cent of the acceptor levels were filled at absolute zero. The spectra were flat from 1 kc–10 mc for temperatures ~120°K. For temperatures below 100°K the lifetime increased, in contrast to (30). The possibility that hole trapping is present or that excited states of the impurity centers play a role is being investigated.

Noise in cases where more donor levels participate in the transitions can also be calculated from the general analysis in the preceding sections. Also, an analysis can be given for the case that intrinsic transitions and impurity transitions occur simultaneously. Since most likely the donors are completely ionized before intrinsic transitions start to be present abundantly, the noise arising from the transitions to and from the donors or acceptors can usually be neglected in these cases. Hence, only the intrinsic transitions cause noise. Consequently, without treating the complete case, it is evident that the noise output of a semiconductor is a complicated function of the temperature. Appreciable noise is found only in the strongly extrinsic and in the near-intrinsic or intrinsic region.

E. Near Intrinsic and Intrinsic Crystals

In these crystals the transition rates are, neglecting first recombination centers:

$$g(n) = \alpha; \qquad r(n) = \beta n^2. \tag{31}$$

The noise can be found in the same way as before. The relation between the density fluctuations and the current fluctuations appropriate to this case is (8b), since the electron and hole densities fluctuate coherently; *i.e.*,

$\Delta n(t) = \Delta p(t)$. The following result is easily obtained in the same manner as in the previous section

$$S_i(f) = 4I^2 \frac{(b+1)^2 n_0 p_0}{(bn_0 + p_0)^2 (n_0 + p_0)} \cdot \frac{\tau}{1 + \omega^2\tau^2}. \quad (32)$$

For intrinsic semiconductors ($n_0 = p_0$) this reduces to

$$S_i(f) = 2I^2\tau / n_0(1 + \omega^2\tau^2). \quad (33)$$

In contrast to the above assumptions usually recombination in germanium and silicon occurs via recombination centers. This changes the problem into a two-variable problem, in which two of the three variables n, p, n_t (where n_t is the number of electrons in the recombination centers) can be taken as the independent ones. The variances $\langle \Delta n^2 \rangle$, $\langle \Delta p \rangle^2$ and the covariance $\langle \Delta n \Delta p \rangle$ were calculated by Burgess [19] and by van Vliet [103]. If, however, the number of electrons in the recombination centers is small compared to the number of majority carriers, we have $\Delta n \approx \Delta p$. Then (32) and (33) apply where τ is the Shockley-Read lifetime [83], [104].

Noise in intrinsic or nearly intrinsic material has been observed by several investigators [39], [55], [94], [79], [43], [40]. In the last two references the noise is compared with (32) and (33) and good agreement is found in most cases. In some cases, the spectrum did not fall off as rapidly as $1/\omega^2$ at high frequencies, indicating that the recombination centers in the bulk or at the surface (see below) are slightly distributed in energy.

F. Surface GR Noise

In many germanium crystals generation and recombination of electrons and holes involve surface centers. This complicates the theory considerably. However, some simplifying argument will be given. First of all, it is noted that the noise is not determined by the local number of carriers. Secondly, as for bulk recombination centers, we will assume that the population of the surface states is not too large. Then the decay of fluctuations can again be described with (10) where τ is a surface lifetime. Although $r(n)$ and $g(n)$ are not specified, the variances $\langle \Delta n^2 \rangle$, $\langle \Delta p^2 \rangle$, and $\langle \Delta n \Delta p \rangle$ will be the same since these expressions can be found from thermodynamic arguments. Consequently, (31) and (32) will be approximately valid. Hyde [43] has pointed out that the spectra may reveal some smaller lifetimes since the solution of the continuity equations for injected carriers give rise to higher order modes [82]. It is also possible to solve the stochastic partial differential equations governing the generation, recombination, and diffusion of carriers exactly, but this would be beyond the scope of this paper.

G. Influence of the Drift of the Carriers on the Noise

In the previous section it was assumed that the probability distribution $W(n)$ for the carriers was not distorted by the presence of the field. This is the case up to very high fields, apart from heat dissipation effects. At much lower field strengths, however, the spectral distribution of the noise may change since carriers are swept to the electrodes of the sample in a time smaller than τ. Let us first consider a semiconductor containing only one type of carriers, say electrons. What will happen if $\tau \gtrsim \tau_d$ where $\tau_d = L/\mu_n E$? Davydov and Gurevich [29] first treated the problem. Their work implies the correlation function

$$\langle \Delta n(t) \Delta n(t+s) \rangle = \langle \Delta n^2 \rangle e^{-s/\tau}(1 - s/\tau_d) \quad (34)$$

for $s \leq \tau_d$ and zero for $s > \tau_d$. The term $(1 - s/\tau_d)$ takes into account that carriers are swept out of the sample in a time τ_d. It has been pointed out before [100], [84], [11], [77] however, that (34) is erroneous if the dielectric relaxation time of the sample is much shorter than τ_d. In that case carriers are not simply swept out at the contacts and their life terminated. As soon as some extra electrons Δn are generated at a spot x and begin to move under the influence of the applied field, electrons start to move everywhere in the crystal; at the negative electrode, electrons flow into the crystal and space charge neutrality is preserved. Hence, a carrier lives virtually its full lifetime τ and the noise is the same as for $\tau \ll \tau_d$. If both electrons and holes are present, things may change, however [62]. Consider, *e.g.*, an n-type sample in which some extra hole-electron-pairs are created somewhere in the middle between the two electrodes. The electrons, starting to move to the positive electrode cannot produce a space charge so that other electrons come in at the negative contact whereas electrons are carried off at the positive electrode. If the holes start to move to the negative electrode they cannot produce a space charge either; their charge will be neutralized mainly by electrons flowing in the opposite direction. Therefore, virtually a hole and a neutralizing electron are flowing to the negative electrode with the so-called ambipolar mobility [98], given by

$$\mu_a = \frac{|n_0 - p_0| \mu_n \mu_p}{\mu_n n_0 + \mu_p p_0}; \qquad \tau_a = \frac{L}{\mu_a E}. \quad (35)$$

The ambipolar mobility is equal to the minority carrier mobility in highly extrinsic material in accordance with the above picture. The correlation function is now given by (34) if we replace τ_d by the ambipolar drift time τ_a. The spectrum can be calculated with the Wiener-Khintchine theorem; the integration involved is found in a paper by Burgess [20]. For $\tau \gg \tau_a$ the spectrum becomes oscillatory

$$S_i(f) = \text{constant} \times \sin^2(\tfrac{1}{2}\omega\tau_a)/(\tfrac{1}{2}\omega\tau_a)^2. \quad (36)$$

The effect has been experimentally verified by Hill. His measurements and an exact theory will be published elsewhere [105].

H. Degenerate Semiconductors

The analysis given in Section II-B is not immediately applicable to degenerate semiconductors like InSb. A more careful averaging process over the electrons in a

degenerate band has to be carried out than is implied by the simple terms $g(n)$ and $r(n)$ in previous sections. The variances $\langle \Delta n^2 \rangle$ and $\langle \Delta p^2 \rangle$ for hole electron pair transitions were calculated by Oliver [68] and by Burgess [22]. The spectrum was recently derived with the approach of irreversible thermodynamics (Section II-C) by the author [104]. It was shown that (16) is still valid. However, g_0 and τ depend on the carrier concentrations in a more complicated way than before. The noise is found to be given by [instead of (32)]:

$$S_i(f) = 4I^2 \frac{(b+1)^2 \xi_n \xi_p n_0 p_0}{(bn_0 + p_0)^2 (\xi_n n_0 + \xi_p p_0)} \cdot \frac{\tau}{1 + \omega^2 \tau^2}; \quad (37)$$

here ξ_n and ξ_p depend on the position of the Fermi level of the degenerate electron and hole distributions. For complete degeneracy of an n-type sample $\xi_n = 3kT/2\epsilon_F$ and $\xi_p = 1$, where ϵ_F is the position of the Fermi level with respect to the bottom of the conduction band. The result (37) has not as yet been verified experimentally since other noise sources dominate in InSb [69], [86].

III. GR Noise in Photoconductors

A. Circuitry and Figures of Merit

A photoconductor is usually incorporated in a simple circuit, consisting of the photoconductive film or crystal, a load resistor, and a battery. The signal can be taken either from the crystal terminals or from the load resistor. If the signal and the rms noise both result from conductivity changes, both these effects are proportional to the current and the signal-to-noise ratio does not depend on the load resistor, providing the thermal noise is negligible. There is, however, another reason to make the load resistor large. Often, 1/f noise of the current carrying contacts is a dominating noise effect. This can be suppressed by measuring the signal between probes on the crystal. If the load resistor is so large that the supply acts as a constant current generator, then resistance fluctuations at the current contacts will not show up as noise between the probes, unless concentration disturbances are swept into the probe region. Therefore, if T_p is the lifetime and $E\mu_p$ the drift velocity of the minority carriers, and d the distance between probes and the adjacent end contacts, we should have $\mu_p E \tau_p \ll d$ (compare Fig. 2).

Since, assuming that no contact noise is present, the signal-to-noise ratio does not depend on the circuitry, we will find it convenient to represent both signal and noise by short-circuited current generators.

We will assume that the light signal is being shopped with an angular frequency $\omega = 2\pi f$ and that the signal passes through an amplifier, with bandwidth $(f - \frac{1}{2}\Delta f, f + \frac{1}{2}\Delta f)$. The signal will be denoted by $i(\omega)$ and the rms noise in this bandwidth by $\sqrt{S_i(\omega)\Delta f}$. The signal-to-noise ratio is then

$$\frac{\text{signal}}{\text{noise}} = \frac{i(\omega)}{\sqrt{S_i(\omega)\Delta f}}. \quad (38)$$

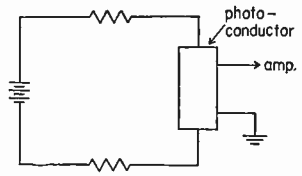


Fig. 2—Probe method for photoconductive cells.

With Jones [46] we will define the responsivity as the signal per unit radiation input. The dimension of the signal has already been fixed (amps). Since photoconductors are quantum detectors we will express the input radiation intensity on the whole area of the cell[1] by J_s. Hence the responsivity is expressed in amps/quanta sec^{-1}:

$$i(\omega) = J_s(\omega)\,\mathfrak{R}(\omega). \quad (39)$$

It should be noted that $\mathfrak{R}$ also depends on the optical frequency of the light signal which will be denoted by ν. The noise equivalent radiation intensity P_{eq} is defined as

$$P_{eq} = [S_i(\omega)\Delta f]^{1/2}/\mathfrak{R}(\omega). \quad (40)$$

As expected, the signal-to-noise ratio (38) is unity when $J_s = P_{eq}$. The noise equivalent power in watt is simply $h\nu P_{eq}$.

$$P_{eq}' = h\nu P_{eq}. \quad (41)$$

It will turn out (Section III-C) that P_{eq} is proportional to $A^{1/2}(\Delta f)^{1/2}$. Therefore, to compare cells it is meaningful to evaluate P_{eq} for $A = 1$ cm^2 and $\Delta f = 1$ cycle per second. The inverse of P_{eq} under these conditions is called the specific sensitivity [46], [74] and can be seen as a figure of merit. In simple cases the signal and noise depend on ω in the same way and P_{eq} does not depend on the response time τ. In other cases, however, P_{eq} does depend on ω and thus on τ. (This is always the case if 1/f noise predominates.) Figures of merit for special reference conditions, including the dependence on τ have been given by Jones in his survey paper on the performance of radiation detectors [46]. If the signal-to-noise ratio is smaller than unity, τ determines the information rate. Petritz [73] and MacQuiston [53] have discussed the performance using an information theory approach.

The value of P_{eq} determined from the observed noise and the measured response according to (40) is not necessarily the limit that can be reached with a particular photoconductive material at a particular temperature. First of all, the detector should not be limited by 1/f noise but by gr noise. This is the case for several

[1] Since we are dealing with fluctuations we get into trouble when we work with radiation per unit area or with carrier densities. *E.g.*, the simple relation $\langle \Delta n^2 \rangle \simeq \langle n \rangle$ gives dimension inconsistencies when applied to densities. Although unorthodox with respect to other fields, our quantities J_s and J_r will refer to intensities on a given cell area A, and likewise our quantities n, p refer to total numbers of electrons and holes in a given cell volume $V = AD'$ in which the absorbed radiation is supposed to be approximately uniform. The conversion of the the results to the usual units is straightforward and left to the reader.

materials today (see Section III-C). Furthermore, as has been especially emphasized by Petritz [73], in most of the present materials fluctuations in the carrier densities are caused by interaction with lattice vibrations (phonons) and not by interaction with the background black body radiation field (photons). The limiting sensitivity is reached when all carrier excitations involve photon absorption. The noise is then closely related to photon fluctuations of the background radiation. Photon noise will be defined here as the fluctuations in background radiation intensity J_r *incident* on the detector and is indicated by $S_{J_r}(f)$. Since the effect is caused by the arrival of the individual photons, we have for the fluctuations in any part ϵ of the stream:

$$S_{\epsilon J_r} = \epsilon S_{J_r}. \tag{42}$$

In a photocell of area A the fraction of the incident intensity effective in producing free carriers, will be denoted by ηJ. The quantum efficiency η is related to the absorption coefficient α and the reflection coefficient R. However, each absorbed quantum need not be effective since part of the energy may be exchanged with phonons. (See also next section.) As an estimate, we may put, however

$$\eta(\nu) \sim |1 - R(\nu)|\,|1 - \exp(-\alpha(\nu)d)| \tag{43}$$

where d is the thickness of the layer. Usually the simplifying assumption is made[2] that the absorption is uniform over an equivalent layer $d' \simeq 1/\alpha$.

The noise in that part of the photon stream that is effective in excitation is found from (42), setting $\epsilon = \eta$. If η_s is the quantum efficiency at the signal frequency then the number of excitations per second due to J_s is $\eta_s J_s$; when this number is equal to twice the rms fluctuations induced by the photon noise we will denote it by Q_{eq}. The factor 2 will be explained in the next section. According to (42) and the above reasoning we have

$$Q_{eq} = \left[2\int \eta(\nu) S_{J_r}(\nu)\,d\nu\,\Delta f\right]^{1/2} \Big/ \eta_s. \tag{44}$$

The integral is to be extended over all frequencies of the background radiation that is permitted to fall on the detector. A performance figure of a particular cell is

$$F = P_{eq}/Q_{eq}. \tag{45}$$

Jones calls this factor the noise figure. It is worthwhile to compare this figure with the noise figure in communication receivers. In that case the limiting input noise is the thermal noise of the transformed antenna resistance [96], which depends on the particular antenna and on its temperature. Likewise, the photon-induced noise with which the actual noise is compared according to (44) depends on the operation of the particular detector (its quantum efficiency, cutoff wavelength, and temperature). The situation is more complicated here, however. A value of F close to unity may indicate that the optimum performance under the particular operating conditions is attained but these conditions themselves may be rather unfavorable. As an example, let us assume that a lead sulphide cell cooled to liquid nitrogen temperature is seeing radiation noise of its room temperature surroundings. Noise figures close to unity have been reported experimentally under these conditions [32], [65]. A better performance could be obtained, however, by introducing a radiation shield with a transmission filter that only passes the signal wavelength, both of which are kept at low temperatures. The cell may then no longer be limited by photon noise, but the noise equivalent power has been reduced below the radiation limit under previous conditions. This will be worked out more quantitatively in Section III-C.

B. Interaction Between Photons and a Solid

Let the photoconductor exchange energy with a black-body radiation field of temperature T_r. We assume thermal equilibrium and also that all transitions involve photon emission or absorption. The density of photons $q(\nu)$ is given by Planck's law

$$q(\nu) = 8\pi(\nu^2/c^3)/(e^{h\nu/kT_r} - 1). \tag{46}$$

Let J_r be again the radiation intensity incident on the detector area A; then

$$J_r(\nu) = cAq(\nu)/4 = A2\pi(\nu^2/c^2)/(e^{h\nu/kT_r} - 1). \tag{47}$$

The spectrum of the photon noise is found with Bose-Einstein statistics [91], [50], [45], [101].

$$S_{J_r}(f) = 2J_r(\nu)[e^{h\nu/kT_r}/(e^{h\nu/kT_r} - 1)]. \tag{48}$$

For Q_{eq} we find from (48) and (44)

$$Q_{eq} = \left[4\int \frac{\eta(\nu)J_r(\nu)(\Delta f)\exp(h\nu/kT_r)d\nu}{\exp(h\nu/kT_r) - 1}\right]^{1/2} \cdot \frac{1}{\eta_s}. \tag{49}$$

The spectrum of the photon noise as expressed by (48) is white as for shot noise. Apparently, this stream is noisier than one would expect from a Poisson distribution since the Bose factor [ ] in (48) is larger than unity. It is remarkable that the noise of the effective photon stream $\eta_r J_r$ reflects these Bose-Einstein fluctuations [compare (42)] whereas the electrons follow Fermi statistics. Several authors have neglected these effects and assumed for simplicity that photons as well as electrons followed Poisson statistics. The consistency is then obvious [73], [77]. However, there is no discrepancy, as we will now show. To that purpose we will adopt Einstein's notation [31], [3], [22]. Let the radiation induce transitions between two levels $\mathcal{E}_1$ and $\mathcal{E}_2$ with

[2] If the signal and the background radiation are composed of different wavelength regions, the inhomogeniety cannot be neglected. This can be one of the reasons why the response time and the time τ from noise measurements are seldom in complete agreement.

carrier occupancy n' and n, respectively; then we may write[3]

$$\frac{dn}{dt} = n'J_rB_{12} - n[J_rB_{21} + A_{21}]. \quad (50)$$

The first term gives the photon induced excitations, the first term in the bracket is the "stimulated emission," and the last term is the spontaneous recombination rate. The quantum efficiency η is, expressed in this notation, the factor that relates the net excitation rate to J_r, hence

$$\eta = (n'B_{12} - nB_{21}). \quad (51)$$

We will identify $n'J_rB_{12}$ with $g^*(n)$ and $n(J_rB_{21}+A_{21})$ with $r^*(n)$ to conform with the notation of Section II-B; the asterisk is added to indicate that these quantities refer to photon induced excitation and to radiative recombination. From (50) we have:

$$\frac{d(\Delta n)}{dt} = \left(\frac{\partial g^*}{\partial n} - \frac{\partial r^*}{\partial n}\right)\Delta n + \left(\frac{\partial g^*}{\partial J_r} - \frac{\partial r^*}{\partial J_r}\right)\Delta J_r - \Delta F_{sp}$$
$$= -\Delta n/\tau^* + (n_0'B_{12} - n_0B_{21})\Delta J_r - \Delta F_{sp}. \quad (52)$$

The interpretation is obvious: the first term gives the regression of any disturbance to equilibrium, τ^* being the radiative lifetime; the second term gives the fluctuations of the induced net excitation rate F_q; and the last term accounts for the random fluctuations in the spontaneous recombination rate F_{sp}. The subscript zero refers again to average values. Note that this equation is of the general form (10), the origin of the random function $f(t)$ being specified in this case. For the noise in the net excitation rate we found already [(48), (42), and (51)]:

$$S_{\eta J_r} = 2(n_0'B_{12} - n_0B_{21})J_r(\nu)e^{h\nu/kT_r}/(e^{h\nu/kT_r} - 1). \quad (53)$$

Following Einstein's treatment the following relations are known

$$n_0/n_0' = K \exp(-h\nu/kT_r); \quad B_{12} = KB_{21}. \quad (54)$$

Hence one shows easily

$$S_{\eta J_r} = 2B_{12}n_0'J_0 \equiv 2g_0^*. \quad (55)$$

Likewise we find for the fluctuations in the spontaneous recombination rate:

$$S_{F_{sp}} = 2n_0A_{21}e^{h\nu/kT_r}/(e^{h\nu/kT_r} - 1)$$
$$= 2n_0(J_rB_{21} + A_{21}) \equiv 2r_0^*. \quad (56)$$

The contributions of (55) and (56) are equal since $g_0^* = r_0^*$. Making a Fourier analysis of the various fluctuating quantities in (52) one finds for S_{nn}:

$$S_{nn}(\omega) = 4g_0^*\tau^{*2}/(1 + \omega^2\tau^{*2}) \quad (57)$$

[3] In Einstein's paper the photon field is represented by the radiation density $q(\nu)$ instead of $J_r(\nu)$. Both quantities are proportional (47) but differ in dimension by [sec]. It should also be noted that the Einstein A's and B's applied to this case are not constants but depend on the occupancy of the levels to which transitions are made and hence on n since $n+n'=$constant.

in complete agreement with (16). We thus proved: 1) the photon induced transitions and the spontaneous transitions give equal contributions to the noise; 2) the noise of the transition rates behaves as if two uncorrelated and completely random electron currents "flow" between the energy states $\mathcal{E}_1$ and $\mathcal{E}_2$ [multiplying with e^2 in (55) gives the shot noise formula $S_i = 2e(eg_0^*) = 2eI$]; 3) there is complete reconciliation between the Bose statistics of the photons and the Fermi statistics of the electrons. The factor 2 in (44) is now explained: *the total noise can never be less than twice the photon induced noise.* In the above treatment we assumed that the photoconductor interacted with background radiation of a single frequency only. Obviously, the result remains the same when more frequencies are present providing the excitation mechanism is the same for all photon energies. Finally, we have to go into the case that the photoconductor is cooled and sees black-body radiation of room temperature. In these cases there is no thermal equilibrium. The absorption process is the same but the emission is usually radiationless. Since the emitted phonons have the same statistics as the photons (see also [34a]) the same result is found. If the absorption act also involves phonons, the detector is obviously not limited by photon noise and the noise equivalent power $P_{eq} \ll Q_{eq}$.

It is instructive to also express Q_{eq} in photoconductor attributes. Let us specify to this purpose the detector as an intrinsic or near-intrinsic semiconductor. Then (Section II-B) the quantity g_0^* can be expressed as $g_0^* = \langle \Delta n^2 \rangle / \tau^* = n_0p_0/\tau^*(n_0+p_0)$. Since a signal J_s gives an increase in excitation rate of η_sJ_s, we find for Q_{eq} from (55) and (56).

$$Q_{eq} = (4g_0^*\Delta f)^{1/2}/\eta_s = [4n_0p_0\Delta f/\tau^*(n_0 + p_0)\eta_s^{-2}]^{1/2}. \quad (58)$$

C. Infrared Photoconductors

Since semiconductors have a bandgap of $\sim$1 ev or smaller, these materials are suitable to detect infrared radiation. A review of the optical properties of germanium and silicon is found in a paper by Burstein, Picus, and Sclar [23]. The long wavelength limits of Ge and Si are 1.8 μ and 1.2 μ, respectively. It is known that the quantum efficiency for *absorbed* quanta is close to unity [37]. At room temperature these detectors see no radiation noise, since the lifetimes usually measured are orders of magnitude less than the radiative lifetimes which are of the order of one second [98a]. Obviously, the photon noise limit Q_{eq} will not be reached at room temperature. Since the absorption coefficient is high ($\approx 10^4$ cm^{-1}) the crystals should be very thin (about 1 micron) in order to reach the radiative limit at all. An alternative is offered by the use of junction diodes which do not have such a large dark current.

PbS, PbSe, and PbTe are intrinsic photoconductors which have been widely used [65a]. GR noise in lead sulphide was measured by Lummis and Petritz [51]. This is somewhat surprising since these materials are applied in the form of microcrystalline films. It is as-

sumed that in these materials the electrons (minority carriers) are mainly trapped. At room temperature the films are limited by dark current gr noise caused by lattice vibrations [51]. At low temperatures the photon limit has been reached for films exposed to radiation of room temperature or lower. Noise equivalent powers as low as 2×10^{-14} watt have been reported [32], [106]. Lead selenide and telluride are less sensitive at comparable temperatures. In general, Q_{eq} increases with decreasing bandgap as was discussed by Petritz [73] [compare also (62b) below].

Impurity semiconductors with deep donors and acceptors have been prepared in great variety in recent years. Germanium doped with nickel, iron, cobalt, manganese, gold, etc., has been prepared by Tyler, *et al.* [90]. These impurities generally give rise to double acceptor levels; n or p-type materials can be obtained by counterdoping. The impurity centers are of the order of 0.2–0.4 ev from the conduction and valence band. Silicon has been doped similarly [28] although less is known than for Ge. To observe photoconductivity the temperature must be so low that most impurities are unionized. For the above materials 77°K is sufficient. Since the absorption constant is of the order of 1 cm^{-1}, crystals can have the usual thickness ($\sim$1 mm). Noise in these crystals has not been reported as far as the author knows. In manganese doped germanium Fassett found that the noise was gr noise [31a] (Section II-D). This noise was measured with the crystal at low temperature and shielded from background radiation. It is known that the resistance decreases when the crystal at 77°K is exposed to room temperature radiation. Whether these crystals are still limited by photon noise when shielded by a cooled transmission filter (Section III-A) is not known yet. The noise in the germanium specimen doped with other impurities is being investigated. An estimate of the sensitivity limit will be made below.

We will calculate the quantities mentioned in Section III-A to predict the behavior of these detectors. For the signal response an equation analogous to (52) can be stated

$$d(\Delta n)/dt = -\Delta n/\tau + \eta_s J_{s0} e^{j\omega t} \tag{59}$$

or, putting $\Delta i = i_0 e^{j\omega t} = (I/n_0)\Delta n$ for a semiconductor with a simple type of carrier, we find:

$$i_0(j\omega + 1/\tau) = (I/n_0)\eta_s J_{s0} \tag{60}$$

from which we find for the responsivity (39)

$$\mathfrak{R}(\omega) = I\eta_s\tau/n_0(1+\omega^2\tau^2)^{1/2}. \tag{61}$$

With the noise given by (28) we obtain for P_{eq}:

$$P_{eq} = \frac{2}{\eta_s}\left[\frac{n_0}{\tau}\left(\frac{1-\lambda}{2-\lambda}\right)\Delta f\right]^{1/2}. \tag{62a}$$

In a similar way one obtains for near-intrinsic semiconductors

$$P_{eq} = \frac{2}{\eta_s}\left[\frac{n_0 p_0 \Delta f}{\tau(n_0+p_0)}\right]^{1/2}. \tag{62b}$$

Since p_0 and n_0 are total numbers P_{eq} is actually proportional to $A^{1/2}(\Delta f)^{1/2}$.

In PbS the result is slightly different since electron trapping and "barrier amplification" may occur [74]. For an intrinsic photoconductor, the noise figure is, according to (58) and (62b)

$$F = (\tau^*/\tau)^{1/2} \tag{63}$$

and this simple formula also holds for other two level cases. As emphasized by Petritz [73] it is not necessary to measure the absolute responsivity to find P_{eq}. If the shape of the spectrum shows that (62a) or (62b) applies, then the quantity P_{eq} may be calculated from n_0 and τ (which follow from the noise spectrum) and from an estimate of η_s. In case the spectrum is not of the above simple form, this indicates that transitions between more than two energy levels occur. The noise and the responsivity can still be calculated if the proper electron mechanism is known but this is usually complicated. Moreover, the noise and responsivity may differ in frequency dependence (next section). Lower sensitivity limits for intrinsic photoconductors have been listed in Petritz [73]. We will now make an estimate of P_{eq} for p-type Ge-Mn, using (62a). [The factor $(1-\lambda)/(e-\lambda)$ has to be omitted in this case; compare Section II-D]. At 77°K we observed: $p_0 \approx 10^9$ holes; $A \approx 0.3$ cm^2 $\tau \approx 10^{-5}$ sec. Putting $\Delta f = 1$ sec^{-1}, $\eta_s \approx 0.2$, $h\nu = 0.16$ ev, we find for P_{eq}:

$$P_{eq} \approx 2.5 \times 10^{-12} \text{ watt.}$$

This detector should be useful up to about 7 microns.

D. Photoconducting Insulators

The best known photoconductive insulator is CdS. Noise and response measurements have been reported by Shulman [84], Böer [9], and by van Vliet and Blok [102]. The latter investigators found that the noise and response were quite different for wavelengths smaller or larger than the absorption limit. The results were explained on account of the two-center model for CdS. Response times varying from 0.1 sec ($\lambda > 5100$ Å) to 10^{-4} sec ($\lambda < 5100$ Å) were observed. The response and the noise were parallel up to about 10 kc. Above that frequency the response dropped faster than the noise. This could be explained on account of electron trapping processes. In accordance with this the signal-to-noise ratio is constant up to 10 kc, even though the response time could be of the order of 1 second. Since CdS has a negligible dark resistance (depending on doping) all excitations are photon induced. Thus, this detector is definitely limited by photon noise. The introduction of concepts like P_{eq} and Q_{eq} is not unambiguous, however. The problem here is not to detect a light signal so small that quasi-thermal equilibrium can still be assumed. The dark current fluctuations, moreover, would have a

complicated behavior since these currents are space charge limited [85]. Usually CdS is illuminated strongly till reasonable conductivities (*e.g.*, 10 KΩ for a cell) are obtained. One can, therefore, define a noise equivalent power depending on a certain amount of "bias light" that plays the same role as background radiation in semiconductors. Since the response as a function of the light intensity is not linear, this value $P_{eq}(J)$ will depend on J in a complicated way. From absolute response measurements we obtained values of P'_{eq} of 10^{-11} watt for the highest light levels which may decrease by several orders of magnitude for the lowest practical light levels.

IV. 1/f Noise in Semiconductor Filaments

A. Experimental Data

When the noise in a semiconductor filament is measured between probes with a constant current generator applied to the crystal end contacts, any noise resulting from the current carrying contacts is eliminated or at least heavily reduced (similar to Fig. 2). It is thus supposed that the remaining low-frequency noise is characteristic of the filament. Whether the noise associated with the contacts is of the same nature as the noise that is attributed to the filament itself has not been systematically investigated. There is no reason to believe, however, that this necessarily should be the case. Here we turn our attention to the crystal 1/f noise.

It seems quite clear at the present time that 1/f noise in germanium single crystals is, at least for the greater part, caused by the surface conditions. Probably the most direct proof was given by Maple, Bess, and Gebbie [54] who reported a 10 to 20 db increase in 1/f noise by switching the filament from a dry nitrogen ambient to one of carbon tetrachloride. The effect of proper etching of a crystal surface on 1/f noise is also well known. Montgomery [60] could produce changes in 1/f noise by concentrating the carriers on the surface by means of a magnetic field.

The frequency range over which 1/f noise extends is quite remarkable. It has been observed down to 2.10^{-4} by Rollin and Templeton [76] for germanium filaments. For point contact diodes it has been found even at 6×10^{-6} cycles per second [1], [33]. As to the upper limit more doubt exists. In carbon resistors and other materials 1/f noise was found up to 1 mc [99], [59] above which thermal noise drowned out the effect. However, this does not mean that 1/f noise in germanium filaments extends that far, since this noise may be of other origin as remarked above. Montgomery [60] measured 1/f noise to quite high frequencies in germanium crystals. In those days germanium crystals were much less perfected than nowadays, however. Hyde [42] observed 1/f noise in two-terminal germanium crystals up to 4 mc. Beyond this frequency the noise changed into a $1/f^2$ spectrum. Since the noise was not measured between probes it is uncertain whether this noise should be considered to be characteristic for the filament. In many recent observations the 1/f noise does not extend beyond 100 to 10,000 cycles per second where it is masked by gr noise. Bess [8] reports an upper turnover frequency of ~1000 cycles per second for crystals at low temperatures. In the experiments of Maple, *et al.* [54], the 1/f noise for a near intrinsic crystal emerged in carbon tetrachloride changed from $1/f^{1.22}$ dependence into a $1/f^3$ dependence at the frequency where the gr noise also started to decrease. This is in accordance with various proposed theories in which a lower limit for the upper turnover frequency is set by $\omega_0=1/\tau_{gr}$ where τ_{gr} is the lifetime of the carriers involved. We return to this in the next section.

Another remarkable effect is the slight temperature dependence of 1/f noise. In germanium little change has been found between liquid nitrogen temperature and room temperature. Templeton and MacDonald [89] measured noise in carbon resistors between 20 cycles per second and 10 kc and found little variation in magnitude. Similar results were reported by Russel [78] for ZnO crystals.

The current dependence of 1/f noise in germanium differs somewhat from case to case. Often, the noise is proportional to the square of the dc current as one would expect for true conductivity fluctuations. Sometimes, however, a I^4 dependence is reported. Brophy [14] has shown that such a higher power dependence is quite often found in plastically deformed crystals. Bess [8] attributes this behavior to edge dislocations. His theory will be discussed in Section IV-C. Since dislocations exist throughout the crystal, the noise is in his theory a combined bulk and surface effect. It was noticed by Brophy [16] that 1/f noise, if due to conductance modulation, should not necessarily depend on the presence of an electric field. He could demonstrate that 1/f noise also occurs when conductivity changes are detected by placing the crystal in a temperature gradient rather than in an electric field, which is an important clue as to the mechanism of 1/f noise.

Significant experiments have been performed in order to identify the carrier that is responsible for 1/f noise. The first experiments performed by Montgomery seemed to indicate that the noise is associated with the minority carriers. In his setup the noise was measured between each pair of three closely spaced terminals, A, B, and C on n-type germanium samples. The noise intensities S_{AB}, S_{BC}, and S_{AC} indicated considerable correlation. From the drift length of the carriers the correlation was also calculated and found to be in good agreement with the results, if the carrier lifetime was that of the minority carriers. In more recent measurements, on the contrary, this correspondence could no longer be established [61]. Another approach to this problem was made by Brophy [13], Rostoker [15], and Bess [7] who measured the Hall effect noise. They concluded that the noise in their samples was mainly due to

majority carriers. In nearly intrinsic material fluctuations in the minority carrier density also contributed. The experimental data could best be fitted by assuming correlation between hole and electron fluctuations, such that

$$\frac{\Delta n}{n_0} = -\frac{\Delta p}{p_0}. \tag{64}$$

Bess [8] has pointed out that such a relation would be expected for slow fluctuations which modulate the Fermi level in a quasi-equilibrium-like fashion. In that case $pn = n_i^2 = \text{constant}$ from which (64) immediately follows. It is interesting to note that Bess [7] also measured the Hall effect noise at frequencies where gr noise predominated; in this case the fact that $\Delta n = \Delta p$, as stated in Section II-E, was corroborated.

Some general conclusions may be drawn from these experimental data. First of all, the effect cannot be caused by random events with a single time constant, since this results in a spectrum of the form $c\tau/(1+\omega^2\tau^2)$. Moreover, the mechanism should explain the fact that the spectrum is usually not exactly 1/f, but of the form $1/f^\alpha$, where α varies somewhere between 0.7 and 1.5 for different materials and specimen. It has been known for a long time [87], [92], [30] that a superposition of $\tau/(1+\omega^2\tau^2)$ spectra can result in a $1/f^\alpha$ law. Formally, one may write

$$S(f) = \int_{\tau_1}^{\tau_2} g(\tau) \frac{\tau}{1+\omega^2\tau^2}\, d\tau. \tag{65}$$

For $g(\tau) = A/\tau$ this results in a 1/f spectrum for $1/\tau_2 < f < 1/\tau_1$. This transfers the problem into the finding of a mechanism for $g(\tau)$. This is not simple either. Electronic transitions between traps and the conduction and valence band as suggested by Baumgartner and Thoma [2] may give long trapping times but this is not observed in the noise since the average *free* time of the carriers is much smaller. The time constant which determines the noise spectrum is always the smaller one of the two time constants involved. This is the main reason that 1/f noise cannot be caused by a superposition of gr noise terms, involving deep traps. Presuming with Brophy that the noise is not inherent in the passage of current but can be attributed to conductivity fluctuations, the two alternatives left are: either the carrier densities themselves are modulated (*e.g.*, by the random creation and disappearance of donor centers [8]), or the *rates* of the carrier transitions are modulated in some way. Bess [8] opposes this idea, arguing that because of detailed balance such fluctuations are smoothed out within a few carrier lifetimes. This may be true for the behavior of the bulk where indeed, for times large in comparison with the carrier lifetime the occupancy of all states and consequently the rate of the transitions only depends on the Fermi level. At the surface, on the contrary, the relative position of the Fermi level depends on the surface charges and can fluctuate. This will be considered in more detail in the next section.

B. McWhorter's and North's Analyses

McWhorter [58], [57] has attributed the noise to the trapping and untrapping of so-called "slow surface states" [47], [49]. Before discussing his theory we first briefly review what is known about the surface of semiconductors such as germanium. It has been known for some time that the energy bands in a semiconductor are curved at the surface due to charges trapped in surface states. By a suitable choice of the ambient it is possible to make the surface either *n*-type or *p*-type irrespective of the bulk conductivity. The space charge region can be described by the parameters ϕ_B and ϕ_S (see Fig. 3).

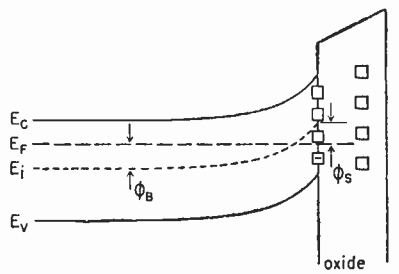


Fig. 3—Band picture for a germanium surface (after Kingston, [47]). The oxide layer is 20–40 Å.

In addition there is an oxide layer at the surface indicated by a surface barrier. The nature of the surface states was particularly investigated with the aid of the field effect [48], [63]. In this effect the conductivity of the germanium sample is modulated by changing the surface charges with the aid of a pulsed or sinusoidally modulated transverse electric field (perpendicular to the surface). The response is usually quite complex. If a pulsed field is applied there is first a relatively rapid response, reaching a value corresponding to some quasi-equilibrium state of the carriers and the surface recombination centers. Then the conductance decays slowly to its original value with a half life for the decay ranging from milliseconds to several seconds, depending on the surface treatment and the gaseous ambient. The effect has been generally analyzed, assuming that there are two groups of surface states, the "fast states" which are responsible for the recombination velocity of the carriers, and the "slow states" which give the tail in the response curve. It is further assumed that the "fast states" are at the germanium-oxide interface, and that the "slow states" are in the oxide layer, or at the outside. McWhorter assumes that free carriers communicate with the slow states by tunneling through the barrier. The attractive feature of this assumption is the temperature independence of this process. McWhorter measured the response to a sinusoidally varying field. In many cases the response could be approximated by

$$\Delta\sigma(\omega) = a \log b\omega \tag{66}$$

for frequencies $f < f_{max}$. If the slow surface states would have a single capture time constant τ, then McWhorter shows that the response should be of the form $j\omega\tau/(1+j\omega\tau)$. Note that this is the same as found in (23). The form (66) can only be explained if we introduce a distribution of τ's:

$$\Delta\sigma(\omega) = a' \int_{\tau_1}^{\tau_2} \frac{g(\tau) j\omega\tau}{1 + j\omega\tau} d\tau. \tag{67}$$

For $g(\tau) \sim 1/\tau$ the result (66) is approximately found. Apparently, this is just the distribution of time constants needed to obtain 1/f noise. It is accordingly very promising to assume that 1/f noise is caused by spontaneous fluctuations in the capture and release of carriers by the slow surface states. If τ would be due to tunneling, then, according to quantum mechanics

$$\tau = \tau_0 \exp (2mV/\hbar^2)^{1/2} w \tag{68}$$

where w is the barrier width, V the barrier height, and $\tau_0 \approx 10^{-12}$ sec. If w varies between 20 and 40 Angstrom, τ varies between 10^{-4} sec and 10^{6} sec. Unfortunately, the field experiments seem to indicate an upper value of f_{max} lower than usually found for 1/f noise. However, the analysis of the field effect for frequencies close to the carrier lifetimes is not unambiguous. Since also the transition of 1/f noise into gr noise is not well known, no discrepancy may exist at all. McWhorter has also given a quantitative calculation of the expected noise. It is felt that this noise can be found in an easier way from the general procedures outlined in Sections II-B and II-C. The slow fluctuations in electrons and holes in the surface region are then easily found. The next step is to solve for the bulk conductivity fluctuations with the aid of Poisson's equation as is also done by McWhorter.

Closely related to this procedure is a theory developed by North [67]. North assumes that the fluctuations in the surface potential ϕ_S are thermal. Hence, the fluctuations in surface recombination velocity s follow from

$$\langle \Delta\phi_s^2 \rangle = 4kTR_{eq}\Delta f \quad \langle \Delta s^2 \rangle = \left(\frac{\partial s}{\partial \phi_s}\right)^2 \langle \Delta\phi_s^2 \rangle. \tag{69}$$

The quantity R_{eq} is the real part of an equivalent impedance into which ϕ_s looks. To calculate R_{eq} an equivalent network is developed by North in which the transition rates serve as conductances and the barrier capacitance and the time constants determine the capacitances. The theory is closely related to that of Section II-C. His basic idea was applied with success by Fongers to noise in transistors and junction diodes [97].

C. Other Theories

Bess [6], [8] has proposed an entirely different interpretation of 1/f noise. In accordance with the observation that the amount of 1/f noise can be changed by plastic deformation Bess assumed that the noise was associated with edge dislocations. Impurities should be diffusing along the edge dislocation line to and from the surface where they undergo some type of Brownian motion. With a highly specialized mathematical model this results in 1/f noise. Although the application of Bess' mathematical model is doubtful, his basic idea to associate the noise with dislocations is very attractive. As pointed out by Morrison [64a] the energy band structure in the neighborhood of a dislocation is similar to that of the surface (Fig. 4) a fluctuation of the trapped charge will thus modulate ϕ_D and the recombination velocity as in previous theories. This effect has

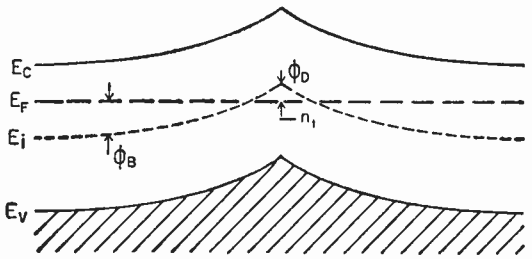


Fig. 4—Band picture at a dislocation (after Morrison, [64a]).

also been worked out by Morrison himself in a different way [64]. He assumes that the transition rates follow a relation of the Elovich type:

$$d(\Delta n_t)/dt = B(e^{b\Delta n_t} - 1) \tag{70}$$

where n_t is the trapped charge and B and b are constants. From this the correlation function and spectrum can be found. This results in a 1/f law over several decades. In contrast to North's theory one must assume very large deviations from thermal equilibrium in order to explain the nonlinear behavior.

Schönfeld [80] has found the interesting result that random events of a $1/\sqrt{i}$ character result in a 1/f spectrum according to Carson's theorem (Section II-A). However, no elementary events of such a form are known.

V. Modulation Effects in Granular Material

A. Proposed Theories

So far we have not mentioned several of the older theories for 1/f noise which were largely based on diffusion mechanisms. If these theories are applicable at all, then they might have some value for granular or microcrystalline material. MacFarlane [56] and Richardson [75] considered the diffusion of atoms or ions over the contact area of the grains. The spectra, however, are not 1/f like over many decades as pointed out by Burgess [17]. Moreover, the region in which a reasonable 1/f approximation is to be found is strongly temperature dependent. Petritz [72] has given a similar theory involving heat diffusion. McWhorter [56], [57] proposes that the tunnel processes also play a role in the

passage of current between contacting grains. Experiments performed on a single mercury-aluminum contact gave support to this idea. Various other ideas have been suggested but nothing definite about the nature of the noise is known yet. The current dependence seems to be somewhat characteristic for the material. Carbon resistors invariably give a I^2 dependence. PbS films often show a stronger current dependence, thermistors gave noise proportional to $I^{1.25}$. Several features are discussed in a wartime report by Harris, Abson, and Roberts [38].

VI. Bibliography

[1] Baker, D. "Flicker Noise in Germanium Rectifiers at Very Low Frequencies," *Journal of Applied Physics*, Vol. 25 (July, 1954), pp. 922–924.

[2] Baumgartner, W., and Thoma, H. U. "Zum Stromrauschen von Halbleitern," *Zeitschrift für angew. Physik*, Vol. 6 (1955), p. 66.

[3] Becker, R. *Theorie der Elektrizität*. Leipzig and Berlin: B. G. Teubner, sec. 70, 1933. Lithoprinted in the U.S.A. by Edwards Brothers, Inc., Michigan, 1945.

[4] Becking, A. G. T. "Fluctuatieverschijnselen by Bolometers," thesis, University of Utrecht, The Netherlands, 1953.

[5] Bernamont, J. "Fluctuations du Potential aux Bornes d'un Conducteur Métallique de Faible Volume, Parcouru par un Courant," *Annales de Physique*, Vol. 7 (1937), pp. 71–140.

[6] Bess, L. "A Possible Mechanism of 1/f Noise Generation in Semiconductor Filaments," *Physical Review*, Vol. 91 (September, 15, 1953), p. 1569.

[7] ——. "Relative Influence of Majority and Minority Carriers on Excess Noise in Semiconductor Filaments," *Journal of Applied Physics*, Vol. 26 (November, 1955), pp. 1377–1381.

[8] ——. "Study of 1/f Noise in Semiconductor Filaments," *Physical Review*, Vol. 103 (July 1, 1956), pp. 72–82.

[9] Böer, K. W., and Junge, K. "Zur Frequenzabhängigkeit von Elektronenschwankungserscheinungen in Halbleitern," *Zeitschrift für Naturforschung*, Vol. 8A (November, 1953), pp. 753–755.

[10] ——. "Einige Bemerkungen zur Gisolfschen Theorie der Elektronenschwankungserscheinungen von Halbleitern," *Annalen der Physik*, Vol. 14 (1954), pp. 87–96; also Vol. 15 (1954), pp. 55–56.

[11] ——, Kummel, U., and Molgedey, G. "Elektronenrauschen von CdS Einkristallen bei Hohen Feldstärken," *Annalen der Physik*, Vol. 17 (1956), pp. 344–356.

[12] Brophy, J. J. "Current Noise in Thermistor Bolometer Flakes," *Journal of Applied Physics*, Vol. 25 (February, 1954), pp. 221–224.

[13] ——, and Rostoker, N. "Hall Effect Noise," *Physical Review*, Vol. 100 (October 15, 1955), pp. 754–756.

[14] ——. "Excess Noise in Deformed Germanium," *Journal of Applied Physics*, Vol. 27 (November, 1956), pp. 1383–1384.

[15] ——. "Excess Noise in *n*-Type Germanium," *Physical Review*, Vol. 106 (May 15, 1957), pp. 675–678.

[16] ——. "Experimental Investigation of Excess Noise in Semiconductors." Presented before the American Physical Society, Boulder, Colo., September 5–7, 1957.

[17] Burgess, R. E. "Contact Noise in Semiconductors," *Proceedings of the Physical Society*, Vol. B66 (April, 1953), pp. 334–335.

[18] ——. "Fluctuations in the Number of Charge Carriers in a Semiconductor," *Physica*, Vol. 20 (November, 1954), pp. 1007–1010.

[19] ——. "Fluctuations of the Number of Electrons and Holes in a Semiconductor," *Proceedings of the Physical Society*, Vol. B68 (September, 1955), pp. 661–671.

[20] ——. "Electronic Fluctuations in Semiconductors," *British Journal of Applied Physics*, Vol. 6 (June, 1955), pp. 185–190.

[21] ——. "The Statistics of Charge Carrier Fluctuations in Semiconductors," *Proceedings of the Physical Society*, Vol. B69 (October, 1956), pp. 1020–1027.

[22] ——. "Statistical Fluctuations in Semiconductors." Presented before the American Physical Society, Boulder, Colo., September 5–7, 1957.

[23] Burstein, E., Picus, G., and Sclar, N. "Optical and Photoconductive Properties of Silicon and Germanium," in *Photoconductivity Conference*, R. G. Breckenridge *et al.*, eds. New York: John Wiley and Sons, Inc., 1956.

[24] Buttler, W. M. "Über das Randschichtrauschen in Halbleitern," *Annalen der Physik*, Vol. 11 (1952), pp. 362–367.

[25] ——, and Muscheid, W. "Die Bedeutung des elektrischen Kontaktes bei Untersuchungen an Kadmium Sulfid-Einkristallen," *Annalen der Physik*, Vol. 14 (1954), pp. 215–219; also Vol. 15 (1954), pp. 82–111.

[26] Callen, H. B. and Welton, T. A. "Irreversibility and Generalized Noise," *Physical Review*, Vol. 83 (July 1, 1951), pp. 34–39.

[27] Christenson, C. J., and Pearson, G. L. "Spontaneous Fluctuations in Carbon Microphones and Other Granular Resistors," *Bell System Technical Journal*, Vol. 15 (April, 1936), pp. 197–223.

[28] Collin, C. B., and Carlson, R. O. *Bulletin of the American Physical Society II*, Vol. 1 (March 15, 1955), p. 127.

[29] Davydov, B., and Gurevich, B. "Voltage Fluctuations in Semiconductors," *Journal of Physics of the USSR*, Vol. 7 (1943), pp. 138–140.

[30] duPré, F. K. "A Suggestion Regarding the Spectral Density of Flicker Noise," *Physical Review*, Vol. 78 (June 1, 1950), p. 615.

[31] Einstein, A. *Physikalisch Zeitschrift*, Vol. 18 (1917), p. 121.

[31a] Fassett, J. R., M.Sc. thesis, University of Minnesota, Minneapolis, Minn., 1958, unpublished.

[32] Fellgett, P. B. "On the Ultimate Sensitivity and Practical Performance of Radiation Detectors," *Journal of the Optical Society of America*, Vol. 39 (November, 1949), p. 970.

[33] Frle, T., and Winston, H. "Noise Measurements in Semiconductors at Very Low Frequencies," *Journal of Applied Physics*, Vol. 26 (June, 1955), p. 716.

[34] Fowler, R. "Statistical Mechanics," Cambridge: University Press, sec. 20.71, 1936.

[34a] Fröhlich, H. "Elektronentheorie der Metalle," Berlin: Springer Verlag, 1936. See especially sec. 13.

[35] Gebbie, H. A. "Excess Noise and Trapping in Germanium," *Physical Review*, Vol. 98 (June 1, 1955), p. 1567.

[36] Gisolf, J. H. "On the Spontaneous Current Fluctuations in Semiconductors," *Physica*, Vol. 15 (September, 1949), pp. 825–832.

[37] Gouscher, F. S. "The Photon Yield of Electron-Hole Pairs in Germanium," *Physical Review*, Vol. 78 (June 15, 1950), p. 816.

[38] Harris, E. J., Abson, W., and Roberts, W. J. Report, Telecommunications Research Establishment, 1946 (unpublished).

[39] Herzog, G. B., and van der Ziel, A. "Shot Noise in Germanium Single Crystals," *Physical Review*, Vol. 84 (December 15, 1951), p. 1249.

[40] Hill, J. E., and van Vliet, K. M. "Generation-Recombination Noise in Intrinsic and Near-Intrinsic Germanium Crystals," *Journal of Applied Physics*, Vol. 29 (February, 1958), pp. 177–182.

[41] Hyde, F. J. "Measurements of Noise Spectra of a Point Contact Germanium Rectifier," *Proceedings of the Physical Society*, Vol. B66 (December, 1953), pp. 1017–1024.

[42] ——. "Excess Noise Spectra in Germanium," *Proceedings of the Physical Society*, Vol. B69 (February, 1956), pp. 242–245.

[43] "Shot Noise in a Germanium Filament," *Report of the Conference on Semiconductors of the Physical Society*, Rugby, England (1956), pp. 57–64.

[44] Johnson, J. B. "The Schottky Effect in Low Frequency Circuits," *Physical Review*, Vol. 26 (July, 1925), pp. 71–85.

[45] Jones, R. C. "The Ultimate Sensitivity of Radiation Detectors," *Journal of the Optical Society of America*, Vol. 37 (November, 1947), pp. 879–890.

[46] ——. "Performance of Radiation Detectors" in *Advances of Electronics*. New York: Academic Press, Vol. 5, 1953.

[47] Kingston, R. H. "Review of Germanium Surface Phenomena," *Journal of Applied Physics*, Vol. 27 (February, 1956), pp. 101–114.

[48] ——, and McWhorter, A. L. "Relaxation Time of Surface States on Germanium," *Physical Review*, Vol. 103 (August 1, 1956), pp. 534–540.

[49] ——, *et al. Surface Physics*. Philadelphia: University of Pennsylvania Press, 1957.

[50] Lewis, W. B. "Fluctuations in Streams of Thermal Radiation," *Proceedings of the Physical Society*, Vol. 59 (January, 1947), pp. 34–40.

[51] Lummis, F. L., and Petritz, R. L. "Noise, Time Constant and Hall Studies on Lead Sulfide Photoconductive Films," *Physical Review*, Vol. 105 (January 15, 1957), pp. 502–508.

[52] Machlup, S. "Noise in Semiconductors; Spectrum of a Two Parameter Random Signal," *Journal of Applied Physics*, Vol. 25 (March, 1954), pp. 341–343.

[53] MacQuiston, R. B. Private communication.

[54] Maple, T. G., Bess, L., and Gebbie, H. A. "Variation of Noise with Ambient in Germanium Filaments," *Journal of Applied Physics*, Vol. 26 (April, 1955), p. 490.

[55] Mattson, R. H., and van der Ziel, A. "Shot Noise in Germanium Filaments," *Journal of Applied Physics*, Vol. 24 (February, 1953), p. 222.
[56] McFarlane, G. G. "A Theory of Contact Noise in Semiconductors," *Proceedings of the Physical Society*, Vol. B63 (October, 1950), pp. 807–814.
[57] McWhorter, A. L. "1/f Noise and Related Surface Effects in Germanium," Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., Report No. 80 (May, 1955), unpublished.
[58] ——. "1/f Noise and Germanium Surface Properties," in *Semiconductor Surface Physics*, R. H. Kingston *et al.*, eds. Philadelphia: University of Pennsylvania Press, 1957.
[59] Miller, P. H., Jr. "Noise Spectrum of Crystal Rectifiers," PROCEEDINGS OF THE IRE, Vol. 35 (March, 1947), pp. 252–256.
[60] Montgomery, H. C. "Electrical Noise in Semiconductors," *Bell System Technical Journal*, Vol. 31 (September, 1952), pp. 950–975.
[61] ——. Private communication to A. L. McWhorter.
[62] ——. Private communication.
[63] ——, and Brown, W. L. "Field-Induced Conductivity Changes in Germanium," *Physical Review*, Vol. 103 (August 15, 1956), pp. 865–870.
[64] Morrison, S. R. "Generation of 1/f Noise by Levels in a Linear or Planar Array," *Physical Review*, Vol. 99 (September 15, 1955), p. 1904.
[64a] ——. "Recombination of Electrons and Holes at Dislocations," *Physical Review*, Vol. 104 (November 1, 1956), pp. 619–623.
[65] Moss, T. S. "The Ultimate Limits of Sensitivity of PbS and PbTe Photoconductive Detectors," *Journal of the Optical Society of America*, Vol. 40 (September, 1950), pp. 603–607.
[65a] ——. "Lead Salt Photoconductors," PROCEEDINGS OF THE IRE, Vol. 43 (December, 1955), pp. 1869–1881.
[66] Mott, N. F., and Gurney, R. W. "Electronic Processes in Ionic Crystals." New York: Oxford University Press, second edition, 1948.
[67] North, D. O. "Theory of Noise Processes in Diodes and Transistors." Presented before the meeting of the American Physical Society, Boulder, Colo., September 5–7, 1957.
[68] Oliver, D. J. "Fluctuations in the Number of Electrons and Holes in a Semiconductor," *Proceedings of the Physical Society*, Vol. B70 (February, 1957), pp. 244–247.
[69] ——. "Current Noise in Indiumantimonide," *Proceedings of the Physical Society*, Vol. B70 (March, 1957), pp. 331–332.
[70] Petritz, R. L. "On the Theory of Noise in *P-N* Junctions and Related Devices," PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1440–1456.
[71] ——. "On the Diffusion Theory of Noise in Rectifiers and Transistors," *Physical Review*, Vol. 87 (July 1, 1952), p. 189.
[72] ——. "Theory of Contact Noise," *Physical Review*, Vol. 87 (August 1, 1952), p. 535.
[73] ——. "The Relation Between Lifetime, Limit of Sensitivity and Information Rate in Photoconductors," in *Photoconductivity Conference*, R. G. Breckenridge *et al.*, eds. New York: John Wiley and Sons, Inc., 1956.
[74] ——. "Theory of Photoconductivity in Semiconductor Films," *Physical Review*, Vol. 104 (December 15, 1956), pp. 1508–1516.
[75] Richardson, J. M. "The Linear Theory of Fluctuations Arising from Diffusion Mechanisms; An Attempt at a Theory of Contact Noise," *Bell System Technical Journal*, Vol. 29 (January, 1950), pp. 117–141.
[76] Rollin, R. V., and Templeton, I. M. "Noise in Semiconductors at Very Low Frequencies," *Proceedings of the Physical Society*, Vol. B66 (March, 1953), pp. 259–261.
[77] Rose, A. "Performance of Photoconductors," in *Photoconductivity Conference*, R. G. Breckenridge *et al.*, eds. New York: John Wiley and Sons, Inc., 1956.
[78] Russell, R. R. Tenth Annual Conference on Physical Electronics, M.I.T., Cambridge, Mass., March 30–April 1, 1950.
[79] Sautter, D., and Seiler, K. "Über das Rauschen von Germanium Einkristallen," *Zeitschrift für Naturforschung*, Vol. 12A (June, 1957), p. 490.
[80] Schönfeld, H. "Beitrag zum 1/f Gesetz beim Rauschen von Halbleitern," *Zeitschrift für Naturforschung*, Vol. 10A (April, 1955), pp. 291–300.
[81] Shockley, W. *Electrons and Holes in Semiconductors*. New York: D. van Nostrand Co., Inc., pp. 342–346, 1950.
[82] *Ibid.*, pp. 319–325.
[83] ——, and Read, W. T. "Statistics of the Recombination of Holes and Electrons," *Physical Review*, Vol. 87 (September 1, 1952), pp. 835–842.
[84] Shulman, G. I. "Shot Noise in CdS Crystals," *Physical Review*, Vol. 98 (April 15, 1955), pp. 384–386.
[85] Smith, R. W., and Rose, A. "Space-Charge-Limited Currents in Single Crystals of Cadmium Sulfide," *Physical Review*, Vol. 97 (March 15, 1955), pp. 1531–1537.
[86] Smits, G. H., *et al.* "Excess Noise in Indium-Antimonide," *Journal of Applied Physics*, Vol. 27 (1956), p. 1385.
[87] Surdin, M. M. "Fluctuations de Courant Thermionique et le Flicker Effect," *Journal de Physique et le Radium*, Vol. 10 (April, 1939), pp. 188–189.
[88] ——. *Revue Générale d'Electricité*, Vol. 47 (1940), p. 97.
[89] Templeton, I. M., and McDonald, D. K. C. "The Electrical Conductivity and Current Noise of Carbon Resistors," *Proceedings of the Physical Society*, Vol. B66 (August, 1953), pp. 680–684.
[90] *E.g.*, Tyler, W. W., and Woodbury, H. H. "Properties of Germanium Doped with Iron; I Electrical Conductivity," *Physical Review*, Vol. 96 (November 15, 1954), pp. 874–886.
Newman, R., and Tyler, W. W. "Properties of Germanium Doped with Iron; II Photoconductivity," *Physical Review*, Vol. 96 (November 15, 1954), pp. 882–886. The same authors have published similar papers on many other impurity elements in germanium. They have appeared in recent volumes of the *Physical Review*.
[91] Tolman, R. C. *The Principles of Statistical Mechanics*. New York: Oxford University Press, 1938.
[92] van der Ziel, A. "On the Noise Spectra of Semiconductor Noise and of Flicker Effect," *Physica*, Vol. 16 (April, 1950), pp. 359–372.
[93] ——. "Shot Noise in Semiconductors," *Journal of Applied Physics*, Vol. 24 (February, 1953), pp. 222–223.
[94] ——. "Simpler Explanation of the Observed Shot Effect in Germanium Filaments," *Journal of Applied Physics*, Vol. 24 (August, 1953), p. 1063.
[95] ——. "Note on the Shot Effect in Semiconductors and Flicker Effect in Oxide Cathodes," *Physica*, Vol. 19 (August, 1953), pp. 742–744.
[96] ——. *Noise*. Englewood Cliffs: Prentice Hall, Inc., 1954.
[97] ——. "Noise in Junction Transistors," this issue, p. 1019.
[98] van Roosbroeck, W. "The Transport of Added Carriers in a Homogeneous Semiconductor," *Physical Review*, Vol. 91 (July 15, 1953), pp. 282–288.
[98a] ——, and Shockley, W. "Photon Radiative Recombination of Electrons and Holes in Germanium," *Physical Review*, Vol. 94 (June 15, 1954), pp. 1558–1561.
[99] van Vliet, K. M., van Leeuwen, C. J., Blok, J., and Ris, C. "Measurements on Current Noise in Carbon Resistors and in Thermistors," *Physica*, Vol. 20 (August, 1954), pp. 481–496.
[100] ——, and Blok, J. "Electronic Noise in Semiconductors," *Physica*, Vol. 22 (March, 1956), pp. 231–242.
[101] ——, ——. "Electronic Noise in Photoconducting Insulators," *Physica*, Vol. 22 (June, 1956), pp. 525–540.
[102] ——, ——, Ris, C., and Steketee, J. "Measurements of Noise and Response to Modulated Light of Cadmium Sulphide Single Crystals," *Physica*, Vol. 22 (August, 1956), pp. 723–740.
[103] ——. "On the Equivalence of the Fokker-Planck Method and the Free Energy Method for the Calculation of Carrier Density Fluctuations in Semiconductors," *Physica*, Vol. 23 (March, 1957), pp. 248–252.
[104] ——. "Irreversible Thermodynamics and Carrier Density Fluctuations in Semiconductors," *Physical Review*, Vol. 110 (April 1, 1950), pp. 50–60.
[105] ——, and Hill, J. E. "Ambipolar Transport of Carrier Density Fluctuations in Germanium," to be published.
[106] Watts, B. N. "Increased Sensitivity of Infrared Photoconductive Receivers," *Proceedings of the Physical Society*, Vol. A62 (July, 1949), pp. 456–457.
[107] Williams, N. H., and Thatcher, E. W. "On Thermal Electronic Agitation in Conductors," *Physical Review*, Vol. 40 (April 1, 1932), p. 121.

Noise in Junction Transistors*

A. VAN DER ZIEL†, FELLOW, IRE

Summary—**This paper gives a survey of the problem of shot noise and flicker noise in junction diodes and junction transistors. After a short introduction in Section I, the theory of shot effect is presented in Section II. First a simplified low-frequency theory is given and the close correspondence with earlier (heuristic) equivalent circuits is indicated. Then the theory is given in a more rigorous form, both from a collective point of view (Petritz, North, and van der Ziel) and from a corpuscular point of view (Uhlir, van der Ziel and Becking). Finally the conditions under which the theory holds are summed up and the possibility of deviations is discussed. Section III gives Fonger's theory of flicker noise in diodes and transistors and incorporates his discussion of base modulation effects into the equivalent noise circuit in a manner that differs somewhat from Fonger's original presentation. Section IV gives the experimental verification of the theory by Guggenbuehl and Strutt, Nielsen, Hanson and van der Ziel and others and also discusses some further experimental material. Finally, the problem of low-noise circuits, the choice of the operating point of the transistor, and the design criteria for low-noise transistors is discussed. Section V extends Fonger's theory of base modulation effects to shot effect and discusses possible consequences of this effect.**

I. Introduction

LET an active four-terminal network be connected to a signal source of internal impedance $Z_s = R_s + jX_s$ or internal admittance $Y_s = 1/Z_s = g_s + jb_s$. The noisiness of the network may then be characterized in many ways. One may, *e.g.*, represent the noise by an equivalent emf e_n in series with the source or by an equivalent noise current generator i_n in parallel to the source; these quantities are defined such that the output noise power of the network is doubled if the noise emf e_n or the noise current generator i_n are introduced. One may then define the *equivalent noise resistance* R_n or the *input equivalent saturated diode current* I_n of the network by the equations

$$\overline{e_n^2} = 4kTR_n\Delta f; \qquad \overline{i_n^2} = 2eI_n\Delta f \tag{1}$$

where T is room temperature, k is Boltzmann's constant, e is the electron charge, and Δf a small frequency interval. Both quantities R_n and I_n may depend upon the internal impedance of the source.

It is more common to introduce the *noise figure* F of the network as the ratio of the total output noise power over the output noise power due to the thermal noise of the source. The latter can be represented by a noise emf $\sqrt{4kTR_s\Delta f}$ in series with the source or by a noise current generator $\sqrt{4kTg_s\Delta f}$ in parallel to the source. Then, according to (1)

$$F = \frac{R_n}{R_s} = \frac{e}{2kT}\frac{I_n}{g_s}. \tag{2}$$

The noise figure always shows a parabolic dependence on R_s and has a minimum value $F_{\min}$ for $R_s = (R_s)_{\min}$.

The smallest available signal power that can be detected against the noise background of an amplifier of noise figure F and bandwidth B is about $FkTB$. One thus wants to make the noise figure F as small as possible under the existing operating conditions.

In many cases it is possible to change the source impedance, as viewed from the input of the amplifier, within a wide range with the help of a lossless matching network. In such cases the amplifier with the lowest value of $F_{\min}$ is the best one. In other cases it is necessary to connect the signal source directly to the amplifier without the benefit of a lossless matching network; in that case the amplifier with the lowest value of $F_{\min}$ may be a rather poor choice. In the case of a low-impedance signal source the amplifier with the lowest noise resistance R_n is the best one. Whereas, with a high-impedance signal source the amplifier with the lowest input equivalent saturated diode current I_n is preferred.

The first step in characterizing the noisiness of an amplifier stage consists of finding the noise sources in their active element and locating the proper positions of these sources in their equivalent circuit. It is then possible to determine the most suitable operating conditions of a given active element, or to design the active element so that it gives the lowest noise figure F under the existing operating conditions.

The representation of the noise properties of an active network by an equivalent circuit is not unique, since a given circuit can be transformed into another one by applying certain network theorems. Usually one tries to find the equivalent circuit that fits closest to the physics of the device.

Since junction transistors have found many applications in amplifier circuits, it is important to have a good understanding of the noise properties of these devices. The noise properties of junction diodes also are of interest for two reasons: their noise properties are closely related to those of junction transistors and certain transistor equations follow directly from the corresponding diode equations. They also are of intrinsic interest because of their use as low-level radiation detectors.

It was found that the noise in these devices consists of two parts, a flicker noise part with a low-frequency noise spectrum and a shot noise part with, at least at low frequencies, a flat spectrum. Flicker noise probably is caused by a modulation mechanism located at the surface of the devices; it can be considerably reduced by appropriate surface treatments and therefore is not a basic limitation. Shot noise is due to the corpuscular character of the current flow and thus represents a basic

* Original manuscript received by the IRE, February 28, 1958; revised manuscript received, April 17, 1958.
† Elec. Eng. Dept., University of Minnesota, Minneapolis, Minn.

limitation, so that it is important to have a good understanding of the phenomenon.

II. Theory of Shot Noise in Junction Diodes and Transistors

The problem may be treated theoretically in two equivalent ways.

1) The collective approach, where the noise is attributed to the random diffusion of minority carriers and to the random recombination and generation of hole-electron pairs.
2) The corpuscular approach, where the shot noise is attributed to a series of random and independent events, *viz.*, the crossing of the emitter and/or the collector junction by the individual current carriers.

Petritz published the first paper on the collective approach, using a lumped-parameter approximation [44]. Later he solved the one-dimensional diode problem more accurately [44a] and obtained a result that is nearly identical with (16) of this paper; unfortunately, this result was not given in an easily applicable form and no detailed account of this work was published. North [42] showed that the mathematical difficulties could be greatly simplified by representing the diffusion and recombination of minority carriers by a distributed *RC* network.[1] Van der Ziel [55] treated both the one-dimensional diode and the one-dimensional transistor in this manner. Solow [49a] extended the theory to two- and three-dimensional geometries in his thesis; the thesis also gives a summary of Petritz's unpublished work. The one-dimensional problem was also solved independently by Becking with the help of the collective method,[2] but his results were not published.

Weisskopf [60] applied the corpuscular approach to crystal diodes; a similar approach is also the (hidden) basis of the earlier heuristic theories of diode and transistor noise [15], [40], [53]. Uhlir [52] extended the method to high frequencies for diodes; van der Ziel and Becking [56] generalized his approach and extended it to transistors.

A. The Low-Frequency Corpuscular Approach

First consider noise in junction diodes. Let the diode have a characteristic:

$$I = I_o(e^{eV/kT} - 1). \tag{3}$$

At low frequencies its admittance Y is a conductance G_o:

$$Y = G = G_o = \frac{dI}{dV} = \frac{e(I + I_o)}{kT}. \tag{4}$$

One may now consider the diode current I to consist of two parts, a part $(I+I_o)$ and a part $-I_o$; the minus sign indicates that the currents flow in opposite directions. Both currents should fluctuate independently and each should show full shot noise (see Section II-E). Hence, if the total noise of the junction is represented by a current generator i in parallel to the junction admittance $Y=G=G_o$, we have

$$\overline{i^2} = 2e(I + I_o)\Delta f + 2eI_0\Delta f. \tag{5}$$

This result should be valid for arbitrary diodes at low frequencies.

Application to point contact diodes showed that reasonable agreement could be obtained between theory and experiment for the case of forward bias, provided that the thermal noise of the contact resistance r of the diode was taken into account; this leads to the equivalent circuit of Fig. 1(a) [51,] [54]. Anderson and van der Ziel applied the equivalent circuit for low frequencies but could not explain their high-frequency data [2]. This is discussed later.

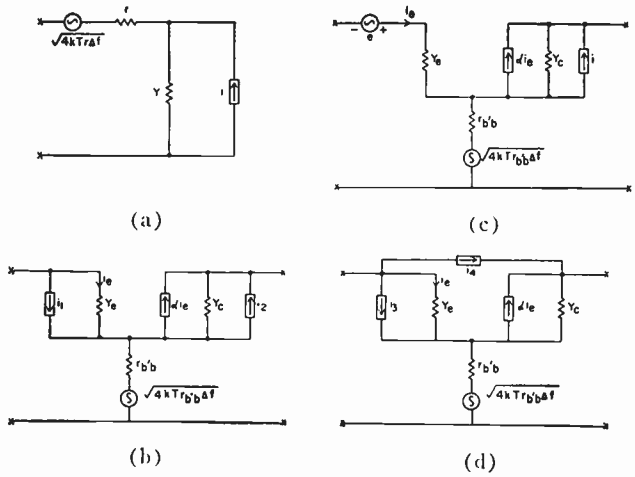


Fig. 1—Equivalent circuits for shot noise. (a) Equivalent circuit of a junction diode. (b) Equivalent circuit of a transistor. (c) Equivalent circuit of Montgomery, Clark, and van der Ziel. (d) Equivalent circuit of Giacoletto.

The above theory of the junction diode is easily extended to transistors. Consider, for example a *p-n-p* transistor; for sake of simplicity it is assumed at first that all current is carried by holes. Let I_e be the emitter current and I_c the collector current and let the collector be biased so that it does not inject holes into the base region. The current I_e can now be considered as consisting of a part (I_c+I_{ee}) due to holes flowing from the emitter to the base and a part $(-I_{ee})$ due to holes flowing from the base to the emitter. Both currents should fluctuate independently and each should show full shot noise (Section II-E). Hence, if the emitter noise is represented by a current generator i_1 in parallel to the emitter junction,

$$\overline{i_1^2} = 2e(I_e + I_{ee})\Delta f + 2eI_{ee}\Delta f. \tag{6}$$

In the collector junction all holes move in the same direction. One would thus expect full shot noise for the

[1] Petritz has used this method in the derivation of his 1953 diode formula [44a]. (R. L. Petritz, private communication.)
[2] A. G. T. Becking, private communication.

collector current I_c. Representing the collector noise by a current generator i_2 in parallel to the collector junction, we have

$$\overline{i_2^2} = 2eI_c\Delta f. \tag{7}$$

These equations follow directly from (5). The emitter noise is obtained by substituting $I=I_e$ and $I_o=I_{eo}$; the collector noise is obtained by putting $(I+I_o)=0$ and $I_o=I_c$. The latter seems strange at first, but it should be remembered that the collector is biased in the back direction and that $(I+I_o)$ corresponds to the hole current injected from the collector into the base, which is zero because of the existing bias conditions. Since (5) is valid for an arbitrary diode, (6) and (7) should also be valid if part of the current is carried by electrons (see below).

In addition, one would expect thermal noise for the series resistance of junctions. The most important thermal noise source is the true base resistance $r_{b'b}$.[3]

The full equivalent circuit thus is as shown in Fig. 1(b). The emitter admittance Y_e shown in this circuit is actually a conductance G_{eo} at low frequencies:

$$Y_e = G_e = G_{eo} = \frac{\partial I_e}{\partial V_e} = \frac{e(I_e + I_{ec})}{kT}. \tag{8}$$

The current generators i_1 and i_2 are strongly correlated.

To calculate the cross correlation $\overline{i_1^* i_2}$ (the asterisk denotes the conjugate complex quantity), we observe that the part $\beta_o(I_e+I_{ee})$ of the hole current (I_e+I_{ee}) injected into the base region by the emitter is collected by the collector; the quantity β_o is the *dc collector efficiency* of the collector junction. If the holes that are generated in the base region and are collected by the collector give a contribution I_{cc} to I_c, then

$$I_c = \beta_0(I_e + I_{ee}) + I_{cc} = \beta_0 I_e + I_{co};\ I_{co} = \beta_o I_{ec} + I_{cc}. \tag{9}$$

The quantity I_{co} is called the *collector saturated current*; it is the collector current for open emitter.[4] The emitter and collector thus have the current $\beta_o(I_e+I_{ee})$ in common.

The cross correlation $\overline{i_1^* i_2}$ is caused by fluctuations in this current. It should have full shot noise, hence

$$\overline{i_1^* i_2} = 2e\beta_o(I_e + I_{ee})\Delta f. \tag{10}$$

The current $\beta_o(I_e+I_{ee})$ is also responsible for the signal transfer properties. The transfer admittance Y_{ce} of the transistor is actually a transfer conductance G_{ceo} at low frequencies:

$$Y_{ce} = G_{ceo} = \frac{\partial I_c}{\partial V_e} = \frac{\beta_o e(I_e + I_{ee})}{kT}. \tag{11}$$

The signal transfer can be represented by a current generator $Y_{ce}v_e$ in parallel to the collector junction, where v_e is the ac emitter voltage. This current generator is also shown in the equivalent circuit of Fig. 1(b)

Next, drop the assumption that all current is carried by holes. There are then also electrons going from emitter to base, from base to emitter, and from collector to base. Each of these electrons contributes either to I_e or to I_c, but no electrons contribute to *both* I_e and I_c. The conditions for full shot noise again exist and hence (6) through (8) remain valid, provided that the currents (I_e+I_{ee}) and $(-I_{ee})$ are now properly redefined; for example, (I_e+I_{ee}) is now partly caused by holes injected from the emitter into the base and partly by electrons injected from the base into the emitter.

In this case let the part $\gamma_o(I_e+I_{ee})$ of the emitter current be due to holes injected into the base by the emitter; γ_o is known as the *dc emitter efficiency*. The part β_o of these holes is collected by the collector; the emitter and collector junctions thus have the current $\gamma_o\beta_o(I_e+I_{ee})$ in common and this current should show full shot noise.

We define the dc current amplification factor α_o

$$\alpha_o = \gamma_o\beta_o \tag{12}$$

of the transistor and observe that the quantity β_o in (9) through (11) should be replaced by α_o in this case. These equations thus become:

$$\overline{i_1^* i_2} = 2e\alpha_0(I_e + I_{ee})\Delta f \tag{13}$$

$$Y_{ce} = G_{ceo} = \alpha_o G_{eo} = \alpha_o \frac{e(I_e + I_{ee})}{kT} \tag{14}$$

$$I_c = \alpha_o I_e + I_{co};\qquad I_{co} = \alpha_o I_{ee} + I_{cc}. \tag{15}$$

B. Extension to High Frequencies

Eqs. (4) through (6), (8), (10) through (12), (14), and (15) cease to be valid at higher frequencies. For example, the diode admittance Y becomes complex and its real part G is no longer equal to G_o at high frequencies. The emitter admittance Y_e of a transistor also becomes complex and its real part G_e is no longer equal to G_{eo}. Finally, the transfer admittance Y_{ce} becomes complex and $|Y_{ce}|$ decreases with increasing frequency. The noise equations (5), (6), and (13) should thus be extended to high frequencies; in addition, the current amplification factor should be redefined. The extension of the noise equations to high frequencies is carried out in Sections II-D and II-E; here we only quote the results and show that they are compatible with the equations derived for low frequencies.

For diodes, the following equation is of general validity:

$$\overline{i^2} = 4kTG\Delta f - 2eI\Delta f \tag{16}$$

where I is taken positive for forward bias and negative

[3] The "true" base resistance $r_{b'b}$ is the base resistance found when the contribution of Early's feedback emf $\mu_{ec}v_e$ (v_e is the ac voltage across the emitter junction) to the "measured" base resistance is subtracted [11]. Early's feedback emf is omitted from the equivalent noise circuit given here, since it does not affect the noise figure of the device [55].

[4] The definition of I_{co} shows that I_{cc} is *not* the collector saturated current but that it is related to it. In the same way I_{ee} is *not* the emitter saturated current.

for back bias. For low frequencies $G=G_o$; substituting (4) for G_o, we obtain (5), so that (5) and (16) are compatible. The new equation fits well with the experimental data, as was shown by Champlin [7], [8] (Section IV-A).

For transistors, the following equations are of general validity:

$$\overline{i_1^2} = 4kTG_e\Delta f - 2eI_e\Delta f \tag{17}$$

$$\overline{i_2^2} = 2eI_e\Delta f \tag{18}$$

$$\overline{i_1^* i_2} = 2kTY_{ce}\Delta f. \tag{19}$$

At low frequencies, $G_e=G_{eo}$; substituting (8) into (17) gives (6) back again. Moreover, $Y_{ce}=G_{ceo}$ at low frequencies; substituting (14) into (19) gives (13) back again. Eqs. (17) through (19) thus seem to be the proper extensions of (6), (7), and (10) [or (13)] for higher frequencies.

We observe with Guggenbuehl and Strutt [23] that (17) and (18) are a direct consequence of (16). For if (16) holds for arbitrary diodes, it will also hold for the emitter diode and for the collector diode. The emitter diode has a conductance G_e and a current I_e; substituting this into (16) yields (17). The collector diode has $G=0$ and $I=-I_c$ (because the collector diode is biased in the back direction, we have to use the minus sign; a diode biased in the back direction has practically zero conductance). Substituting this into (16) yields (18).

Eqs. (16) through (19) reduce to thermal noise if the proper amounts of shot noise power are *added* for forward bias and *subtracted* for back bias. Guggenbuehl and Strutt [23] have given arguments in favor of such a procedure and have used it to derive these equations from thermal noise considerations only. One may consider this as a first attempt towards a thermodynamical derivation of the equations. A rigorous derivation of these equations along these lines should be based upon the principles of irreversible thermodynamics.

At low frequencies, the current amplification factor α_o may be defined as $\alpha_o=G_{ceo}/G_{eo}$, according to (14). It thus seems logical to define the high-frequency amplification factor α as

$$\alpha = \frac{Y_{ce}}{Y_e}. \tag{20}$$

The value $|\alpha|$ decreases with increasing frequency; generally

$$\alpha = \frac{\alpha_o}{1 + jf/f_o} \tag{20a}$$

where f_o is the α-cutoff frequency.

C. Other Equivalent Circuits

The equivalent circuit of Fig. 1(b) is nearly identical with two other equivalent circuits that were developed earlier on a more or less heuristic basis by Montgomery, Clark, and van der Ziel [40], [53] and by Giacoletto [15].

The first proposed the equivalent circuit of Fig. 1(c). It contains two independent noise sources, a noise emf e_e in series with the emitter and a noise current generator i in parallel to the collector junction. Both noises were assumed to be uncorrelated, and[5]

$$\overline{e_e^2} = 2kTR_{eo}\Delta f\left(\frac{I_e + 2I_{ee}}{I_e + I_{ee}}\right);$$

$$\overline{i^2} = 2e\alpha_0(1-\alpha_0)I_e\Delta f + 2eI_{co}\Delta f, \tag{21}$$

where $R_{eo}=1/G_{eo}$. The signal transfer properties of the transistor are represented in this circuit by the current generator $\alpha_o i_e$, where i_e is the current passing through the emitter junction.

The circuit of Fig. 1(b) is easily transformed into the one of Fig. 1(c); this leads to

$$i = (i_2 - \alpha_0 i_1); \qquad e_e = i_1 R_{eo}, \tag{21a}$$

according to (14). The quantity e_e^2 is calculated from (6) and (8) and the expression for $\overline{i^2}$ follows from (6), (7), and (13). Finally, it is indeed true that e_e and i are practically uncorrelated, for

$$\overline{e_e^* i} = R_{eo}\overline{i_1^*(i_2 - \alpha i_1)} = 2e\alpha_o I_{ee}R_{eo}\Delta f \tag{21b}$$

for $I_e \gg I_{ee}$ this is small in comparison with $\sqrt{\overline{e_e^2}\cdot\overline{i^2}}$.

Giacoletto [15] represented the noise by two uncorrelated current generators i_3 and i_4; i_3 was connected in parallel to the emitter junction and i_4 was connected between the emitter and the collector junction, whereas

$$\overline{i_3^2} = 2eI_b\Delta f; \qquad \overline{i_4^2} = 2eI_c\Delta f, \tag{22}$$

where $I_b=(I_e-I_c)$ is the base current. This equivalent circuit is shown in Fig. 1(d).

The circuit of Fig. 1(b) is easily transformed into the one of Fig. 1(d); this leads to

$$i_1 = i_3 + i_4; \qquad i_2 = i_4 \tag{23}$$

so that $\overline{i_4^2}$ follows directly from (7), whereas

$$\overline{i_3^2} = \overline{(i_1 - i_2)^2} = 2eI_b\Delta f + 4e(I_{ee} + I_{ee})\Delta f \tag{24}$$

according to (6), (7), (13), and (15). This corresponds to (22), if $2(I_{ee}+I_{ee}) \ll I_b$. Moreover, i_3 and i_4 are practically uncorrelated, since

$$\overline{i_3^* i_4} = \overline{(i_1^* - i_2^*)i_2} = -2eI_{ee}\Delta f \tag{25}$$

according to (7), (13), and (15), which is small in comparison with $\sqrt{\overline{i_3^2}\cdot\overline{i_4^2}}$ as long as I_{ee} is small.

These discussions show that the three equivalent circuits are interchangeable, except for the minor details just mentioned.

[5] Montgomery and Clark [40] gave the expression for $\overline{e_e^2}$ except for the (usually unimportant) factor $(I_e+2I_{ee})/(I_e+I_{ee})$ that was added by van der Ziel. They also gave the second term in $\overline{i^2}$; the first term was added by van der Ziel [53].

D. The Collective Approach

The operation of the junction diode and the junction transistor is based upon the injection and extraction of minority carriers. The flow of these carriers is by diffusion; they disappear sooner or later by recombination. These two processes are studied in detail in the collective approach. As mentioned in the beginning of Section II the diffusion and recombination of the minority carriers can be represented by a distributed RC network; for a one-dimensional model this corresponds to a distributed line without distributed inductance. That this is indeed the case is most easily seen by comparing the differential equation for a distributed line of series resistance R, parallel conductance G, and parallel capacitance C (all per unit length) with the differential equations describing a one-dimensional diffusion problem in which drift is negligible in comparison with diffusion.[6]

Let the minority carriers be holes with a charge e, a diffusion constant D_p, and a lifetime τ_p. Let p_n be the equilibrium hole concentration, p the total hole concentration, and $p'=p-p_n$ the excess hole concentration (all per unit length) and let i_p be the hole current. In the transmission line let E be the voltage on the line and I the current. Then, according to van der Ziel, the following correspondence holds between the diffusion problem and the transmission line problem: E corresponds to p', I corresponds to i_p, R corresponds to $1/(eD_p)$, G corresponds to (e/τ_p), and C corresponds to e. A diffusion problem can be solved now by first translating it into a transmission line problem, solving that by standard methods, and then translating back to the diffusion problem.

In this model the noise is caused by recombination fluctuations and diffusion fluctuations. In a section of length Δx the first effect can be represented by a fluctuating current Δi_{px} disappearing in the section Δx, whereas the diffusion fluctuations give rise to a fluctuating hole density in that section. Van der Ziel [55] showed that

$$\overline{\Delta i_{px}^2} = 2e^2\Delta f(p + p_n)\Delta x/\tau_p \tag{26}$$

$$\overline{\Delta p_x^2} = 4p\Delta f\Delta x/D_p. \tag{27}$$

Van der Ziel proved (26) from shot noise considerations and showed in an indirect manner that the expression for $\overline{\Delta p_x^2}$ had to have the form (27); otherwise a junction diode at zero bias would not give full thermal noise at all frequencies. Petritz [44a], [49a] derived expressions for these noise sources with the help of the Kolmogaroff-Fokker-Planck equation. He was able to give a rigorous proof of (27), whereas he obtained

$$\overline{\Delta_{ipx}^2} = 4e^2\Delta f p\Delta x/\tau_p \tag{26a}$$

instead of (26).[7] In the opinion of this author, the discrepancy between (26) and (26a) is due to the fact that the Kolmogaroff-Fokker-Planck equation has to be applied with caution to the recombination fluctuations if $p \neq p_n$, whereas the shot noise method remains fully applicable in that case.

The quantities Δi_{px} and Δp_x are, of course, independent, since they represent independent fluctuations. In the transmission line analogy Δp_x corresponds to a distributed series noise emf and Δi_{px} corresponds to a distributed parallel noise current generator. The fluctuations in the sections Δx can be treated as independent as long as Δx is large in comparison with the free path length of the carriers. The final result is obtained by adding the contributions of all sections Δx quadratically.

Solow [49a] has extended this treatment to two and three dimensions; his theory includes the effects of surface recombination velocity. He uses Petritz's bulk noise generators and derives surface generators with the help of the Kolmogaroff-Fokker-Planck equation.

A simple proof of (26) may be given with the help of the shot noise method; it starts from the diffusion equation

$$\frac{\partial p}{\partial t} = -\frac{p'}{\tau_p} - \frac{1}{e}\frac{\partial i_p}{\partial x}. \tag{28}$$

For stationary current flow $\partial p/\partial t = 0$ and hence, according to (28), a current $ep'\Delta x/\tau_p$ disappears between x and $(x+\Delta x)$ by recombination. Because of the equilibrium concentration p_n an additional current $ep_n\Delta x/\tau_p$ disappears for the same reason, which is balanced by the appearance of a current $ep_n\Delta x/\tau_p$ by pair generation. The total current *disappearing* in the section Δx is thus $ep\Delta x/\tau_p$ and the total current *appearing* in that section is $ep_n\Delta x/\tau_p$. Both currents should fluctuate independently and each should show full shot noise. The Fourier component Δi_{px} of this fluctuation therefore is given by

$$\overline{\Delta i_{px}^2} = 2e\left(\frac{ep\Delta x}{\tau_p}\right)\Delta f + 2e\left(\frac{ep_n\Delta x}{\tau_p}\right)\Delta f$$

which corresponds to (26).

Knowing the noise sources, it is not difficult to calculate the noise currents in the leads short-circuiting the electrodes and to prove (16) through (19). The equations thus are valid at all frequencies for a one-dimensional model in which all current is carried by holes.

Eqs. (16) through (19) do not give any reference to the model; therefore it was expected that they should be

[6] The fact that the problem is one-dimensional implies that the recombination must be *volume* recombination, not *surface* recombination.

[7] Petritz used (26a) and (27) for noise generators in his derivation of the diode result [44a]. By a slight rearrangement of terms his expression can be written as

$$\bar{i}^2 = 4kTG\Delta f - 2eI\Delta f\left(\frac{2G}{2G+G_0}\right). \tag{16a}$$

The difference between (16a) and (16) arises entirely from the difference in the expression for the recombination source [(26a) instead of (26)].

of general validity. The discussion of Section II-E shows that this is indeed the case.

E. Extension of the Corpuscular Theory [56]

The basic assumptions underlying the theory of Sections II-A and II-B will now be verified and the results of those sections extended to higher frequencies. To do so, the problem of current flow in *p-n* junctions has to be investigated in greater detail.

Consider first an *n*-type semiconductor sample with an ohmic contact. If electrons are injected into the material, space-charge neutrality will be re-established in a very short time, of the order of magnitude of the dielectric relaxation time of the material (about 10^{-12} seconds for germanium). This is achieved by a small displacement of the other electrons; at the same time electrons will leave through the ohmic contact to make the material externally neutral. If holes are injected into the material, space-charge neutrality will again be established in a very short time by a slight rearrangement of the electrons; at the same time electrons will enter through the ohmic contact to make the material externally neutral. The injected holes now spread out slowly by diffusion and disappear by recombination, but this does not cause any current in the external lead to the ohmic contact. In both cases, therefore, current occurs only at the instant that the carriers are injected into the material.

Now consider a *p-n* junction with two ohmic contacts. A very short current pulse will occur in the external circuit if a hole *enters* into the *n* region through the space-charge region or when a hole leaves the *n* region through that region. The two current pulses have opposite polarity and the displaced charge per pulse is $\pm e$. The duration of the pulse is determined by the diffusion time of the carriers through the space-charge region and is, in general, very short. Similar considerations hold for electrons entering or leaving the *n* region.

It is interesting to note that the individual current pulses are independent and that each pulse transfers a charge $\pm e$ in the external circuit. It is thus allowed to assign full shot noise to the various currents and, because the individual current pulses are so short, this should be the case for all frequencies of practical interest.

If the applied voltage is changed, then the minority carrier concentration at the boundaries of the transition region follows the applied voltage practically instantaneously, since the diffusion time through the transition region is so small. However, the subsequent diffusion of the minority carriers is a very slow process; it is responsible for the high-frequency behavior of the diodes and the transistors. To understand this high-frequency behavior, consider devices in which all current is carried by holes and split the carriers into different groups.

In a *p-n* junction diode one has to split the carriers into three groups[8] (Fig. 2).

Group 1: Holes flowing from the *p* region into the *n* region and recombining there. They give very short, random, and independent single current pulses and carry a current $(I+I_o)$; their contribution to $\overline{i^2}$ is therefore equal to the first term in (5) for all frequencies of practical interest. Moreover, because the rate of diffusion of the holes of Group 1 across the space-charge region follows the applied voltage practically instantaneously, this group gives a contribution $e(I+I_o)/kT$ to the junction admittance Y at all frequencies.

Group 2: Holes flowing from the *p* region into the *n* region and returning to the *p* region before having recombined. They give independent and random *double* current pulses, each consisting of two single, short current pulses of opposite polarity with the second one being delayed by a random delay time with respect to the first one. This group of holes is responsible for the high-frequency behavior.

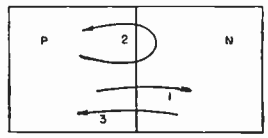


Fig. 2—The holes taking part in the conduction process in a junction diode are divided into three groups.

Group 3: Holes generated in the *n* region and diffusing into the *p* region; they give rise to very short, random, independent, single current pulses carrying a total current $(-I_o)$; thus they give a contribution equal to the second term of (5) for all frequencies of practical interest. This group of holes does not contribute to the diode admittance Y, since the current $(-I_o)$ is independent of the applied voltage.

The admittance Y thus consists of a part G_o caused by the holes of Group 1, an unknown part Y_2 due to the holes of Group 2, and a part $j\omega C_T$ due to the capacitance C_T of the space-charge region.[9] Putting

$$Y = G + jB = G_0 + Y_2 + j\omega C_T \tag{29}$$

we have

$$Y_2 = (G - G_0) + j(B - \omega C_T). \tag{29a}$$

[8] One might object that it is not known in advance whether a hole will belong to Group 1 or Group 2. It is sufficient for our argument that it will *either* belong to Group 1 *or* to Group 2. It is also unnecessary to describe processes in which the hole under discussion crosses the space-charge region several times. For, if a hole enters (or re-enters) the *p* region, another hole will leave through the ohmic contact to maintain space-charge neutrality; the hole can then no longer be distinguished from the other holes in that region.

[9] The applied ac voltage changes the width of the space-charge region periodically with time and the charge stored in that region varies in the same rhythm; the region thus acts as a capacitance. Because the charge transfer follows the applied voltage practically instantaneously, this effect gives a contribution $j\omega C_T$ (with constant C_T) to the admittance Y for all frequencies of practical interest.

The holes of Group 2 thus give a contribution $(G-G_o)$ to the diode conductance G. Without going into a detailed calculation, we see that the high-frequency behavior of the admittance is due to the holes of Group 2. These holes return to the p region by diffusion, which is a thermal process; the noise caused by these holes should thus be thermal noise of the conductance $(G-G_o)$, so that Group 2 gives a contribution $4kT(G-G_o)\Delta f$ to $\overline{i^2}$. Adding this to (5) we obtain (16) after substituting (4). By making a Fourier analysis of the random, independent current pulses of the individual holes of Group 2, van der Ziel and Becking proved in a rigorous manner that the holes of Group 2 give indeed the contribution $4kT(G-G_o)\Delta f$ to $\overline{i^2}$.

This result holds for all geometrical configurations and its validity does not depend upon the mode of recombination of the injected carriers.

The condition that the current is carried by holes may now be dropped; since the current carriers give independent pulses, (16) remains valid if part of the current is carried by electrons.

In p-n-p transistors one has to split the holes into 5 groups (Fig. 3):

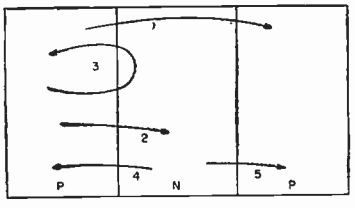


Fig. 3—The holes taking part in the conduction process in a junction transistor are divided into five groups.

1) Holes injected into the base region and collected by the collector.
2) Holes injected into the base region and recombining in that region with a free electron.
3) Holes injected into the base region and returning to the emitter.
4) Holes generated in the base region and collected by the emitter.
5) Holes generated in the base region and collected by the collector.

The validity of (17) and (18) already follows from the general validity of (16). A similar proof may also be given by making a careful analysis of the contributions of the holes of groups 1) to 5) to $\overline{i_1^2}$ and $\overline{i_2^2}$. A similar analysis also shows that the equations remain true if part of the current is carried by electrons.

Finally (19) has to be proved. We observe that only the holes of group 1) contribute to both the emitter current I_e and the collector current I_c and that no electrons contribute to both I_e and I_c. Hence only the holes of group 1) contribute to $\overline{i_1^* i_2}$ and to the signal transfer admittance Y_{ce}.

If a small ac voltage v_e is applied to the emitter, then the ac current in the short-circuited collector is $Y_{ce}v_e$ and the ac emitter current is $Y_e v_e$. The part $\alpha_o G_{eo} v_e$ of this emitter current comes from the holes of group 1); since the rate of diffusion of the holes of group 1) follows the ac emitter voltage practically instantaneously, this contribution to $Y_e v_e$ is the same for all frequencies of practical interest. At low frequencies the ac collector current follows the emitter voltage practically instantaneously, so that $Y_{ce}=G_{ceo}=\alpha_o G_{eo}$ as mentioned before. If all holes of group 1) had the same diffusion time τ through the base region, one would have:

$$Y_{ce} = G_{ceo}e^{-j\omega\tau}. \tag{30}$$

Because the diffusion of holes through the base region is a random process, there will be a distribution $h(\tau)d\tau$ in diffusion times and, as a consequence

$$Y_{ce} = \int_0^\infty G_{ceo}e^{-j\omega\tau}h(\tau)d\tau. \tag{30a}$$

This explains the decrease in $|Y_{ce}|$ at high frequencies.

Now turn to the cross-correlation $\overline{i_1^* i_2}$. Let i_{11} and i_{21} be the contributions of the holes of group 1) to i_1 and i_2, then $\overline{i_1^* i_2}=\overline{i_{11}^* i_{21}}$. Since individual current pulses are independent and since all holes of group 1) will ultimately pass *both* the emitter junction *and* the collector junction we have in analogy with (13)

$$\overline{i_{11}^2} = \overline{i_{21}^2} = 2e\alpha_0(I_e + I_{ee})\Delta f = 2kTG_{ceo}\Delta f. \tag{31}$$

If all holes had the same diffusion time through the base region, one would thus expect

$$\overline{i_1^* i_2} = \overline{i_{11}^* i_{21}} = \overline{i_{11}^2}e^{-j\omega\tau} = 2kTG_{ceo}\Delta f e^{-j\omega\tau}. \tag{32}$$

Introducing again the distribution $h(\tau)d\tau$ in diffusion times τ we thus obtain, by substituting (30a)

$$\overline{i_1^* i_2} = \int_0^\infty 2kTG_{ceo}\Delta f e^{-j\omega\tau}h(\tau)d\tau = 2kTY_{ce}\Delta f \tag{32a}$$

which is identical with (19).

F. *Validity of the Theory*

The above general proof of the validity of (16) through (19) is based upon the following (implicit and explicit) assumptions:

Assumption 1: It was explicitly assumed that the individual current pulses were independent and occurred at random. At high injection levels the injected carriers give rise to an appreciable space charge; this means that the above assumption is not satisfied under that condition, because space charge implies interaction between individual carriers. A violation of this assumption does not necessarily lead to a large deviation from (16) through (19).

Assumption 2: It was explicitly assumed that a single current pulse displaced a charge $\pm e$ in the external cir-

cuit. This cannot be correct if the space-charge region(s) of the junction(s) have an appreciable trap density. At low injection levels these traps are only partly filled so that part of the carriers diffusing through the space-charge region will get trapped there. Those that get trapped will not displace the full charge $\pm e$, so that the average charge displaced per pulse is less than this amount and the low-frequency noise is less than full shot noise. At higher injection levels practically all the traps are permanently filled and the noise equals full shot noise. Champlin has detected this effect in silicon diodes [7].

Assumption 3: The series resistance of the junctions, partly caused by the finite conductivity of the bulk material and partly due to the contact resistance of the junctions, was neglected. This effect is usually taken into account by introducing these resistances into the equivalent circuit and ascribing full noise to them [Fig. 1(a)–1(d)]. Moreover, these series resistances are strongly current dependent and this gives rise to interesting modulation effects that were discovered by Fonger [13]. The influence of these effects upon flicker noise is discussed in Section III, and upon shot noise in Section V.

As far as shot noise is concerned, we shall first neglect modulation effects altogether for two reasons. The first is that these effects have not been taken into account before, so that it is difficult to discuss earlier work when this effect is taken into account. Moreover, it will be shown that it is often warranted to neglect these effects for shot noise, because of peculiar coincidences.

To understand most of the earlier work, represent the shot noise by an emf e_e in series with the emitter and a current generator i in parallel with the collector junction (Fig. 4). As is easily seen,

$$i = i_2 - \alpha i_1 \qquad e_e = i_1 Z_e \tag{34}$$

which is the proper extension of (21a). The signal transfer properties of the transistor are now represented by the current generator αi_e, where $i_e = v_e/Z_e$ is the current flowing in the emitter junction. Furthermore, introduce the emitter impedance $Z_e = 1/Y_e$, the collector impedance $Z_c = 1/Y_c$, and the base impedance $Z_{b'b} = r_{b'b} + jX_{b'b}$ (to take into account that the base impedance may be complex at high frequencies); $r_{b'b}$ should show thermal noise. Fig. 4 is the high-frequency extension of Fig. 1(c).

Substituting (17) through (19), we have

$$\overline{i^2} = \overline{(i_2^* - \alpha^* i_1^*)(i_2 - \alpha i_1)} = 2e(I_c - |\alpha|^2 I_c)\Delta f$$
$$= 2e[(\alpha_0 - |\alpha|^2) I_e + I_{co}]\Delta f \tag{35}$$

and, if $I_e \gg I_{ee}$, so that $G_{eo} \simeq eI_e/kT$:

$$\overline{e_e^2} = \overline{i_1^2}\,|Z_e|^2 = 2kT(2G_e - G_{eo})\Delta f\,|Z_e|^2 \tag{36}$$

which are the high-frequency extensions of (21). Finally, if $I_e \gg I_{ee}$, we also have

$$\overline{e_e^* i} = 2kT\alpha(G_{eo} - Y_e^*)\Delta f. \tag{37}$$

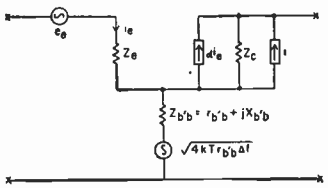


Fig. 4—Extension of Montgomery, Clark, and van der Ziel's circuit to higher frequencies.

The correlation is thus practically zero at lower frequencies but may have an appreciable value at higher frequencies.

III. Flicker Noise in Diodes and Transistors

We now turn to the problem of flicker noise. Its causes are not yet fully understood, though several investigators have worked on the problem. Fonger has discovered several noise sources and has found their proper place in the equivalent circuit. This allows a discussion of its effect in circuit applications. Here we follow Fonger's theory with some slight modifications.

A. General Characteristics of Flicker Noise [13]

According to Fonger there are two types of flicker noise, both with a low-frequency spectrum: surface noise and leakage noise.

We discuss surface noise first. It is now known that there are two types of energy levels at the surface of a semiconductor: "slow" states and "fast" states; the first act mainly as traps for the majority carriers and the latter as recombination centers for minority carriers [33]–[35]. The fluctuating occupancy of the slow states modulates the conductivity; this is the cause of flicker noise in bulk material. In addition, it modulates the capture cross section of the recombination centers; this is the cause of surface noise in diodes and transistors. The fluctuating current of minority carriers disappearing at the surface causes a fluctuating current to flow through the junction (or junctions) and modulates the series resistance of the junction (or junctions).

Leakage is caused by a thin conducting film bypassing the junction; it occurs at the perimeter of the junction and gives rise to a dc leakage current I_L and a leakage conductance g_L, which increase strongly with increasing bias. Spontaneous fluctuations in g_L cause leakage noise.

Surface noise is very sensitive to the ambient atmosphere; it is, *e.g.*, quite large in a humid atmosphere. It may be considerably reduced by proper surface treatment, it increases strongly with increasing current, and it is most prominent for junctions biased in the forward direction. Leakage noise is also very sensitive to the ambient atmosphere. Proper heat treatment can reduce it to such an extent that it becomes negligible for bias voltages less than a few volts; for that reason leakage noise is usually negligible for forward bias but may become quite important for large back bias.

Other studies, especially on diodes biased in the back direction, were carried out by Kennedy [31] and Mc-

Whorter [35], [38]; their results agree in general with Fonger's. McWhorter distinguishes between flicker noise and channel noise. A "channel" is a surface layer having a conductivity opposite to that of the bulk material. Channels formed in reverse bias diodes cause excess back current and a considerable increase in noise. An increased channel length increases the effective junction area (and hence the back current) and the length of its perimeter (and hence the leakage current); the back current increases linearly with the channel length due to both effects. Channel effects can also be diminished by proper surface treatments.

B. Flicker Noise in Diodes

The influence of modulation effects on the series resistance is discussed first. Consider a junction diode carrying a dc current I; let R be the dc series resistance of the junction. Because of the current dependence of R, the ac series impedance Z_s of the junction differs from the dc resistance R, since the ac current flowing through R will modulate R. We may thus split Z_s into a dc part R and a modulation part Z_{mb}. At low frequencies $Z_{mb}=R_{mb}$ is real and negative and

$$R_{mb}=I\frac{\partial R}{\partial I};\qquad Z_s=r=R+I\frac{\partial R}{\partial I},\tag{38}$$

where r is the ac resistance of the junction. R_{mb} is negative, since R decreases with increasing current; hence r is smaller than R. At high frequencies Z_{mb} (and hence Z_s) becomes complex; there are strong indications that Z_{mb} becomes inductive at high frequencies. At the frequencies where flicker noise is important, (38) may be used.

Now turn to the surface noise. Since the series resistance of the junction is strongly current dependent, fluctuations in the rate of generation and recombination of hole electron pairs at the surface will randomly modulate this resistance; because of the flow of dc current this modulation will show up as noise. Let this noise be described by a current generator i_s across the junction. In addition, fluctuations in the surface recombination rate will modulate the dc resistance in two ways:[10]

1) Directly. This is described by a noise emf e_b in series with R; e_b should be partly correlated with i_s.
2) Indirectly, through the current generator i_s. Fonger describes this by an additional emf i_sR_{mb} in series with R_{mb}. An equivalent representation connects the current generator i_s across both the junction impedance R_0 and the modulation impedance R_{mb}, as shown in Fig. 5(a); this demonstrates more clearly how the current generator i_s modulates the dc resistance R.

[10] This splitting of the modulation effect is, of course, somewhat arbitrary. It means that we try to take the modulation effect into account by relocating the current generator i_s. That part of the modulation effect that is not taken into account by this procedure is incorporated into the emf e_b. The sole justification of this noise schematic is that e_b happens to be very small for flicker noise.

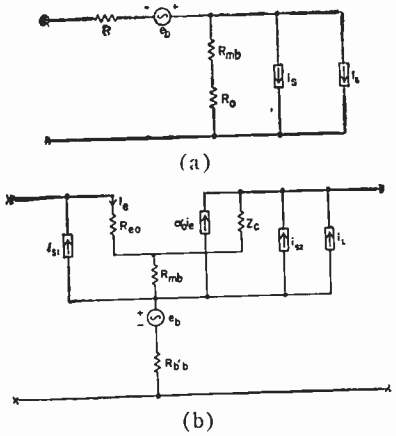


Fig. 5—Equivalent circuits for flicker noise. (a) Equivalent flicker noise circuit for a junction diode. (b) Equivalent flicker noise circuit of a transistor.

Since R_{mb} is negative at the frequencies of interest for flicker noise, the influence of the current generator i_s will be zero if $(R_o+R_{mb})=0$. Fonger found indeed that the low-frequency noise of a diode went through a deep minimum at a certain current I, in agreement with the above prediction; his results also indicated that the effect of the noise emf e_b was quite small and usually negligible. To enhance the modulation effect, Fonger used transistors with a large base resistance in diode connection (emitter and collector in parallel); then the effect became easily observable. For normal diodes, the effect is much less pronounced and may only be noticeable at very large currents [2].

The leakage noise can be described by a current generator i_L. It is not immediately clear where this current generator should be located. If the leakage noise modulates the dc resistance R as effectively as the surface noise, i_L should be connected across both R_o and the modulation resistance R_{mb}; if the leakage noise does not modulate the dc resistance R appreciably, the current generator i_L should be connected across R_o only. Fonger has tried to discriminate between these two possibilities by measuring leakage noise in transistors (see the following section) but his experiments were inconclusive. This means that it makes little difference either way which one of the two possibilities is chosen in practice. Fonger has chosen the second possibility and his approach is followed here. He finds that $\overline{i_L^2}$ is proportional to the square of the dc leakage current I_L; since I_L increases strongly with increasing back bias, the leakage noise can be considerably reduced by bringing the back bias closer to zero.

Fig. 5(a) gives the full equivalent circuit of the diode. As was said before, surface noise predominates strongly over leakage noise for forward bias; the current generator i_L may then be neglected. The leakage noise is only observable for diodes biased in the back direction. In that condition i_s is quite small (because the dc current is so small) and i_L usually predominates over i_s. The junction impedance for back bias is so high that resistance modulation effects may be neglected. In that case the

modulation resistance R_{mb} may be eliminated from the equivalent circuit, which completely removes the uncertainty in the location of the current generator i_L.

C. Flicker Noise in Transistors

The influence of modulation effects on the series resistances of the junctions is discussed first. Here the base resistance is strongly current dependent. Because of the resistance modulation, the ac base impedance $Z_{b'b}$ may be split into a dc part $R_{b'b}$ and a modulation part Z_{mb}. At low frequencies $Z_{mb}=R_{mb}$ is real and

$$Z_{mb} = R_{mb} = I_b \frac{\partial R_{b'b}}{\partial I_b};$$

$$Z_{b'b} = r_{b'b} = R_{b'b} + I_b \frac{\partial R_{b'b}}{\partial I_b}, \tag{39}$$

where $r_{b'b}$ is the ac base resistance. Since $R_{b'b}$ decreases with increasing I_b, $r_{b'b}<R_{b'b}$. At high frequencies Z_{mb}, and hence $Z_{b'b}$ becomes complex; in analogy with the diode case one would expect Z_{mb} to become inductive. For frequencies at which flicker noise is important, however, one may safely assume that $Z_{mb}=R_{mb}$.

We now turn to the noise and discuss first the surface noise component. The minority carriers disappearing at the surface give rise to two current generators i_{s1} and i_{s2} connected across the emitter and the collector junction. This noise source also modulates the dc base resistance $R_{b'b}$ in two ways: 1) directly, as described by the noise emf e_b; 2) indirectly, by means of the current generators i_{s1} and i_{s2}. Fonger describes this with the help of two noise emf's $i_{s1}R_{mb}$ and $i_{s2}R_{mb}$, but it is better represented by connecting i_{s1} across R_{eo} and R_{mb} and the current generator i_{s2} across Z_c and R_{mb}. The noise emf e_b is, of course, correlated with i_{s1} and i_{s2}; its influence is usually so small that it can be neglected.

The leakage noise can be described by a current generator i_L connected across the collector. If base resistance modulation is important for this noise source, i_L should be connected across Z_c and R_{mb}; if it is unimportant, i_L should be connected across Z_c only. We follow Fonger who has chosen the first possibility. The best transistors show negligible leakage noise for $|V_c|<10$ volts; to avoid leakage noise in poorer units, it is recommended that $|V_c|$ be kept considerably smaller.

The full equivalent circuit for flicker noise in transistors thus is as shown in Fig. 5(b). In this equivalent circuit the current generator $\alpha_o i_e$ is also connected across Z_c and R_{mb} and not across Z_c only, since this current generator should also modulate the base resistance; one would expect the current generators i_{s2}, i_L and $\alpha_o i_e$ to be connected in the same manner.

Now represent the flicker noise by an emf e_e in series with the emitter and a current generator i in parallel with the collector impedance Z_c. This corresponds to the equivalent circuit of Fig. 4, but with different values for i and e_e:

$$i = i_{s2} + \alpha_0 i_{s1} + i_L; \quad e_e = -i_{s1}(R_{e0} + R_{mb}) + e_b. \tag{40}$$

If e_b and i_L are negligible, i and e_e will be practically fully correlated. We return to this later.

In one respect there is a considerable difference between shot noise and flicker noise. In shot noise, even though the current generators i_1 and i_2 are strongly correlated, e_e and i are nearly uncorrelated, but in flicker noise e_e and i are strongly correlated. This difference comes about because i_1 and i_2 on the one hand, and i_{s1} and i_{s2} on the other hand, have an opposite phase relationship. One may also put it as follows. Shot noise mainly *circulates* through the transistor and the current generator i is the *difference* between i_2 and αi_1; flicker noise generated at the surface flows *from* the base surface towards both junctions and the current generator i is the *sum* of i_{s2} and $\alpha_o i_{s1}$.

IV. Experimental Verification of the Theory and Circuit Applications

Here the experimental data on diode noise and transistor noise are reviewed and the theory is applied to transistor circuits.

A. Semiconductor Diodes [2], [7], [8], [17], [18], [30], [35a], [43], [49]

We have already dealt with the work on low-frequency noise in some detail; now shot noise and application of the theory to photodiodes is discussed.

For shot noise we define the noise ratio n of the junction conductance as

$$\overline{i^2} = n \cdot 4kTG\Delta f. \tag{41}$$

Let $G_{oo}=eI_o/kT$ be the low-frequency junction conductance for zero bias and $G_o=e(I+I_o)/kT$, the low-frequency conductance of the biased diode. Substituting (16) into (41), we obtain

$$n = 1 - \frac{G_o}{2G} + \frac{G_{oo}}{2G}. \tag{41a}$$

For forward bias ($G_o>G_{oo}$) the junction conductance has a noise ratio n varying between $\frac{1}{2}$ and 1, going to unity for very high frequencies, since $G>G_o$ in that case. Assuming full thermal noise of the ac series resistance r thus gives that the noise ratio of the whole device is between $\frac{1}{2}$ and 1; this agrees roughly with Anderson and van der Ziel's early data [2].

Probably the most accurate measurements were made by Champlin who used an ac bridge circuit with the junction in one arm, a variable RC network in the other arm, and noise diodes connected across the junction and across the RC network [7], [8]. First, the RC network was adjusted so that the bridge circuit was balanced at the frequency at which the noise measurement was to be performed and then the saturated current of one of the noise diodes was adjusted so that the two arms of the bridge circuit gave equal amounts of noise power.

In most cases he observed full thermal noise of the ac series resistance plus full shot noise of the junction. The frequency dependence of the junction noise agreed well with (41a). In some silicon junction diodes at low currents the noise was less than full shot noise; the cause of that result was mentioned previously (see Section II-F).

In p-n junctions used as photoelectric cells the junction is biased in the back direction. The light then generates hole-electron pairs; the holes are collected by the p region and the electrons by the n region. Full shot noise should be associated with this photocurrent; in addition, some low-frequency noise will be generated, but this may not be very large if the surface is properly treated. Shot noise in semiconductor photoelectric cells has been observed by Slocum and Shive [49] and by Pearson, Montgomery, and Feldmann [43]. The latter group found shot noise down to 80 cycles in a dry atmosphere. In a humid atmosphere the noise was low-frequency noise and the noise power at 100 cycles was a factor 3×10^5 above shot noise. This shows the importance of the ambient atmosphere.

In p-n junction photocells operated under open-circuited condition the photovoltage biases the junction in forward direction to such an extent that the forward current exactly balances the photocurrent I. Gianola [16] investigated silicon photovoltaic cells and found low-frequency noise at low frequencies; the mean square value $\overline{e^2}$ of the open-circuit noise voltage was proportional to the photocurrent I for small values of I and inversely proportional to I for large values of I. His result is understandable if for the flicker noise the current generator i (i_s or i_L, probably i_s) has a mean square value $\overline{i^2}$ that is proportional to I. If R_o is the internal resistance of the cell, then $\overline{e^2}=\overline{i^2}R_o^2$. For small I the resistance R_o is independent of I and hence $\overline{e^2}$ is proportional to I in that case; for larger values of I the resistance R_o varies as $1/I$ (since $R_o\simeq kT/eI$) and hence $\overline{e^2}$ varies as $1/I$. It is shown in the next section that the proportionality of $\overline{i^2}$ to the current I occurs more often.

Hyde [29] has found spectra of the form const $/(1+\omega^2\tau^2)$ in point contact diodes under reverse bias conditions. Such spectra are probably due to traps; they might also be expected for some types of junction diodes.

An interesting noise phenomenon associated with avalanche breakdown is found in silicon p-n junctions biased in the back direction. The breakdown seems to occur at very tiny discharge spots (microplasmas) that emit light. The noise is generated as pulses, many millivolts high, occurring at random and at a rate that depends very strongly upon the current [9], [10], [36], [37], [45].

B. Shot Noise in Transistors

The validity of the shot noise theory may be tested by verifying the equivalent circuit of Fig. 4, that is, by determining $\overline{i^2}$, $\overline{e_e^2}$, and the cross correlation $\overline{e_e^*i}$. Extensive tests have been carried out by Nielsen, by Guggenbuehl and Strutt, and by Hanson and van der Ziel and others [6], [19]–[21], [23]–[28], [41], [50].

One way consists in expressing the noise figure F in terms of $\overline{i_1^2}$, $\overline{i_2^2}$, and $\overline{i_1^*i_2}$. This was done by Guggenbuehl and Strutt [23], who found

$$F=\frac{e}{2kTR_s}\left[\frac{I_e}{|\alpha|^2}|Z_s+r_{b'b}+Z_e|^2-I_e|Z_s+r_{b'b}|^2\right]. \quad (42)$$

The merit of this equation is that it expresses the noise in terms of the macroscopic parameters of the transistor. However, if one finds deviations between theory and experiment, detection of the source of the discrepancy is not so easy. In that case it is better to introduce with Hanson and van der Ziel [28] the noise conductance g_{s1}, the noise resistance R_{s1}, and the correlation impedance $Z_{sc}=(R_{sc}+jX_{sc})$ as follows. First e_e is split into a part e_e' that is uncorrelated with i and a part e_e' that is fully correlated with i. One then defines

$$\frac{\overline{i^2}}{|\alpha|^2}=4kTg_{s1}\Delta f;\qquad \overline{e_e'^2}=4kTR_{s1}\Delta f;$$
$$Z_{sc}=\frac{\alpha e_e'}{i}=\frac{\overline{\alpha e_e i^*}}{\overline{i^2}}, \quad (43)$$

according to (34). Substituting (17) through (19) into (35) one obtains

$$g_{s1}=\frac{e}{2kT}\left[\frac{(\alpha_0-|\alpha|^2)I_e+I_{co}}{|\alpha|^2}\right]. \quad (43a)$$

At low frequencies, the noise resistance R_{s1} for $I_e\gg I_{co}$ is

$$R_{s1}=\tfrac{1}{2}R_{eo}, \quad (43b)$$

where R_{eo} is the low-frequency emitter resistance.

Calculating the noise figure F and expressing it in terms of g_{s1}, R_{s1}, Z_{sc}, and the other transistor parameters, we obtain

$$F=1+\frac{(r_{b'b}+R_{s1})}{R_s}+\frac{g_{s1}}{R_s}|Z_s+Z_e+Z_{b'b}+Z_{sc}|^2, \quad (44)$$

where R_s is the resistive part of the source impedance Z_s.

Nielsen neglects the correlation between e_e and i, which amounts to putting $Z_{sc}=0$, and further assumes that $R_{s1}\simeq\frac{1}{2}R_{eo}$ at all frequencies. One then obtains, if $Z_{b'b}=r_{b'b}$,

$$F=1+\frac{(r_{b'b}+\frac{1}{2}R_{eo})}{R_s}+\frac{g_{s1}}{R_s}|Z_s+Z_e+r_{b'b}|^2. \quad (44a)$$

Nielsen found reasonable agreement between theory and experiment and showed how the theory could be used for the design of low-noise transistor circuits. This indicates that his approximations were warranted; we shall see later why that should be the case.

Expression (44a) shows that the frequency depend-

ence of the noise figure comes mainly from the frequency dependence of g_{s1}. To understand this, we substitute $\alpha=\alpha_0/(1+jf/f_0)$; (43a) then becomes

$$g_{s1}=\frac{e}{2kT}\left[\frac{(\alpha_0 I_e+I_{co})(1+f^2/f_0^2)-\alpha_0^2 I_e}{\alpha_o^2}\right] \tag{45}$$

from which it follows that g_{s1} has increased by a factor 2 at the frequency $f=f_o\sqrt{1-\alpha_o}$. The noise figure of the transistor is thus reasonably constant for $f<f_o\sqrt{1-\alpha_o}$ and increases rapidly with increasing frequency for $f>f_o\sqrt{1-\alpha_o}$. For frequencies below the α-cutoff frequency f_o, the quantity g_{s1} increases with increasing frequency mainly because of the frequency dependence of the term $(\alpha_o-|\alpha|^2)$ in (43a); for frequencies above f_o, the quantity g_{s1} increases mainly because the factor $|\alpha|^2$ in the denominator of (43a) goes to zero. In order to obtain good noise figures at high frequencies, it is important to use transistors with a high α-cutoff frequency.

Guggenbuehl and Strutt's (42) takes the correlation between e_e and i into account. The source reactance X_s should now be chosen such that F is a minimum. In (44a) this is the case if $X_s=-X_e$; in (44) the expression

$$\frac{I_c}{|\alpha|^2}(X_s+X_e)^2-I_e X_s^2 \tag{46}$$

should be made a minimum. Differentiation shows this to be the case if

$$X_s=-\frac{I_c}{(I_c-|\alpha|^2 I_e)}X_e. \tag{46a}$$

Substitution of (46a) into (42) shows that the correlation should have an appreciable effect on the noise figure at higher frequencies. This is discussed later.

Guggenbuehl and Strutt also found reasonable agreement between theory and experiment [23]. They noticed the strong increase in noise figure with increasing frequency and found, for constant source resistance R_s, that they had to go to very low values of I_c to attain minimum noise figures at high frequencies. At least a major part[11] of their results can be explained by the frequency dependence of g_{s1} and by the facts that g_{s1} decreases with decreasing I_c whereas R_e and R_{s1} increase with decreasing I_c. The minimum noise figure thus occurs at the value of I_c where any further decrease in F due to g_{s1} is offset by the increase in F due to R_e and R_{s1}.

We now turn to Hanson and van der Ziel's results [28]. Starting with (44) and considering F as a function of the source reactance X_s, F is a minimum if

$$X_s+X_e+X_{b'b}+X_{sc}=0 \tag{47}$$

in which case

$$F=1+\frac{(r_{b'b}+R_{s1})}{R_s}+\frac{g_{s1}}{R_s}(R_s+R_e+r_{b'b}+R_{sc})^2$$
$$=A+\frac{B}{R_s}+CR_s. \tag{48}$$

The quantities A, B, and C may be determined experimentally by measuring F as a function of R_s; calculating from (48) we also have

$$A=1+2g_{s1}(R_e+r_{b'b}+R_{sc});$$
$$B=r_{b'b}+R_{s1}+g_{s1}(R_e+r_{b'b}+R_{sc})^2;\quad C=g_{s1}. \tag{48a}$$

Having deduced the values of A, B, and C from the measurements, we may express the following quantities in terms of A, B, and C:

$$(R_e+r_{b'b}+R_{sc})=\frac{(A-1)}{2C}$$
$$(r_{b'b}+R_{s1})=\frac{4BC-(A-1)^2}{4C}$$
$$g_{s1}=C. \tag{49}$$

Unfortunately, it often happens that $(A-1)$ is only small and/or that $4BC$ and $(A-1)^2$ differ relatively little. In that case the quantities $(R_e+r_{b'b}+R_{sc})$ and $(r_{b'b}+R_{s1})$ are only inaccurately known and hence the values of R_{s1} and R_{sc} can only be determined inaccurately too.

Theoretically, [6] R_{sc} is zero at low frequencies and passes through a maximum for higher frequencies. The quantity X_{sc} should have a rather broad maximum around $f=f_o\sqrt{1-\alpha_o}$ and should have an appreciable value at those frequencies, so that tuning for minimum noise figure should give a marked noise figure improvement. Finally, R_{s1} should be of the order of $\frac{1}{2}R_e$ at low frequencies and should decrease rapidly with increasing frequency.

Hanson and van der Ziel [28] found that the experimental values of g_{s1} agreed very well with the theoretical expectations. In most transistors the experimental values of R_{s1} and R_{sc} agreed with the theoretical values within the (rather large) limits of experimental error. In some transistors with a low value of α_o the value of R_{sc} differed markedly from zero; this effect might possibly be attributed to base modulation (Section V). Moreover, they found that only little improvement in noise figure could be obtained by properly adjusting the source reactance X_s; this probably indicates that the experimental value of X_{sc} was considerably smaller than the expected theoretical value. This might either be caused by the reactive component of the base impedance $Z_{b'b}$ or by base modulation effects (Section V).

Now we may also understand why Nielsen could ignore the correlation effect. If X_s is adjusted for minimum noise figure, then the quantity R_{sc} in A and B may be omitted if R_{sc} is small in comparison with $(R_e+r_{b'b})$.

[11] Guggenbuehl and Strutt maintain that the frequency dependence of g_{s1} cannot *fully* explain the observations at high injection levels. According to Section II-F, deviations between theory and experiment are not impossible at high injection levels, but no quantitative theory exists at present.

This is especially true at low frequencies, where the terms $2g_{s1}(R_e+r_{b'b}+R_{sc})$ and $g_{s1}(R_e+r_{b'b}+R_{sc})^2$ are often quite small; even at higher frequencies, where these two terms are larger, the omission of R_{sc} from the equations may not cause too large an error. Moreover, if $R_{s1}<r_{b'b}$, it does not make much difference whether one puts $R_{s1}=\frac{1}{2}R_{eo}$ or a smaller value. For good information on R_{sc} and R_{s1} we thus need accurate measurements of A, B, and C.

These can be achieved as follows. The quantities B and C may be determined accurately by a direct method. Since the total noise resistance R_n of the whole circuit is equal to FR_s, the quantity B corresponds to the noise resistance R_{no} for zero source impedance

$$B = R_{no} = r_{b'b} + R_{s1} + g_{s1}(R_e + r_{b'b} + R_{sc})^2. \quad (50)$$

The quantity C is related to the equivalent input saturated diode current I_n for large source impedance R_s. If we define the equivalent input saturated diode current I_n by an equivalent current generator $\sqrt{\overline{i_n^2}}=\sqrt{2eI_n\Delta f}$ in parallel to R_s, then obviously

$$\overline{i_n^2} = 2eI_n\Delta f = F\cdot 4kT\Delta f/R_s;$$

or

$$I_n = \frac{2kT}{e}\left[\frac{A}{R_s}+\frac{B}{R_s^2}+C\right]. \quad (51)$$

For large values of R_s, we have $I_n=I_{n\infty}$ independent of R_s, or

$$I_{n\infty} = \frac{2kT}{e}C, \quad \text{or} \quad C = \frac{e}{2kT}I_{n\infty}. \quad (51a)$$

The quantity $(A-1)$ may finally be determined from the minimum noise figure $F_{\min}$. According to (48) the minimum noise figure $F_{\min}$ is attained if $R_s=\sqrt{B/C}$, in which case

$$F=F_{\min}=A+2\sqrt{BC}; \text{ or } (A-1)=(F_{\min}-1-2\sqrt{BC}). \quad (52)$$

The accuracy with which $(A-1)$ can be determined, depends upon the difference between $F_{\min}$ and $(1+2\sqrt{BC})$.

As mentioned before, the measurement of the noise conductance g_{s1} allows the determination of whether (35) is correct. This equation can also be checked very accurately by inserting a large impedance in series with the emitter (input open) and determining the equivalent output saturated diode current I_{eq} of this circuit. To do so, one connects a noise diode in parallel to the output and determines the diode current for which the output noise power is doubled. According to Fig. 4 we have

$$\overline{i^2}|Z_c|^2 + 4kTr_{b'b}\Delta f = 2eI_{eq}\Delta f|Z_c + r_{b'b}|^2. \quad (53)$$

Unless the frequency is very high, it may be assumed that $|Z_c|$ is large in comparison with $r_{b'b}$. Retaining only the terms in Z_c^2 in that case, we have

$$\overline{i^2} = 2eI_{eq}\Delta f \quad (54)$$

or, substituting (35)

$$I_{eq} = (\alpha_0 - |\alpha|^2)I_e + I_{co}. \quad (55)$$

If I_{co} is very small, then this equation tells us that I_{eq} is equal to $(1-\alpha_0)I_e$ for small frequencies, practically twice as large for $f=f_o\sqrt{1-\alpha_0}$, equal to $\frac{1}{2}I_e$ at the α-cutoff frequency f_o, and equal to I_e above the cutoff frequency. If I_e is quite small, I_{eq} should be equal to I_{co} at all frequencies. All these predictions were well verified by Hanson and van der Ziel's measurements; their results indicated that the measurement of I_{eq} as a function of frequency might be used to determine the α-cutoff frequency f_o [28].

We note that the quantities I_{eq} and $I_{n\infty}$ are closely related, since[12]

$$g_{s1} = C = \frac{e}{2kT}I_{n\infty} = \frac{e}{2kT}\frac{I_{eq}}{|\alpha|^2},$$

we have

$$I_{eq} = I_{n\infty}|\alpha|^2. \quad (56)$$

Hanson and van der Ziel found good agreement between the experimental values of g_{s1} and I_{eq} at relatively low frequencies.

At high frequencies, Hanson and van der Ziel sometimes found values of I_{eq} that differed from these predictions. Some transistors, for example, had $I_{eq}<I_e$ for large currents and $I_{eq}>I_e$ for small current at frequencies at which one would expect $\overline{i^2}=2eI_e\Delta f$. This could be attributed to the fact that $|Z_c|$ was no longer large in comparison with $r_{b'b}$; calculating $\overline{i^2}$ from (53), using the observed value of I_{eq}, it was found that $\overline{i^2}\simeq 2eI_e\Delta f$ even in this case. This apparent deviation between theory and experiment is thus caused by the fact that the output terminals are not connected directly to Z_c but are connected through the base resistance $r_{b'b}$ that also has noise associated with it.

In the case discussed by Guggenbuehl and Strutt [23] it was important to go to low emitter current. Hanson [27] has reported a few cases in which the capacitive feedback between emitter and collector gave a considerable increase in the apparent current amplification factor of the device.[13] This resulted in a considerable decrease in the noise figure that could be greatly reduced by going to *larger* emitter currents.

A similar condition occurs in drift transistors. Here the α-cutoff frequency increases with increasing emitter current; for high-frequency applications of drift transistors, the emitter current therefore should not be chosen too small.

[12] This means that the simultaneous measurement of I_n and I_{eq} might be used to determine the value of $|\alpha|^2$ under the exact operating conditions of the circuit; this sometimes may be useful.

[13] It is thus important to screen input and output of transistor amplifiers at high frequencies.

C. Flicker Noise in Transistors [1], [3], [5], [13], [39], [58], [61]

As was already mentioned in Section III, one has to discriminate between surface noise and leakage noise. Both have a low-frequency spectrum. We saw that leakage noise was most easily reduced by bringing the collector bias closer to zero; by giving the device the proper treatment, the effect can be reduced still further. Surface noise could also be reduced by proper surface treatment.

Another interesting feature of surface noise is the difference between *p-n-p* and *n-p-n* transistors; the latter show considerably more flicker noise than the former [5], [6a]. This reflects differences in the physical characteristics of the surface layer of the base regions of the two types of transistors. The reduction of surface noise in *p-n-p* transistors has apparently proceeded farther than in *n-p-n* transistors.

Another interesting point is that the current dependence of the surface noise resembles the current dependence of shot noise. Measuring the equivalent output saturated diode current I_{eq} with open input, one has, for frequencies where shot noise predominates,

$$I_{eq} = (I_{eq})_s = 2I_e\alpha_0(1 - \alpha_0)\Delta f \tag{57}$$

if the collector saturated diode current I_{co} is negligible. This dependence is especially characteristic for tetrode transistors where α_o can be changed over a wide range by changing the current bias of the base region. Yajima [61] showed that a similar expression holds for these transistors at frequencies where flicker noise predominates

$$(I_{eq})_f = \frac{\text{const}}{f} I_e\alpha_0(1 - \alpha_0)\Delta f. \tag{58}$$

He concluded from this result that flicker noise in junction transistors arises at least partly from the recombination process of injected carriers in the base and nearby surface. This agrees with Fonger's ideas.

For triode transistors for which (58) is also valid, $(I_{eq})_f$ should be proportional to I_e, since α_o is practically independent of I_e in that case. Many transistors indeed show such a relationship, but deviations do occur in some types.

In (40) the flicker noise was represented by an input emf e_e and an output current generator i. Neglecting leakage noise, which is allowed in good units, and ignoring the noise emf e_b, which is also allowed according to Fonger, gives that e_e and i are practically fully correlated. Splitting e_e into a part e_e'' that is uncorrelated with i and a part e_e' that is fully correlated with i, and representing the flicker noise by an equivalent emf in series with the input yields:

$$\begin{aligned} e_{nf} &= e_e + \frac{1}{\alpha_o}(Z_s + R_{eo} + r_{b'b}) \\ &= e_e'' + \frac{i}{\alpha_o}\left(Z_s + R_{eo} + r_{b'b} + \frac{\alpha_o i_e'}{i}\right). \end{aligned} \tag{59}$$

We now define the flicker noise resistance R_{nf} by the equation $\overline{e_{nf}^2} = 4kTR_{nf}\Delta f$; furthermore we introduce three constants, the emitter noise resistance R_{f1}, the flicker noise conductance g_{f1}, and the correlation resistance R_{fc} by

$$\overline{e_e''^2} = 4kTR_{f1}\Delta f; \quad \frac{\overline{i^2}}{\alpha_o^2} = 4kTg_{f1}\Delta f; \quad R_{fc} = \frac{\alpha_o e_e'}{i} \tag{60}$$

as a characterization of the flicker noise properties of the transistor: R_{f1} and g_{f1} should vary as $1/f$ and, in view of what was said above, g_{f1} is proportional to I_e. According to (40)

$$R_{fc} = -(R_{eo} + R_{mb})\frac{\alpha_o i_{s1}}{(i_{s2} + \alpha_0 i_{s1})}. \tag{60a}$$

Substituting into (59), we have

$$R_{nf} = R_{f1} + g_{f1}\left|Z_s + R_{eo} + r_{b'b} + R_{fc}\right|^2. \tag{61}$$

For minimum flicker noise resistance, the source reactance X_s should be chosen such that

$$X_s^2 \ll (R_{eo} + r_{b'b} + R_{fc})^2. \tag{61a}$$

It is thus not sufficient that X_s is small in comparison with the input resistance of the transistor circuit, which may be quite large if feedback is applied; one has to satisfy (61a). If a source is thus capacitively coupled to the input of a low-noise (audio) transistor amplifier, the coupling capacitance should be chosen sufficiently large.

The emitter noise resistance R_{f1} is zero if e_e and i are fully correlated for flicker noise. If $\sqrt{R_{nf}}$ is then plotted as a function of R_s, one should obtain a straight line. Chenette [5] has carried out the experiment. Measuring R_{nf} as a function of the source resistance R_s, he found that the linear relationship was well satisfied; this indicates that e_e and i are indeed practically fully correlated and that $R_{f1} \simeq 0$. The straight line intercepts the zero axis at the point

$$\begin{aligned} R_s &= -(R_{eo} + r_{b'b} + R_{fc}) \\ &= -\left[R_{b'b} + (R_{eo} + R_{mb})\frac{i_{s2}}{(i_{s2} + \alpha_o i_{s1})}\right] \end{aligned} \tag{62}$$

since R_{fc} is given by (60a) and $r_{b'b} = (R_{b'b} + R_{mb})$ according to Section III. The values for R_s observed by Chenette roughly agree with the theoretical expectations.

D. Applications to Low-Noise Circuits [3], [12], [19], [32], [39], [47], [50], [58], [59]

Most of our circuit discussions held for a grounded base circuit. It is easily shown, however, that *a grounded base and a grounded emitter circuit have the same noise figure F if the source impedances used in the two circuits are identical* [23]. The grounded emitter circuit usually is recommended, since it allows a much simpler interstage coupling and a much higher gain per stage.

As mentioned before, the noise figure due to shot noise is reasonably flat for frequencies up to $f_o\sqrt{1-\alpha_o}$, where f_o is the α-cutoff frequency. It therefore is impor-

tant to use transistors with high cutoff frequency in low-noise applications.

Let us first consider the low-frequency case and let us further assume that $R_{s1}=\frac{1}{2}R_{eo}$, $R_{sc}\simeq 0$, and $R_{f1}\simeq 0$. The total noise resistance of the circuit, according to (48) and (61), is then

$$R_n = R_s + r_{b'b} + \tfrac{1}{2}R_{eo} + g_{s1}(R_s + R_{eo} + r_{b'b})^2 + g_{f1}(R_s + R_{eo} + r_{b'b} + R_{fc})^2. \quad (63)$$

In these equations g_{f1} and g_{s1} usually depend linearly on I_e, whereas R_{eo} is inversely proportional to I_e.

First consider the case where flicker noise predominates. For minimum noise resistance at a given value of the source impedance R_s, the last term in (63) should then be made a minimum. For $R_s \gg (R_{eo}+r_{b'b}+R_{fc})$ this leads to the condition that g_{f1} must be as small as possible, which is the case for very small emitter currents. For $R_s \ll (R_{eo}+r_{b'b}+R_{fc})$ the quantity $g_{f1}(R_{eo}+r_{b'b}+R_{fc})^2$ should be made a minimum; experimentally it is found that this also leads to quite small emitter currents.

If shot noise predominates and R_s is given, one should make $[(R_s+r_{b'b}+\frac{1}{2}R_{eo})+g_{s1}(R_s+r_{b'b}+R_{eo})^2]$ a minimum. We observe that the first term predominates over the second one if $g_{s1}(R_s+r_{b'b}+R_{eo})<1$. For that reason R_n does not depend so strongly upon R_s as in the previous case, unless R_s is large. For very large R_s one should make g_{s1} as small as possible; this leads again to very small values of I_e. If $g_{s1}(R_s+r_{b'b}+R_{eo})\ll 1$, one makes R_{eo} as small as possible; this leads to relatively large values of I_e. For low source impedances R_s the requirements for low flicker noise resistances and low shot noise resistances are thus opposite.

Low-noise p-n-p transistors, operating at emitter currents of about 0.5 ma and fed from a source with low source resistance R_s ($R_s<50$ ohms), may have noise resistances as low as 1000 ohms at 10 cycles. At higher frequencies, where shot noise predominates, the noise resistance may be as low as 100–150 ohms. This is considerably better than vacuum pentodes, which have noise resistances of about 10^5–10^6 ohms at 10 cycles and about 10^3 ohms at frequencies where shot noise predominates. Low-noise audio and subaudio amplifiers, operating from a source of low impedance, should thus use low-noise transistors instead of vacuum tubes.

For signal sources with very large source impedances R_s, the vacuum tube is much better. The reason is that the noise resistance R_{nt} of the tube is independent of the source impedance, whereas the noise resistance of the transistors varies as R_s^2 for large R_s. The total noise resistance of the vacuum tube circuit is thus (R_s+R_{nt}), whereas the total noise resistance of the transistor circuit is approximately $[R_s+(g_{s1}+g_{f1})R_s^2]$ for large R_s. Hence if:

$$R_s > \sqrt{\frac{R_{nt}}{g_{s1}+g_{f1}}}$$

the vacuum tube is better than a transistor. Nevertheless, in some applications it may happen that the transistor amplifier, though inferior to the vacuum tube amplifier, is still adequate for the purpose for which it is used.

Volkers and Pedersen [58] were the first to notice the low noise resistance of the transistor for low source impedances. The units with which they worked had considerable leakage noise; to eliminate this, they had to operate the transistor at nearly zero bias. They called this mode of operation the "hushed" operation. In modern transistors the leakage noise is much smaller so that the collector bias does not have to be chosen so close to zero; it is still important, however, not to make $|V_c|$ too large.

We now consider the case where the circuit can be adjusted for minimum noise figure. If flicker noise predominates strongly and if F_{shot} is the noise figure due to shot noise,

$$F = F_{\text{shot}} + \frac{R_{nf}}{R_s} = F_{\text{shot}} + g_{f1}\frac{(R_s + R_{eo} + r_{b'b} + R_{fc})^2}{R_s} \quad (64)$$

which has its minimum value if the last term is a minimum,

$$F_{\min} = F_{\text{shot}} + 4g_{f1}(R_{eo} + r_{b'b} + R_{fc})^2$$

for

$$(R_s)_{\min} = (R_{eo} + r_{b'b} + R_{fc}. \quad (64a)$$

In Section IV-C the plot showing R_{nf} as a function of R_s intercepted the zero axis at $R_s=-(R_s)_{\min}$. Considered as a function of I_e, the minimum noise figure is smallest if I_c is quite small.

For frequencies where shot noise predominates, the value of F was given by (48) and its minimum value $F_{\min}$ was shown in (52). Substituting the values of A, B, and C given by (48a), putting $R_{s1}=\frac{1}{2}R_{eo}$, and neglecting R_{sc} yields

$$F_{\min} = 1 + 2g_{s1}(R_{eo} + r_{b'b}) + 2\sqrt{g_{s1}(\tfrac{1}{2}R_{eo} + r_{b'b}) + g_{s1}^2(R_{eo} + r_{b'b})^2} \quad (65)$$

for

$$R_s = \frac{1}{g_s}\sqrt{g_{s1}(\tfrac{1}{2}R_{eo} + r_{b'b}) + g_{s1}^2\,(R_{eo} + r_{b'b})^2}. \quad (65a)$$

Depending on the relative magnitude, one thus has to minimize $g_{s1}(\frac{1}{2}R_{eo}+r_{b'b})$ or $g_{s1}(R_{eo}+r_{b'b})$; since these conditions do not differ so strongly, we shall here minimize $g_{s1}(\frac{1}{2}R_{eo}+r_{b'b})$. This is easily done, for g_{s1} and R_{eo} depend upon the emitter current I_e in an opposite manner; g_{s1} decreases with decreasing I_e, whereas $R_{eo}\simeq kT/eI_e$ increases. We have from (43a)

$$g_{s1}(\tfrac{1}{2}R_{eo} + r_{b'b}) = \frac{1}{4}\left(\frac{1-\alpha_0}{\alpha_0}\right)\left[I_e + \frac{I_{co}}{\alpha_0(1-\alpha_0)}\right]\cdot\left[\frac{1}{I_e} + \frac{2er_{b'b}}{kT}\right]. \quad (66)$$

Assuming α_o to be independent of I_e, this has a minimum value.

$$\frac{1}{4}\left(\frac{1-\alpha_o}{\alpha_o}\right)\left[1+\sqrt{\frac{I_{co}}{\alpha_o(1-\alpha_o)}\cdot\frac{2er_{b'b}}{kT}}\right]^2 \qquad (66a)$$

for

$$I_e=\sqrt{\frac{I_{co}}{\alpha_0(1-\alpha_0)}\cdot\frac{kT}{2er_{b'b}}}. \qquad (66b)$$

Substituting into (65) yields

$$F_{\min}\simeq 1+2\sqrt{g_{s1}\left(\tfrac{1}{2}R_{eo}+r_{b'b}\right)}$$

$$=1+\sqrt{\frac{1-\alpha_o}{\alpha_o}}+\sqrt{\frac{I_{eo}}{\alpha_0^2}\cdot\frac{2er_{b'b}}{kT}}. \qquad (67)$$

Taking, for example, $I_{co}\simeq 1$ μa, $\alpha_o=0.98$, and $r_{b'b}=100$ ohms, we have a minimum noise figure of about 1.25 at $I_e\simeq 80$ μa. This shows that rather low noise figures can be obtained by proper choice of the source impedance and the emitter current.[14]

The design conditions for low-noise transistors are that the quantities $(1-\alpha_o)$ and $I_{co}r_{b'b}$ should be made as small as possible and that $(2eI_{co}r_{b'b}/kT)<(1-\alpha_o)$.

Next we investigate the noise figure close to the cutoff frequency and determine how R_s and I_e should be chosen in order to make the noise figure F a minimum. The minimum noise figure is again given by (65), if we neglect the correlation resistance R_{sc} (which is probably allowed). Furthermore, we assume that $g_{s1}(R_{s1}+r_{b'b})\ll g_{s1}^2(R_e+r_{b'b})^2$ and take into account that $R_e\simeq R_{eo}$ up to the cutoff frequency. In that case

$$F_{\min}\simeq 1+4g_{s1}(R_{eo}+r_{b'b}) \text{ for } R_s\simeq(R_{eo}+r_{b'b}). \qquad (68)$$

We now minimize this expression as a function of the emitter current I_e. In a good transistor $\alpha_o\simeq 1$ so that $|\alpha|^2\simeq\frac{1}{2}$ at the cutoff frequency. One then has at that frequency

$$g_{s1}=\frac{e}{2kT}(I_e+2I_{eo}); \qquad R_{eo}\sim\frac{kT}{eI_e}$$

so that

$$F_{\min}\simeq 1+2(I_e+2I_{eo})\left(\frac{1}{I_e}+\frac{er_{b'b}}{kT}\right) \qquad (68a)$$

at that frequency. This has to be minimized as a function of I_e. The minimum value is:

$$F_{\min}\simeq 1+2\left(1+\sqrt{2I_{eo}\cdot\frac{er_{b'b}}{kT}}\right)^2$$

for

$$I_e=\sqrt{2I_{eo}\cdot\frac{kT}{er_{b'b}}}. \qquad (69)$$

As in our previous example, take $I_{eo}=1$ μa and $r_{b'b}=100$ ohms, then $F_{\min}\simeq 3.4$ at $I_e\simeq 20$ μa. The minimum noise figure is now obtained for a much smaller emitter current than in the previous case, in agreement with Guggenbuehl and Strutt's results [23]. Moreover, the example shows that the task of designing transistors that have a reasonable noise figure at the cutoff frequency is not a hopeless one, though it is impossible to have $F_{\min}<3$ in this case (corresponding to about 5 db).

The design condition for transistors with a low noise figure at the α-cutoff frequency is thus that $I_{eo}r_{b'b}$ should be made small. The minimum obtainable noise figure at the cutoff frequency will not change very much if α_o depends upon I_e.

The condition which must be satisfied for $I_{eo}r_{b'b}$ is actually less stringent than in the previous case. For we now have to require only that $(2eI_{eo}r_{b'b}/kT)\ll 1$, whereas it had to be $<(1-\alpha_o)$ in the previous case. Viewed in that light, the requirement of a low noise figure at the cutoff frequency does not pose any additional restrictions upon the product $I_{eo}r_{b'b}$.

If we look at the circuit from a noise resistance point of view, (63) has to be modified somewhat. If we now put $R_{s1}\simeq 0$ and $R_{sc}\simeq 0$, which may not be too far from the truth, we have for the cutoff frequency

$$R_n\simeq R_s+r_{b'b}+g_{s1}(R_{eo}+R_s+r_{b'b})^2$$

$$=(R_s+r_{b'b})+\frac{1}{2R_{eo}}(R_{eo}+R_s+r_{b'b})^2 \qquad (70)$$

since

$$R_e\simeq R_{eo}=\frac{kT}{eI_e} \quad\text{and}\quad g_{s1}\simeq\frac{eI_e}{2kT}=\frac{1}{2R_{eo}}. \qquad (70a)$$

Considered as a function of R_{eo}, this has a minimum value

$$R_n\simeq 3(R_s+r_{b'b}) \text{ if } R_{eo}\simeq(R_s+r_{b'b}), \text{ or}$$

$$I_e=\frac{kT}{e(R_s+r_{b'b})}. \qquad (70b)$$

Putting $R_s=0$ and $r_{b'b}=100$ ohms, this corresponds to $I_e=0.25$ ma and $R_n\simeq 300$ ohms. At low frequencies at the same current, g_{s1} is quite small, hence

$$R_n\simeq(R_s+r_{b'b}+\tfrac{1}{2}R_{eo})=\tfrac{3}{2}(R_s+r_{b'b}) \qquad (70c)$$

so that $R_n\simeq 150$ ohms if $R_s=0$. This shows that at the frequency f_o there is an optimum noise resistance for small source impedance R_s and that the noise resistance only increases by a factor 2 by going to the cutoff frequency. *The low-noise properties of a transistor operating from a low impedance source thus hold quite well for frequencies up to the cutoff frequency.*

[14] This calculation is incorrect if α_o depends on I_e. This is, for example, the case in silicon transistors, where α_o decreases strongly with decreasing I_e for small emitter currents. This probably will not change the design conditions very much.

At very large source impedances R_s, the noise resistance $R_n \simeq g_s R_s^2$; close to the cutoff frequency R_n thus increases rapidly with increasing frequency unless I_e is of the order of I_{co}, the collector saturated current. Whether or not this noise performance can be tolerated in practical cases depends upon the required value of the equivalent input saturated diode current I_n [see (56)].

We finally turn to the grounded collector [6], [41] circuit (Fig. 6). In order to calculate its noise figure, we use open-circuit output and obtain after some calculation, if the correlation impedance Z_{ic} is neglected,

$$F = 1 + \frac{r_{b'b} + R_{s1}}{R_s} + \frac{|\alpha|^2 g_{s1}}{R_s}(R_s + r_{b'b})^2. \quad (71)$$

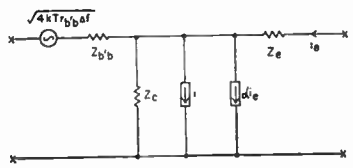


Fig. 6—Equivalent circuit of a transistor in grounded collector connection.

For low frequencies, $|\alpha|^2 \simeq 1$ and this is only slightly better than the other two circuits. For very high frequencies ($f > f_o$) the last term is smaller and the grounded collector circuit has a much lower noise figure than the other two circuits; unfortunately, this is not very useful, since the circuit gives a power gain less than unity at those frequencies.

To investigate how much improvement can be obtained at lower frequencies, we determine F_{min}.

$$F_{min} = 1 + 2|\alpha|^2 g_{s1} r_{b'b} + 2\sqrt{|\alpha|^2 g_{s1}(r_{b'b} + R_{s1}) + |\alpha|^4 g_{s1}^2 r_{b'b}^2}. \quad (72)$$

If $|\alpha|^2 g_{s1} r_{b'b} \ll 1$, this may be written, if $R_{s1} = \frac{1}{2}R_{eo}$:

$$F_{min} \simeq 1 + 2\sqrt{|\alpha|^2 g_{s1}(r_{b'b} + \tfrac{1}{2}R_{eo})}. \quad (72a)$$

If $|\alpha| \simeq 1$, this is practically identical with the value

$$F_{min} \simeq 1 + 2\sqrt{g_{s1}(r_{b'b} + \tfrac{1}{2}R_{eo})}$$

obtained for the other two circuits from (65) under corresponding conditions.

As far as the noise figure is concerned, the grounded collector circuit is thus practically identical with the other circuits. Since its power gain is much smaller, it should be checked carefully whether the noise of the following stage becomes important. In most cases the other two circuits will be preferred.

V. Modulation Noise Caused by Shot Effect

In view of Fonger's successful treatment of flicker noise, it seems logical to apply his approach also to the modulation noise caused by shot effect. This has not been done before; but, since the various noise sources cannot be expected to behave in different manners, modulation noise caused by shot effect should also exist.

Fonger assumed that *not the ac but the dc* series resistance of the junctions should show thermal noise. This means that the bulk material, quite apart from the fluctuations in the rate of injection, recombination, and escape of the carriers will also show noise because of the random motion of the carriers; this corresponds to thermal noise of the dc resistance. We thus follow Fonger's suggestion.

First turn to the junction diode. An ac current passing through the diode will change the rate of injection, recombination, and escape of the carriers; this is the cause of the modulation impedance Z_{mb}. Fluctuations in these rates will cause the shot noise described by the current generator i [Fig. 1(a)] connected in parallel to the junction; they will also modulate the dc series resistance R of the diode in two ways.[15]

1) Directly, as indicated by the noise emf e_b in series with R.
2) Indirectly, through the current generator i; this is taken into account (as in the flicker noise case) by connecting i across both the impedances Z and Z_{mb}.

The full equivalent circuit is thus as shown in Fig. 7(a). In the case of low-frequency noise Fonger found that the noise emf e_b had negligible influence; this does not necessarily have to be true for the shot noise case. Moreover, e_b and i should be partly correlated. The correlation does not necessarily have to be a complete one, since part of e_b may be caused by hole-electron pairs that never cross the junction; they only modulate the dc resistance, but do not contribute to i.

If e_b is neglected, the open-circuit noise emf is

$$e = i(Z + Z_{mb}) + \sqrt{4kTR_{b'b}\Delta f}. \quad (73)$$

At low frequencies, $Z_{mb} = R_{mb}$ and $R_{mb} = I(\partial R/\partial I)$ is negative, whereas $Z = R_o = 1/G_o$; the shot noise term thus should be zero if $(R_o + R_{mb}) = 0$. This effect has not yet been detected for shot noise.

If we now introduce the noise resistance R_n by equating $\overline{e^2} = 4kTR_n\Delta f$ and bear in mind that $\overline{i^2} = 2kT\Delta f/R_o$ for $I \gg I_n$ at relatively low frequencies, we obtain

$$R_n = R + \frac{1}{2}\frac{(R_0 + R_{mb})^2}{R_o} = \frac{1}{2}R_o + r + \frac{R_{mb}^2}{2R_o}. \quad (73a)$$

In most diodes the last term is negligible in comparison with the other two. The total noise power is then practically equal to the full shot noise power of the junction plus full thermal noise of the ac resistance; this agrees with Champlin's experimental data [7], [8]. Deviations should be expected for diodes with a large

[15] For the meaning of these two modes of modulation see footnote 10.

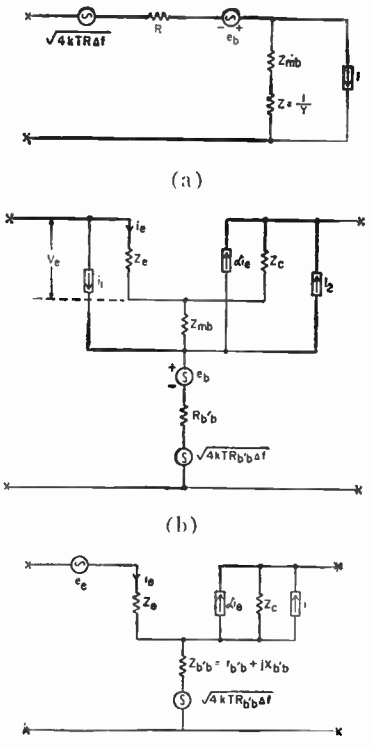


Fig. 7—Modulation noise caused by shot effect. (a) Equivalent circuit of a junction diode. (b) Equivalent circuit of a transistor. (c) Modification of the equivalent circuit of Fig. 4.

dc resistance R operating at large currents (small R_o). Attempts are being made at the University of Minnesota to detect such deviations.

The same ideas can be directly applied to transistors and lead to the equivalent circuit of Fig. 7(b). Here the current generator i is connected in parallel to Z_e and Z_{mb}, and the current generator i_2 is connected in parallel to Z_c and Z_{mb}. The noise emf e_b in series with the dc base resistance $R_{b'b}$ should be partly correlated with i_1 and i_2; it is probably not always warranted to ignore e_b (contrary to the flicker noise case).

As far as the noise figure is concerned,[16] we may replace the circuit of Fig. 7(b) by the one of Fig. 7(c) with

$$i = (i_2 - \alpha i_1); \quad e_e = i_1(Z_c + Z_{mb}) - e_b. \tag{74}$$

This circuit is nearly equivalent to the one shown in Fig. 4. The only differences are that the dc base resistance now shows thermal noise and that e_e has a different value. The current generator i apparently is not affected by the base modulation. The circuit of Fig. 7(c) should only be used if it may be assumed that $|Z_c| \gg |Z_{b'b}|$; the same condition was imposed upon Fig. 4.

Defining again the quantities g_{s1}, R_{s1}, and R_{sc} as in (43), we may write the noise figure F as in (48), but with values for A, B, and C different from the ones given in (48a)

$$A = 1 + 2g_{s1}(R_e + r_{b'b} + R_{sc});$$
$$B = R_{b'b} + R_{s1} + g_{s1}(R_e + r_{b'b} + R_{sc})^2; \quad C = g_{s1}; \tag{75}$$

where $r_{b'b}$ is again the ac base resistance. In these equations R_{s1} and R_{sc} are slightly different, whereas in the first term of B the ac resistance $r_{b'b}$ has been replaced by the dc resistance $R_{b'b}$.

The quantity g_{s1}, which gives the most important contribution to the noise figure F, has not changed, and the other terms in A, B, and C have not changed very much; it may thus be assumed that the theory of Section IV-B remains essentially correct even if base modulation noise is taken into account.

Now consider some of the changes expected at relatively low frequencies. First let e_b be assumed negligible, then

$$e_e'' = i_1(R_{eo} + R_{mb});$$

or

$$R_{s1} = \frac{\overline{i^2}(R_{eo} + R_{mb})^2}{4kTf} = \frac{1}{2} R_{eo}\left(\frac{R_{eo} + R_{mb}}{R_{eo}}\right)^2. \tag{76}$$

Substituting into the expression for B and putting $(R_{b'b} + R_{mb}) = r_{b'b}$, we have

$$B = R_{no} = \frac{1}{2} R_{eo} + r_{b'b} + \frac{1}{2}\frac{R_{mb}^2}{R_{eo}} + g_{s1}(R_{eo} + r_{b'b} + R_{sc})^2. \tag{77}$$

Without base modulation we would have had, since $R_{s1} = \frac{1}{2}R_{eo}$,

$$B = R_{no} = \tfrac{1}{2}R_{eo} + r_{b'b} + g_{s1}(R_{eo} + r_{b'b} + R_{sc})^2. \tag{77a}$$

Admittedly, the R_{sc} values may be slightly different in the two cases, but since the last terms in (77) and (77a) are small in comparison with the other ones, this will not cause a considerable difference. In that case the two B values differ mainly by the amount $\frac{1}{2}R_{mb}^2/R_{eo}$, which should be noticeable for transistors with a large dc base resistance $R_{b'b}$ at large emitter currents (small R_{eo}). If e_b is not negligible, the difference may be either larger or smaller, depending upon the correlation between e_b'' and i_1.[17] Coffey [6a] seems to have found indications that the measured noise resistance R_{no} for zero source impedance is somewhat larger than would be expected from (77a).

Now investigate the correlation resistance R_{sc} at relatively low frequencies. In that case

$$R_{sc} = \alpha_0 \frac{\overline{e_e i^*}}{\overline{i^2}} = \alpha_0(R_{eo} + R_{mb}) \frac{\overline{i_1 i^*}}{\overline{i^2}} + \alpha_0 \frac{\overline{e_b i^*}}{\overline{i^2}}$$
$$= \alpha_0 \frac{\overline{e_b i^*}}{\overline{i^2}} \tag{78}$$

since $\overline{i_1 i^*} = 0$ at relatively low frequencies, according to

[16] The circuits of Fig. 7(b) and 7(c) do not have equal response to the input signal, because of the different location of the current generator αi_e in the two cases. But since this change affects the signal and the noise in the same manner, the two circuits have identical noise figures. The effects of the base modulation are thus incorporated into the noise sources e_e and i.

[17] e_b'' is that part of e_b that is uncorrelated with i.

(17) and (19). At low frequencies one would thus expect a measurable correlation resistance only if e_b is not negligible and if a considerable correlation exists between i and e_b. Hanson and van der Ziel [28] found a measurable correlation resistance in transistors with a low value of α_o (and hence with a large base current). This is probably a base modulation effect; if it is, it would indicate that e_b is not always negligible.

The influence of shot noise base modulation requires further study. Work in progress at the University of Minnesota is aimed at obtaining a better understanding of the effect.

Bibliography

[1] Amakasu, K., and Asano, M. "Temperature Dependence of Flicker Noise in *N-P-N* Junction Transistors," *Journal of Applied Physics*, Vol. 27 (October, 1956), p. 1249.

[2] Anderson, R. L., and van der Ziel, A. "On the Shot Effect of *P-N* Junctions," IRE Transactions on Electron Devices, Vol. ED-1 (November, 1952), pp. 20–24.

[3] Bargellini, P. M., and Herscher, M. B. "Investigations of Noise in Audio Frequency Amplifiers Using Junction Transistors," Proceedings of the IRE, Vol. 43 (February, 1955), pp. 217–226.

[4] Becking, A. G. T., Groendijk, H., and Knol, K. S. "The Noise Factor of Four-Terminal Networks, *Philips Research Reports*, Vol. 10 (October, 1955), pp. 349–357.

[5] Chenette, E. R. "Measurement of Correlation between Flicker Noise Sources," this issue, p. 1304.

[6] Coffey, W. N. "Behavior of Noise Figure in Junction Transistors," Proceedings of the IRE, Vol. 46 (February, 1958), pp. 495–496.

[6a] ——. Private communication.

[7] Champlin, K. S. "A Study of Shot and Thermal Noise in Silicon *P-N* Junction Diodes." Unpublished Master of Science Thesis, University of Minnesota; September, 1955.

[8] ——. "Bridge Method of Measuring Noise in Low-Noise Devices at Radio Frequencies," Proceedings of the IRE, Vol. 46 (April, 1958), p. 779.

[9] Chynoweth, A. G., and McKay, K. G. "Photon Emission from Avalanche Breakdown in Silicon," *Physical Review*, Vol. 102 (April, 1956), pp. 369–376.

[10] ——. "Internal Field Emission in Silicon *P-N* Junction," *Physical Review*, Vol. 106 (May, 1957), pp. 418–426.

[11] Early, J. M. "Effects of Space-Charge Layer Widening in Junction Transistors," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1401–1406.

[12] Englund, J. W. "Noise Considerations for *P-N-P* Junction Transistors," in *Transistors I*. Princeton: RCA Laboratories, 1956, pp. 309–321.

[13] Fonger, W. H. "A Determination of L/F Noise Sources in Semiconductor Diodes and Transistors," in *Transistors I*. Princeton: RCA Laboratories, 1956, pp. 239–297.

[14] Freedman, L. A. "Design Considerations of the First Stage of Transistor Receivers," *RCA Review*, Vol. 18 (June, 1957), pp. 145–162.

[15] Giacoletto, L. J. "The Noise Factor of Junction Transistors," in *Transistors I*. Princeton: RCA Laboratories, 1956, pp. 296–308.

[16] Gianola, U. F. "Photovoltaic Noise in Silicon Broad Area *P-N* Junctions," *Journal of Applied Physics*, Vol. 27 (January, 1956), pp. 51–54.

[17] Guggenbuehl, W., and Strutt, M. J. O. "Messungen der spontanen Schwankungen bei Stromen mit Verschiedenen Trägern in Halbleitersperrschichten," *Helvetica Physica Acta*, Vol. 28, No. 7 (1955), pp. 694–704.

[18] ——, ——. "Experimentelle Bestätigung der Schottky'schen Rauschformeln an neueren Halbleiterflächendioden im Gebiet des weissen Rauschspektrums," *Archiv der elecktrischen Übertragung*, Vol. 9 (March, 1955), pp. 103–108.

[19] ——, ——. "Experimentelle Untersuchung und Trennung der Rauschursachen in Flächentransistoren," *Archiv der electrischen Übertragung*, Vol. 9 (June, 1955) pp. 259–269.

[20] ——, ——. "Theorie des Hochfrequenzrauschens von Transistoren bei kleinen Stromdicten," *Nachrichtentech, Fachberichte, Beihefte der N.T.Z.*, Vol. 5, 1956, pp. 30–33.

[21] ——, Schneider, B., and Strutt, M. J. O. "Messungen über das Hochfrequenzrauschen von Transistoren," *Nachrichtentech, Fachberichte, Beiheft der N.T.Z.*, Vol. 5, (1956), pp. 34–36.

[22] ——. "Theoretische Ueberlegungen zur physikalischen Begründung des Ersatzschaltbildes von Halbleiterflächendioden bei hohen Stromdichten," *Archiv der elecktrischen Übertragung*, Vol. 10 (November, 1956), pp. 433–435.

[23] ——, and Strutt, M. J. O. "Theory and Experiments of Shot Noise in Semiconductor Junction Diodes and Transistors," Proceedings of the IRE, Vol. 45 (June, 1957), pp. 839–857.

[24] ——, ——. "Transistors in high-frequency amplifiers," *Electronic and Radio Engineer*, Vol. 34 (July, 1957), pp. 258–267.

[25] ——, "Beiträge zur Kenntnis des Halbleiterrauschens mit besondere Berücksichtigung für Kristaldioden und Transistoren." Unpublished Ph.D. dissertation, Eidgenössischen Technischen Hochschule, Zurich, 1955.

[26] Hanson, G. H. "Shot Noise in *P-N-P* Transistors," *Journal of Applied Physics*, Vol. 26 (November, 1955), pp. 1338–1339.

[27] ——. "An Experimental Investigation of Noise in Transistors." Unpublished Ph.D. dissertation, University of Minnesota, 1957.

[28] ——, and van der Ziel, A. "Shot Noise in Transistors," Proceedings of the IRE, Vol. 45 (November, 1957), 1538–1542.

[29] Hyde, F. J. "Measurements of Noise Spectra of a Point Contact Germanium Rectifier," *Proceedings of the Physical Society B*, Vol. 66 (December, 1953), pp. 1017–1024.

[30] ——. "Measurement of Noise Spectra of a Germanium *P-N* Junction Diode," *Proceedings of the Physical Society B*, Vol. 69 (February, 1956), Pt. 2, pp. 231–241.

[31] Kennedy, D. "Gaseous Ambients and Diode Noise." Unpublished paper presented at the 1954 IRE-AIEE Conference on Semiconductor Device Research, University of Minnesota.

[32] Keonjian, E., and Schaffner, J. S. "An Experimental Investigation of Transistor Noise," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1456–1460.

[33] Kingston, R. H. "Review of Germanium Surface Phenomena," *Journal of Applied Physics*, Vol. 27 (February, 1956), pp. 101–114.

[34] ——, and McWhorter, A. L. "Relaxation Time of Surface States on Ge," *Physical Review*, Vol. 103 (August, 1956), pp. 534–540.

[35] ——, *et al. Surface Physics*. Philadelphia: University of Pennsylvania Press, 1957.

[35a] Lummis, F. L., and Petritz, R. L. "On Noise in *P-N* Junction Rectifiers: II Experiment," *Physical Review*, Vol. 91 (July, 1953) p. 231.

[36] McKay, K. G., and McAfee, K. B. "Electron Multiplication in Silicon and Germanium," *Physical Review*, Vol. 91 (September, 1953), pp. 1079–1084.

[37] ——. "Avalanche Breakdown in Silicon," *Physical Review*, Vol. 94 (May, 1954), pp. 877–884.

[38] McWhorter, A. L., and Kingston, R. L. "Channels and Excess Reverse Current in Grown Germanium *P-N* Junction Diodes," Proceedings of the IRE, Vol. 42 (September, 1954), pp. 1376–1380.

[39] Montgomery, H. C. "Transistor Noise in Circuit Applications," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1461–1471.

[40] ——, and Clark, M. A., "Shot Noise in Junction Transistors." *Journal of Applied Physics*, Vol. 24 (October, 1953), 1337–1338.

[41] Nielsen, E. C. "Behavior of Noise Figure in Junction Transistors," Proceedings of the IRE, Vol. 45 (July, 1957), pp. 957–963.

[42] North, D. O. "A Physical Theory of Noise in Transistors." Unpublished paper presented at the 1955 IRE-AIEE Conference on Semiconductor Device Research, University of Pennsylvania. Philadelphia, Pa.

[43] Pearson, G. L., Montgomery, H. C., and Feldmann, W. L. "Noise in Silicon *P-N* Junction Photocells," *Journal of Applied Physics*, Vol. 27 (January, 1956), pp. 91–92.

[44] Petritz, R. L. "On the Theory of Noise in *P-N* Junctions and Related Devices," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1440–1456.

[44a] ——. "On Noise in *P-N* Junction Rectifiers and Transistors: I Theory," *Physical Review*, Vol. 91 (July, 1953), pp. 204, 231. See especially the correction of p. 231 shown on p. 204.

[45] Rose, D. J. "Microplasmas in Silicon," *Physical Review*, Vol. 105 (January, 1957), pp. 413–418.

[46] Rothe, H., and Dahlke, W. "Theory of Noisy Four-Poles," Proceedings of the IRE, Vol. 44 (June, 1956), pp. 811–818.

[47] Ryder, R. M., and Kircher, R. J. "Some Circuit Aspects of the Transistor," *Bell System Technical Journal*, Vol. 28 (July, 1949), pp. 367–400.

[48] Shockley, W. *Electrons and Holes in Semiconductors*. New York: D. van Nostrand Co., Inc., 1950, p. 342.

[49] Slocum, A., and Shive, J. N. "Shot Dependence of *P-N* Junction Phototransistor Noise," *Journal of Applied Physics*, Vol. 25 (March, 1954). p. 406.

[49a] Solow, M. "Theory of Noise in a Multidimensional Semiconductor with a *P-N* Junction." Thesis. Catholic University of America, 1957. Also, Silver Spring: U. S. Naval Ordnance Laboratory, Navord 5762.

[50] Stephanson, W. L. "Measurements of Junction Transistor Noise in the Frequency Range 7–50 KC/S," *Proceedings of the IEE*, Vol. 102B (November, 1955), pp. 753–756.
[51] Torrey, H. C., and Whitmer, C. A. *Crystal Rectifiers*. New York: McGraw-Hill Book Co., Inc., 1948.
[52] Uhlir, A. "High Frequency Shot Noise in *P-N* Junctions," PROCEEDINGS OF THE IRE, Vol. 44 (April, 1956), pp. 557–558. Erratum, Vol. 44 (November, 1956), p. 1541.
[53] van der Ziel, A. "Note on Shot and Partition Noise in Junction Transistors," *Journal of Applied Physics*, Vol. 25 (June, 1954), pp. 815–816.
[54] ——. *Noise*. Englewood Cliffs: Prentice-Hall, Inc., 1954, ch. 8.
[55] ——. "Shot Noise in Junction Diodes and Transistors," PROCEEDINGS OF THE IRE, Vol. 43 (November, 1955), pp. 1639–1646; and Vol. 45 (July, 1957), p. 1011.
[56] ——, and Becking, A. G. T. "Theory of Junction Diode and Junction Transistor Noise," PROCEEDINGS OF THE IRE, Vol. 46 (March, 1958), pp. 589–594.
[57] ——. *Fluctuation Phenomena in Semiconductors*. London: Thornton Butterworth, Ltd., to be published.
[58] Volkers, W. K., and Pedersen, N. E. "The 'Hushed' Transistor Amplifier," *Tele-Tech*. Vol. 14 (December, 1955), pp. 82–84, 156–158; and Vol. 15 (January, 1956), p. 70.
[59] Wallace, R. L., and Pietenpol, W. J. "Some Circuit Properties and Applications of *N-P-N* Transistors," *Bell System Technical Journal*, Vol. 30 (July, 1951), pp. 530–563. Also, PROCEEDINGS OF THE IRE, Vol. 39 (July, 1951), pp. 753–757.
[60] Weisskopf, V. F. "On the Theory of Noise in Conductors, Semiconductors and Crystal Rectifiers." NDRC No. 14-133, May, 15, 1953, unpublished.
[61] Yajima, T. "Emitter Current Noise in Junction Transistors," *Journal of the Physical Society, Japan*, Vol. 11 (October, 1956), pp. 1126–1127.

The Effects of Neutron Irradiation on Germanium and Silicon*

G. C. MESSENGER†, ASSOCIATE MEMBER, IRE, AND J. P. SPRATT†

Summary—**The known effects of neutron irradiation upon majority and minority carrier properties of germanium and silicon are reviewed, and used as a basis to derive a theoretical expression for the dependence of grounded-emitter current gain of a transistor upon accumulated neutron dose. This theoretical expression assumes a Shockley-Read recombination mechanism in the base of the transistor; the crystal defects introduced by bombardment act as recombination sites. A number of germanium and silicon transistors were irradiated at different facilities; the observed changes in transistor parameters are explained in terms of the theory. This explanation enables determination for germanium of certain basic quantities in recombination theory, *viz.*, the position in the forbidden band of the recombination site ($E_C - E_t = 0.23$ ev), and the capture cross section of the site for hole and electron capture ($\sigma_p = 1.0 \times 10^{-15}$ cm², and $\sigma_n \simeq 4 \times 10^{-15}$ cm²).**

INTRODUCTION

THE electrical properties of semiconducting materials are extremely dependent upon disorder in lattice structure. Nuclear bombardment of these materials greatly increases this disorder through the creation of Frenkel defects, *i.e.*, regions in which atoms have been knocked from their sites in the lattice and placed in interstitial positions. Such defects produce energy levels in the forbidden band of a semiconductor and thus cause changes in the electrical properties. James and Lark-Horovitz[1] have proposed a model for the energy levels associated with Frenkel defects that provides a qualitative, and to some extent a quantitative, explanation of the observed changes in germanium and silicon as a result of bombardment. This model predicts that the interstitial atom will act as a donor, and the vacancy as an acceptor. It predicts further that for semiconductors with high dielectric constants, such as germanium and silicon, there will occur in the forbidden band two levels (*viz.*, the first and second ionization potentials) corresponding to the interstitial atom, and also two levels corresponding to the vacancy. When Frenkel defects are formed, the electrons donated by the interstitial atom are redistributed among the states of lower energy, so that the interstitial may be single or even doubly ionized. Correspondingly, the vacancy may have one or two electrons in it.

Cleland[2,3] and co-workers at Oak Ridge National Laboratories have attempted to explain radiation-induced changes in the majority-carrier properties of *n* and *p*-type germanium in terms of the James-Lark-Horovitz model. Agreement between the theory and these experiments was good at room temperatures; all four levels were located and identified. The results of these experiments show that the interstitial atom introduces one level approximately 0.2 ev below the conduction band and one level 0.18 ev above the valence band,

* Original manuscript received by the IRE, March 14, 1958. The work described was carried out under a subcontract with Boeing Airplane Co., sponsored by the Air Force under Prime Contract No. AF-33 (600)-35050.
† Res. Div., Philco Corp., Philadelphia, Pa.
[1] H. M. James and K. Lark-Horovitz, "Localized electronic states in bombarded semiconductors," *Z. Physik Chem.*, vol. 198, nos. 1–4, pp. 107–126; 1956.
[2] J. W. Cleland, J. H. Crawford, and J. C. Pigg, "Fast-neutron bombardment of *n*-type Ge," *Phys. Rev.*, vol. 98, pp. 1742–1750; June 1955.
[3] J. W. Cleland, J. H. Crawford, and J. C. Pigg, "Fast neutron bombardment of *p*-type germanium," *Phys. Rev.*, vol. 99, pp. 1170 1181; August, 1955.

while the two levels due to the vacancy lie 0.066 ev and 0 ev above the valence band (see Fig. 1). The conductivity behavior of n and p-type germanium under neutron bombardment can be successfully explained in terms of this distribution of energy levels.

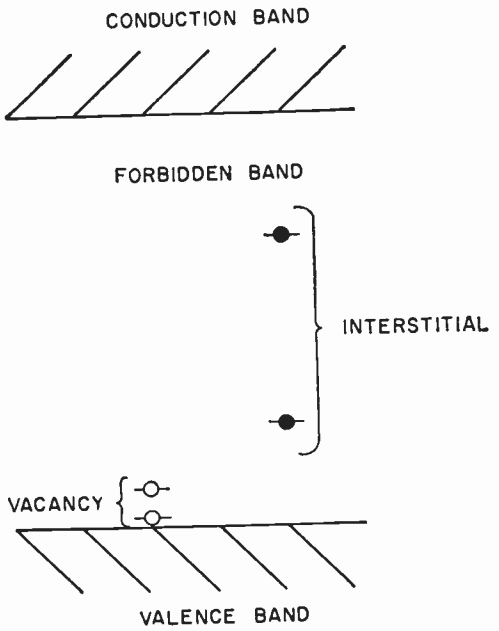


Fig. 1—Level diagram for Frenkel defect in germanium.

In addition to changing majority-carrier properties, nuclear bombardment would also be expected to affect minority-carrier properties of semiconductors, since the lattice disorder may act as a recombination center. Many investigators[4,5] have shown that bombardment does affect minority-carrier lifetime drastically. The effect of nuclear irradiation upon minority-carrier lifetime in bars of n and p-type germanium has been determined quantitatively by Curtis *et al.*[6,7] Radiation introduces recombination sites at a linear rate, so that the induced lifetime is inversely proportional to radiation dose. A dependence of this induced lifetime upon initial carrier concentration was observed by Curtis and was explained in terms of the Hall[8] and Shockley-Read[9] model for minority-carrier recombination. A unique determination of the position of the recombination site in the forbidden band was not possible solely on the basis of Curtis' work, but by assuming that the assignment of levels given by Cleland was correct and that therefore the important recombination site would be the level corresponding to the first ionization potential of the interstitial, Curtis was able to show that E_t, the position of the level, is 0.23 ev below the conduction band.

Knowledge of the bulk properties of irradiated semiconductors permits the changes in parameters of transistors subjected to bombardment to be deduced, if the relationships between transistor parameters and bulk properties of the constituent semiconductors are known. Loferski,[10] using Webster's[11] equation for the grounded-emitter current gain of a transistor, has shown that $\Delta 1/\beta$ should be directly proportional to neutron dose, due to lifetime degradation in the base region. This decrease in β is also inversely proportional to the square of the electrical base width: thus, for a given radiation flux level, the thinner the base, the smaller the change in β. Loferski also reported that observed changes in surface-recombination velocity of transistors upon neutron bombardment were transient, and did not contribute appreciably to permanent damage.

The work of these investigators has provided a basis for the present detailed experimental analysis of the effects of fast neutrons on transistors; this analysis is presented in the following sections.

Description of Experiment

Radiation Sources

The radiation sources employed in this study were a cyclotron, a nuclear reactor, and a critical assembly. The 60-inch cyclotron at the University of Washington in Seattle provided a high energy spectrum (peaking at approximately 10 mev) by focusing 20-mev deuterons on a beryllium target, which stripped the deuteron of a proton and allowed the neutron to pass through. The Materials Testing Reactor at Arco, Idaho, provided, in addition to fast neutrons, a high thermal and epithermal flux. The third radiation source, the "Godiva," a critical assembly at Los Alamos, N. M., provided a fission spectrum.

The use of several neutron sources for these experiments posed the problem of correlating dosimetry data; expected effects would depend upon the neutron spectrum employed, due to the following relationship.

$$\text{The number of Frenkel defects} = \int \frac{\text{the number of defects}}{\text{neutron}}(E) \times \text{neutrons}\,(E)dE.$$

The number of defects introduced per neutron as a function of neutron energy increases from the threshold for displacement (about 1 kev) until saturation (about

[4] C. D. Florida, F. R. Holt, and J. H. Stephen, "Irradiation of transistors," *Nature*, vol. 173, pp. 397–398; February, 1954.
[5] W. L. Brown, R. C. Fletcher, and K. A. Wright, "Traps produced by electron bombardment of germanium at low temperature," *Phys. Rev.*, vol. 96, p. 834; November, 1954.
[6] O. L. Curtis, *et al.*, "Effect of irradiation on the hole lifetime of n-type germanium," *J. Appl. Phys.*, vol. 28, pp. 1161–1165; October, 1957.
[7] O. L. Curtis, Jr., and J. W. Cleland, Annual Meeting of the American Physical Society, New York, N. Y.; January 29–February 1, 1958.
[8] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, p. 387; July, 1952.
[9] W. Shockley and W. T. Read, "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835–842; September, 1952.
[10] J. T. Loferski, "Analysis of the effect of nuclear radiation on transistors," *J. Appl. Phys.*, vol. 29, pp. 35–40; January, 1958.
[11] W. M. Webster, "On the variation of junction-transistor current-amplification factor with emitter current," PROC. IRE, vol. 42, pp. 914–920; June, 1954.

2-4 mev); $N(E)$, the neutron spectrum, depends upon the neutron source. Dosimetry measurements for these experiments were performed and correlated by members of the Applied Physics Staff, Pilotless Aircraft Division, Boeing Airplane Company, under the direction of Dr. D. A. Hicks. Both neutron threshold detector foils and transistor voltage punch-through dosimeters were used. The results presented here are in terms of neutrons above 4 kev in a "Godiva" spectrum. This spectrum was chosen to fit the requirements of systems designers; it can be related to other spectra by a multiplicative constant. Neutron dose was measured in terms of neutrons per square centimeter (nvt).

Measurements of Effect of Radiation on Transistors

Transistors provide a definite advantage over semiconductor bulk material for studying radiation effects since they permit measurement of base lifetime in the range from several hundred microseconds down to the millimicrosecond range, whereas determination of lifetime by photoconductivity in bars of semiconductor material is in general limited to the range greater than 1 microsecond.

One hundred and twenty transistors of 25 types were employed in these experiments. The group included both germanium and silicon transistors. The aim was to provide a range of bulk and device parameters as wide and as representative as possible. Prior to irradiation, electrical and mechanical quantities essential to the expected analysis were measured; the quantities included:

r_b' = extrinsic base resistance
C_c = collector capacitance
f_{max} = maximum frequency of oscillation
V_{PT} = punch-through voltage
I_{c0} = collector saturation current
V_{DE} = emitter-diode-breakdown voltage
V_{DC} = collector-diode-breakdown voltage
A_c = area of the collector
A_e = area of the emitter.

These quantities were remeasured after irradiation, when possible. The dynamic measurements consisted of measuring grounded-emitter current gain for the 120 transistors as a function of fast neutron dose. Collector-diode characteristics with the base open were taken for 29 germanium transistors.

After the dynamic measurements were completed, the transistors were removed from their headers and were etched and remeasured to detect surface effects, if any. Germanium p-n-p graded-base transistors were also probed to determine whether the high-resistivity region around the collector junction had converted from n to p-type.

Theory of the Effects of Radiation upon Transistors

Lifetime Damage

In attempting to obtain an expression for the grounded-emitter current gain of transistors that have been permanently damaged by fast neutron irradiation, we shall revise the equation for grounded-emitter current gain by adapting Shockley-Read statistics to the bulk-recombination process. Further, we shall assume that at low neutron doses the only semiconductor parameter affected by the radiation is bulk lifetime in the base region. Experimental results indicate that the effect of radiation upon surface properties is primarily transient; therefore, in considering permanent damage, surface-recombination velocity can be neglected.

The following expressions from Shockley-Read statistics are necessary in this analysis.

$$U = \frac{dp}{dt} = \frac{p}{\tau_0}\frac{(1 + cp)}{(1 + ap)} \tag{1}$$

$$\tau_0 = \tau_{n0}\frac{(p_0 + p_1)}{(n_0 + p_0)} + \tau_{p0}\frac{(n_0 + n_1)}{(n_0 + p_0)} \tag{2}$$

$$\tau_\infty = \tau_{n0} + \tau_{p0} = \tau_0\frac{a}{c} \tag{3}$$

$$\frac{a}{c} = \frac{(1 + R)(n_0 + p_0)}{n_0 + n_1 + R(p_0 + p_1)} \tag{4}$$

$$R = \frac{\tau_{n0}}{\tau_{p0}} = \frac{C_p N_t}{C_n N_t} = \frac{v_{Tp}\sigma_{cp}}{v_{Tn}\sigma_{cn}} = 1.3\frac{\sigma_{cp}}{\sigma_{cn}}. \tag{5}$$

Here:

U = recombination rate
p = added hole concentration
n_0, p_0 = initial electron and hole concentration
n_1, p_1 = electron and hole concentration when Fermi level falls at trap level
τ_0 = low-level lifetime
τ_{n0}, τ_{p0} = lifetime of electrons and holes injected into highly p and n-type material, respectively
τ_∞ = high-level lifetime
a, c = constants of recombination process
R = ratio of capture probability
C_n, C_p = capture probability of electrons and holes
v_{Tn}, v_{Tp} = average thermal velocity of electrons and holes
σ_{cn}, σ_{cp} = capture cross section of recombination site for electrons and holes
N_t = density of recombination sites.

Webster's[11–13] expression for grounded-emitter current gain is

$$1/\beta = \frac{SWA_s}{D_p A_e}g(Z) + \left[\frac{\sigma_b W}{\sigma_e L_e} + \frac{1}{2}\left(\frac{W}{L_b}\right)^2\right](1 + Z). \tag{6}$$

Here the first term describes surface recombination. The second term describes injection efficiency and vol-

[12] A. Rittmann, G. C. Messenger, R. A. Williams, and E. Zimmerman, "Microalloy transistor," IRE Trans. on Electron Devices, vol. ED-5, pp. 49–54; April, 1958.
[13] N. H. Fletcher, "Note on the variation of junction transistor current amplification factor with emitter current," Proc. IRE, vol. 44, pp. 1475–1476; October, 1956.

ume recombination. The factor $(1+Z)$ multiplying W^2/L_b^2 treats bulk recombination as a bimolecular process. Applying the Shockley-Read statistics to describe the variation of lifetime with injection level and combining the effects of injection efficiency and surface recombination into a single term, $1/\beta_0$, we get

$$1/\beta = 1/\beta_0 + \frac{W^2}{2D_p\tau_0} h(Z) \tag{7}$$

$$h(Z) = H(Z)g(Z)$$

$$H(Z) = \left[\frac{c}{2a} - \left(\frac{c}{a} - 1\right)\left\{\frac{\ln\left(1+\frac{aZ}{c}\right)}{\left(\frac{aZ}{c}\right)^2} - \frac{1}{\frac{aZ}{c}\left(1+\frac{aZ}{c}\right)}\right\}\right],$$

$$g(Z) = (1+Z)/(1+2Z)$$

$$Z = \frac{WI_e}{qD_pA_en_0}. \tag{8}$$

Here W is base width, D_p is hole-diffusion constant, I_e is emitter current, and A_e is emitter area. The function $h(Z)$, which combines the current dependence of D_p and τ_0, is such that for injection levels of interest,

$$h(Z) = \frac{1}{2} + \frac{Z}{6}\left(1 - \frac{2a}{c}\right). \tag{9}$$

For $(aZ/c)\gg 1$, the high-level case, the expression is

$$h(Z) = \frac{c}{4a}. \tag{10}$$

The variation of $h(Z)$ is shown in Fig. 2.

Utilizing the expression for the current-gain frequency cutoff,

$$f_{c\alpha} = \frac{1.22D_p}{\pi W^2}. \tag{11}$$

Eq. (5) may be written

$$1/\beta = 1/\beta_0 + \frac{0.2}{f_{c\alpha}\tau_0} h(Z). \tag{12}$$

Assuming that lifetime degrades with flux according to

$$\frac{1}{\tau_0} = \frac{1}{\tau_i} + \frac{\phi}{K} \tag{13}$$

where τ_i is the value of lifetime before any irradiation, then

$$1/\beta = 1/\beta_0 + \frac{0.2}{f_{c\alpha}}\left(\frac{1}{\tau_i} + \frac{\phi}{K}\right) h(Z). \tag{14}$$

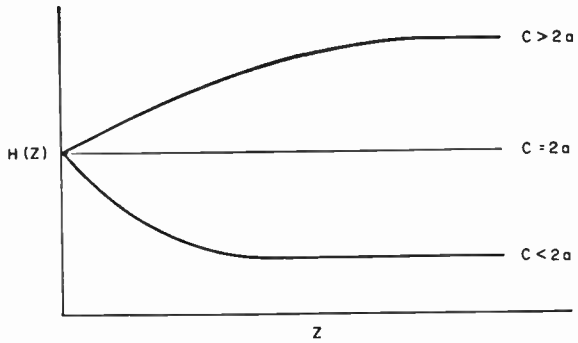


Fig. 2—Variation of $h(Z)$ with Z for various values of the ratio, c/a.

This is the basic equation which explains the dependence of current gain on both flux and injection level. Two useful relations for obtaining the ratio a/c can be deduced from this equation if we assume a linear approximation to $h(Z)$

$$\frac{\partial 1/\beta}{\partial I_e} = \left.\frac{\partial 1/\beta}{\partial I_e}\right|_{\phi=0} + \frac{0.2}{f_{c\alpha}}\left(\frac{1}{\tau_i} + \frac{\phi}{K}\right)\frac{1}{6}\frac{\partial Z}{\partial I_e}\left(1 - \frac{2a}{c}\right) \tag{15}$$

$$\frac{\partial^2 1/\beta}{\partial\phi\partial I_e} \div \frac{\partial 1/\beta_M}{\partial\phi} = \frac{1}{3}\frac{\partial Z}{\partial I_e}\left(1 - \frac{2a}{c}\right). \tag{16}$$

Here $1/\beta_M$ is the value of $1/\beta$ at zero emitter current, if the dependence of surface effects upon emitter current is neglected.

By combining (3), (5), and (13), a relation between damage constant, K, and the statistics of the recombination process is obtained.

$$K = \frac{n_0 + n_1 + Rp_0 + Rp_1}{(n_0 + p_0)v_{Tp}\sigma_{cp}\frac{\partial N_t}{\partial\phi}}. \tag{17}$$

Resistivity Damage

At relatively high flux levels, the change in effective donor density of the semiconductor material is given by

$$N_D = N_{D0} + K_1\phi \tag{18}$$

where K_1 is the rate of change of donor concentration with neutron dose.

The punch-through voltage of a transistor is defined by

$$V_{PT} = \frac{W^2qN_D}{2\epsilon} = CN_D \tag{19}$$

where ϵ is the dielectric constant and N_D is the donor density. Combining (18) and (19) provides (20) for the punch-through voltage as a function of flux, which can be used to experimentally determine K_1.

$$V_{PT} = V_{PT_0}\left(1 + \frac{K_1\phi}{N_{D0}}\right). \tag{20}$$

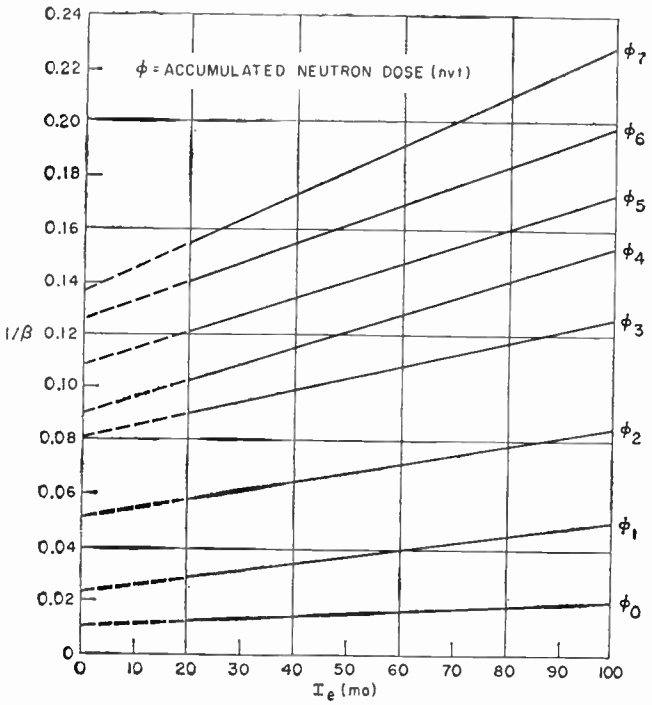


Fig. 3—Plot of $1/\beta$ vs I_e at different neutron doses for a typical germanium transistor.

General Results

Lifetime-Damage Constant

It was found experimentally for all the transistors tested that $1/\beta_M$ increases linearly with accumulated neutron dose according to (14). Fig. 3 shows curves for a typical germanium transistor which indirectly verify (13). It was also found that the quantity K, the lifetime-damage constant, as expected, was a function of base type, and semiconductor type, as indicated by (17). No conclusions could be reached about the dependence of K upon base resistivity, however, since the spread in base resistivity in available transistors is slight, and any such dependence would be masked by experimental error. Since a large percentage of commercially available transistors have a resistivity close to 2 ohm-cm, it is a useful oversimplification to determine average values of K for p and n-base silicon and germanium; these values

TABLE I

Constants of the Recombination Process in Irradiated Germanium and Silicon

	K (nvt-μsec)	$E_C - E_t$	σ_p(cm²)	σ_n(cm²)	$\partial N_t/\partial\phi$
nGe	$5.0\pm2.0\times10^{13}$	0.23 ev	1.0×10^{-15}	$\sim4\times10^{-15}$	3
pGe	$2.4\pm0.4\times10^{13}$	0.23 ev	1.0×10^{-15}	$\sim4\times10^{-15}$	3
nSi	$2.8\pm0.8\times10^{12}$	—	—	—	—
pSi	$3.2\pm1.1\times10^{12}$	—	—	—	—

are listed in Table I. In addition, because of the different relative positions of trap levels and of the Fermi level in silicon, its K is not expected to vary with resistivity.

The first and most obvious comment is to note that K for germanium is an order of magnitude higher than K for silicon. K is higher in n-type germanium than in p-type germanium of comparable resistivity, which is consistent with theoretical expectations on the basis of (15), provided R is close to 1. It is also noted that this higher value of K in n-type germanium does not agree with Loferski's assumption.

Damage constants for germanium p-n-p graded-base units averaged 0.5×10^{13} nvt-μsec. Preliminary calculations indicate that the expected increase in K (because of the built-in field) is more than offset by the expected decrease in K (because the average resistivity is much lower than the 2 ohm-cm characteristic of homogeneous-base units). Since K in silicon is probably not a function of resistivity, some improvement in K is expected from a built-in field. Preliminary experimental results indicate that this is probably true.

Characteristics of the Recombination Site in Germanium

It has been shown that, of the four levels introduced into the forbidden band of germanium by bombardment, the recombination processes should be controlled by the level corresponding to the first ionization potential of an interstitial atom.[6] This level is known to lie above midband so that

$$\frac{a}{c} = (1 + R)(n_0 + p_0)/n_0 + n_1 + Rp_0. \qquad (21)$$

Using (21) for a/c, we see that it is possible to determine n_1 and therefore E_t, the position of the level, by two methods. The first consists of determining a/c for both n and p-base germanium transistors for neutron doses so small that conductivity changes can be neglected. This will furnish two sets of equations in two unknowns, n_1 and R, and therefore will allow us to evaluate both these quantities. Such an analysis was performed using germanium transistors, and the results show that

$$E_t = 0.23 \pm 0.02 \text{ ev below the conduction band.} \qquad (22)$$

Although the determination of R from this equation is considerably less accurate, its most probable value is 0.3. The values given by Cleland and Curtis for E_t are 0.20 and 0.23 ev, respectively, while Curtis estimates R to be in the neighborhood of 1.0.

The second method of determining E_t consisted of irradiating transistors to very high fluxes, so that conductivity begins to change. The contribution to $(\partial 1/\beta)/\partial I_e$ due to conductivity modulation by emitter current is independent of base resistivity, so that (15) is valid even at flux levels high enough to change base resistivity. Applying (15) and (21) to p-base germanium transistors, it should be possible to make the quantity $(\partial I/\beta)/\partial I_e$ go negative for high-neutron doses since p_0 increases with flux. Such an effect was actually produced on one n-p-n transistor which we were able to irradiate to the high level necessary. The resultant family of curves shows the slope going through zero

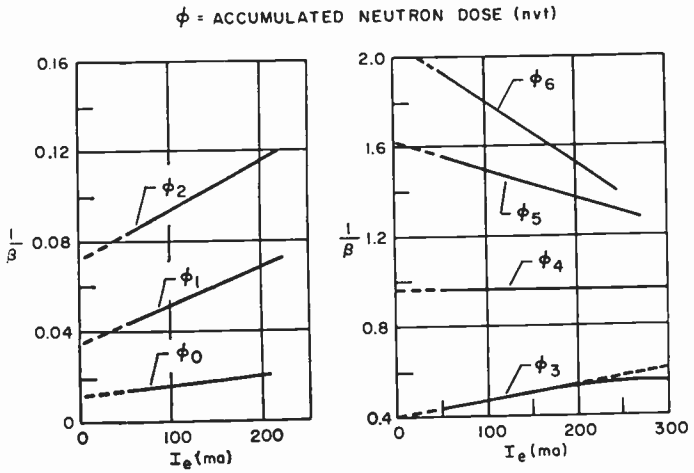


Fig. 4—Plot of $1/\beta$ vs I_e at different flux levels for a germanium n-p-n transistor that exhibited slope becoming negative.

(Fig. 4). Since $h(Z)$ levels off relatively early for $1/2 \leq a/c \leq 1$, it cannot be determined which curve corresponds to which value of a/c. However, if we assume that the observed case of zero slope corresponds to $a/c = 1$, we need make no assumption about the magnitude of R. Therefore, at this value of radiation dose,

$$\frac{\partial 1/\beta}{\partial I_e} = \frac{\partial 1/\beta}{\partial I_e}\bigg|_{\phi=0}, \qquad \frac{a}{c} = 1; \tag{23}$$

therefore,

$$n_1 = p_0 = p_0|_{\phi=0} + \frac{\partial p_0}{\partial \phi}\,\phi. \tag{24}$$

The value of $\partial p_0/\partial \phi$ is taken as 1.5/cc/nvt from Cleland's[14] data,[3] so that E_t is 0.2 ev below the conduction band.[15] Recent Russian experiments[16] have also shown that it is possible to achieve a value of 1 for a/c in a p-n junction by varying the resistivity of the sample.

The lifetime-damage constant is related to the bulk properties of the germanium by (17). $\partial N_t/\partial \phi$ can be determined from measurements of the initial rate at which the majority-carrier concentration changes in moderate-to-low resistivity n-type germanium. For such material, two electrons are removed for each Frenkel defect, so that by obtaining $\partial N_D/\partial \phi$ we can obtain $\partial N_t/\partial \phi$. G. L. Keister[17] of Boeing Airplane Company has performed experiments on transistor voltage punch-through dosimeters using the same spectra employed in these experiments, and his results, based on (20), show that

$$\frac{\partial N_t}{\partial \phi} = 3/\text{cc/nvt}. \tag{25}$$

Using this figure and the value obtained above for n_1, we can determine σ_p (which is the capture cross section of the recombination site for holes) from experimentally determined values for K using (18). σ_p was computed, and the value obtained was

$$\sigma_p = 1.0 \times 10^{-15}\ \text{cm}^2. \tag{26}$$

σ_n could not be determined accurately due to the occurrence of R in the equation. However, since

$$R = 1.3\,\frac{\sigma_p}{\sigma_n} \tag{27}$$

and since R was in the neighborhood of 0.3, σ_n is probably about 4×10^{-15} cm^2.

Surface-Recombination Velocity (s)

As has been stated, the primary effect of irradiation upon the surface of a transistor is a transient one. However, the permanent effects can also be studied, and two different methods were used to do this. The first consists of plotting $1/\beta_M$ vs nvt for some thin-base transistors. For such units, a large flux was required to produce β degradation due to lifetime damage, so that variations in $1/\beta_M$ at lower fluxes could be attributed to surface-recombination velocity. Several high-frequency p-n-p germanium transistors exhibited variations in $1/\beta_M$ which show that s varies with flux as shown in Fig. 5. Other experimenters[18] have reported an increase in β at low-flux levels, followed by degradation due to lifetime damage. This result indicates that it may be possible to obtain very low values of s by using radiation techniques.

The other technique used to determine effects of irradiation on s consisted of removing the transistors from their cases after they had been bombarded, etching them, and remeasuring their grounded emitter characteristics. If the latter had improved appreciably, the damage could be attributed to surfaces. No such effect was noticed. Silicon transistors showed no change in characteristics at all. Changes brought about in germanium were approximately equal to those produced in unexposed units used as controls to check the etch.

These results indicate that, aside from an interesting effect at low values of flux, the effect of irradiation upon

[14] Cleland's value is actually 0.75/cc/nvt. Comparison of our experimental values for $\partial N_D/\partial \phi$ on n-base germanium with Cleland's experimental values for $\partial N_D/\partial \phi$ on n-base germanium disclosed a factor of two discrepancy. This was attributed to a difference in dosimetry, so that the value of 0.75/cc/nvt was multiplied by two to correct for this difference.

[15] The value for m_e used in this calculation is that $m_e = 0.51$ m, given by B. Lax, H. J. Zeiger, R. N. Dexter, and E. S. Rosenblum, "Directional properties of the cyclotron resonance in germanium," *Phys. Rev.*, vol. 93, pp. 1418–1420; March, 1954.

[16] M. I. Iglitsyn, Y. A. Kontsevoy, and A. I. Sidorov, "The lifetime of non-equilibrium charge-carriers in germanium at arbitrary injection levels," *J. Tech. Phys. (U.S.S.R)*, vol. 27, pp. 2461–2468; November, 1957.

[17] Unpublished data.

[18] B. Reich and G. E. Pavlik, "A survey of the nuclear radiation effects on semiconductor materials," *AGET News Bull.*, vol. 1, pp. 8–17; July, 1957.

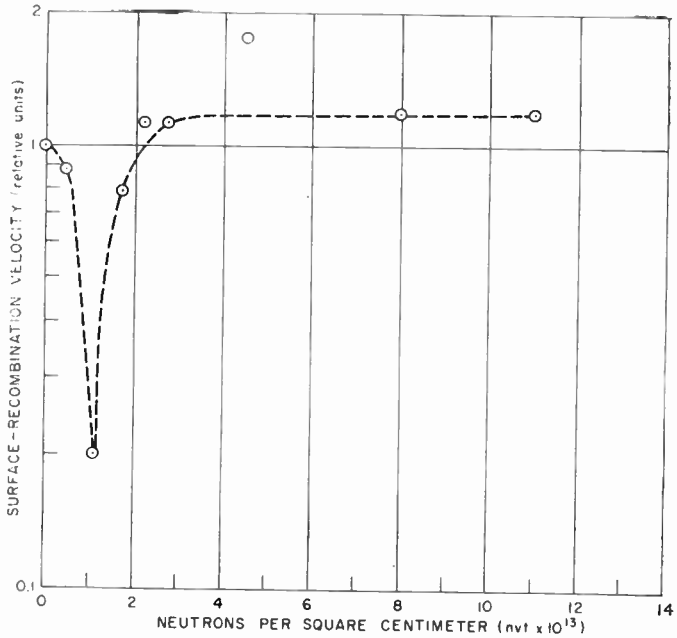


Fig. 5—Relative surface-recombination velocity as a function of fast neutron dose for a germanium p-n-p transistor.

surface-recombination velocity of transistors is small compared to bulk lifetime effects.

Resistivity Conversion in P-N-P Graded-Base Germanium Units

Since the collector-saturation current of p-n-p germanium graded-base transistors increased so markedly, and since some transistors actually appeared to have shorted from collector to emitter, it was decided to determine whether or not conversion of n-type germanium to p-type had occurred near the collector junction. Several of the units were removed from their cases, and a thermoelectric probe was placed on the collector side of the wafer. The base lead of the transistor was used as the reference point.

On two 2N247 transistors, there were definite indications that conversion had occurred; one L-5405 transistor exhibited slight p-type behavior on the base wafer near the collector dot, but the thermoelectric current was rather small, so that no final conclusion could be made in this case. These indications were reproducible, whether a hot or a cold probe was used. Quantitative measurements of the removal rate were difficult to obtain from these units since the initial carrier concentration was not known. However, if we assume that $n_0 \leq 7\times10^{13}$ corresponding to $\rho \geq 20$ ohm-cm under the collector, and if the average value of $\partial N_D/\partial N_t$ is -0.5, then the neutron dose to which these units were subjected would have been sufficient to cause conversion.

Collector-Saturation Current

The collector-saturation current, I_{c0}, was not obtained directly as a function of neutron dose, but its behavior can be inferred from the observed dependence on neutron dose of collector-saturation current measured with the base lead open. This latter quantity is proportional to the product βI_{c0}, and it was roughly constant over a range of neutron dosage for homogeneous-base germanium units. Therefore, it can be concluded that I_{c0} increases as β decreases with flux, so that the product is constant. This is to be expected from the dependence of I_{c0} on semiconductor bulk parameters.

Discussion

By combining the Shockley-Read statistics and transistor-design equations, constants have been derived which describe the damage mechanism for germanium and which agree very well with results obtained by other workers using other methods. Constants for silicon are being derived by a similar method. However, if, as expected, the recombination process is more complicated in silicon, additional information, such as data on the dependence of radiation-induced lifetime upon temperature, may be needed.

Conclusions

By modifying Webster's equation for grounded-emitter current gain through the application of Shockley-Read statistics to the lifetime term, we have derived the dependence of current gain in transistors on injection level and radiation level. Transistor measurements interpreted in terms of this theory gave values of 0.23 ev for E_t, 1.0×10^{-15} cm^2 for σ_p, and 4×10^{-15} cm^2 for σ_n. The dependence of current gain on the injection and radiation levels describes quantitatively the variation in transistor parameters with flux level. The surface damage was negligible compared to bulk damage except for an interesting decrease in surface-recombination velocity at relatively low flux levels.

Preliminary experiments on germanium indicate that the lifetime-damage constant is smaller for graded-base units than for homogeneous-base units. The high-resistivity material beneath the collector is quite likely to convert at flux levels lower than those which cause β degradation.

Acknowledgment

The authors are indebted particularly to Dr. D. A. Hicks, Dr. D. V. Keller, J. B. Robison, J. R. Orr, A. K. Durkee, and Mrs. B. M. Clarke of the Applied Physics Staff, Boeing Airplane Company, for dosimetry measurements used in this experiment, and also for the electrical measurements performed at the various facilities. The results of these experiments, which were performed jointly by the authors and the above-mentioned members of the Applied Physics Staff, were supplied for comparison with the theoretical predictions.

Irradiation of *P-N* Junctions with Gamma Rays: A Method for Measuring Diffusion Lengths*

R. GREMMELMAIER†

Summary—The photovoltaic effect in *p-n* junctions can be used to measure the diffusion lengths of minority carriers in a semiconductor. The short-circuit current in an irradiated *p-n* junction is $I_k = egL$, where e is the electron charge, g the generation rate (number of electron-hole pairs per unit volume and unit time generated by the radiation), and L a quantity which equals the diffusion length if the position of the *p-n* junction is suitably chosen. If g is known, the diffusion length can be calculated from the short-circuit current. The generation rate can easily be calculated if the junction is irradiated by γ rays from a Co^{60} source. The method will be described more closely in this paper. Diffusion lengths were measured in Si, GaAs, and InP containing a *p-n* junction. In GaAs diffusion lengths up to 8 μ were measured, and up to 130 μ in InP.

I. Introduction

THERE are a number of methods which can be employed to measure the lifetime or diffusion length of the minority carriers in a semiconductor. Of special interest for the production of rectifiers and transistors are those methods which allow one to measure the lifetime in specimens containing, for example, an alloyed or diffused *p-n* junction, because the lifetime depends not only on the nature of the raw material, but also on any further treatment of the material. One method suitable for this purpose is that of directly observing the decay with time of the density of injected minority carriers.[1] However, this method can only be employed with difficulty if the lifetime is very short and if, in addition, numerous traps are present as is the case in GaAs, for example. In this case another method, which is based on the photovoltaic effect in *p-n* junctions, can be used.

In the case of the photovoltaic effect, the short-circuit current through the junction is carried by the current carriers which can reach the *p-n* junction by means of diffusion. These are on the average the carriers generated within a diffusion length on each side of the *p-n* junction.[2] Therefore, if the generation rate g is known, the diffusion lengths can be calculated from the short-circuit current. The calculation can be easily made if g is constant. This is the case with radiation which is weakly absorbed in the semiconductor, *e.g.*, with very fast electrons or γ rays. The monoenergetic γ rays of a radioactive isotope are very suitable for the investigation; especially favorable is the radiation of Co^{60}. The γ-ray source can be easily handled; activities in the order of 1 curie or less are sufficient for the measurements. The intensities can be calculated from the distance between source and junction. The absorption coefficients for γ rays of this energy range are well known, so that the generation rate can be calculated. This method will be described in more detail in the following.[3]

II. Relation between Diffusion Length and Short-Circuited Current

The photovoltaic effect in *p-n* junctions has already been dealt with by various authors.[2,4] The following discussion is based on the model schematically illustrated in Fig. 1. There a *p-n* junction is shown in a crystal, the

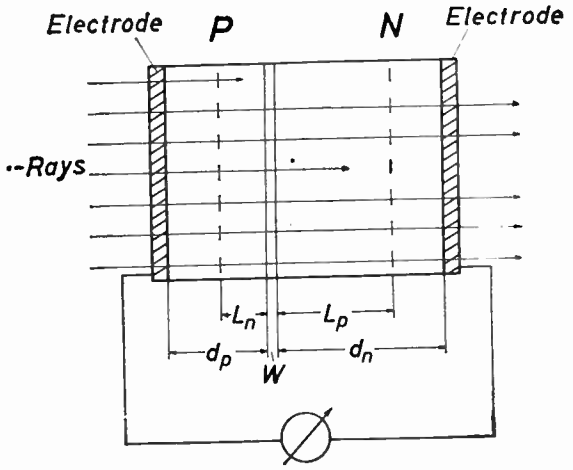


Fig. 1—Model of a *p-n* junction.

thickness of which should be small compared with the mean free path of the γ rays. Let the width of the space-charge region be W, and the distance of the junction from the electrodes (being assumed to make an ohmic contact) d_p and d_n. L_n and L_p are the diffusion lengths of the electrons in the p region and of the holes in the n region, respectively. The generation rate should be constant in the whole range. (The calculation of g will be carried through in the next section).

If the diffusion lengths are large compared with the width W of the space-charge region, the short-circuit current per unit area is given by[4]

* Original manuscript received by the IRE, March 3, 1958.
† Research Lab., Siemens-Schuckertwerke, Erlangen, Germany.
[1] S. R. Lederhandler and L. G. Giacoletto, "Measurement of minority carrier lifetime and surface effects in junction devices," PROC. IRE, vol. 43, pp. 477–483; April, 1955.
[2] R. L. Cummerow, "Photovoltaic effect in *p-n* junctions," *Phys. Rev.*, vol. 95, pp. 16–21; July 1, 1954.
[3] R. Gremmelmaier, *Phys. Verhandlungen*, vol. 7, pp. 196; 1957.
[4] R. Wiesner, "Halbleiterprobleme III," Friedr. Vieweg & Sohn, Braunschweig, pp. 59–74; 1956.

$$i_k = eg(L_1 + L_2). \quad (1)$$

In (1) $L_1 = \alpha_1 L_n$, where α_1 is a function of d_p/L_n. Correspondingly, $L_2 = \alpha_2 L_p$, where α_2 is a function of d_n/L_p. Of practical interest are the following two cases:

1) $d_p \gg L_n$ and $d_n \gg L_p$, *i.e.*, the electrodes are many diffusion lengths away from the *p-n* junction. In this case $L_1 = L_n$ and $L_2 = L_p$, *i.e.*,

$$i_k = eg(L_n + L_p). \quad (1a)$$

Only the sum of the diffusion lengths is obtained by measuring the short-circuit current. If, however, the mobility ratio is great (as, for example, in GaAs) and the lifetimes in the *n* and *p* regions are comparable, one term will predominate. Then one obtains approximately the diffusion length of the carriers with the higher mobility, and from this one can calculate the lifetime.

2) The junction is situated close to the one electrode, while the other electrode is several diffusion lengths away, *i.e.*,

$$d_n \ll L_n, d_p \ll L_p, d_n \gg L_p \text{ or } d_n \ll L_p, d_n \ll L_n, d_p \gg L_n.$$

Then

$$i_k = egL_p \quad \text{or} \quad i_k = egL_n. \quad (1b)$$

In this way *one* diffusion length is obtained directly from the measurement. This result can be easily obtained with alloyed and diffused *p-n* junctions if the diffusion length is not too small. Conditions are even more favorable if the measurements can be made on a surface-barrier layer. Then one diffusion length only is obtained from the short-circuit current even if this diffusion length is very small.

The lower limit for the application of this method is reached when the diffusion lengths become comparable with the width W of the space-charge region. All the minority carriers generated by the irradiation in the space-charge region are drawn by the inner electric field of this region into the neighboring region before they can recombine. Consequently, the more exact equation for the short-circuit current is

$$i_k = eg(L_n + L_p + W). \quad (2)$$

If the diffusion lengths are very small, $i_k \approx egW$. Here it is possible to determine the width of the space-charge region from a capacity measurement and then to calculate the diffusion length from the short-circuit current. The accuracy of this method will, however, no longer be good for $L < W$. If W is of the order of some tenths of a micron, it will still be possible to measure diffusion lengths of 1 μ with sufficient accuracy.

In some cases it is easier to measure the open-circuit voltage than the short-circuit current. Since, with small radiation intensities, the open-circuit voltage is proportional to the short-circuit current (for voltages under $\frac{1}{10}\ kT/e$), the diffusion length can be easily determined from the open-circuit voltage too. It is only required to determine, in addition, the zero-voltage junction resistance R_0 of the *p-n* junction from the volt ampere characteristic without irradiation.

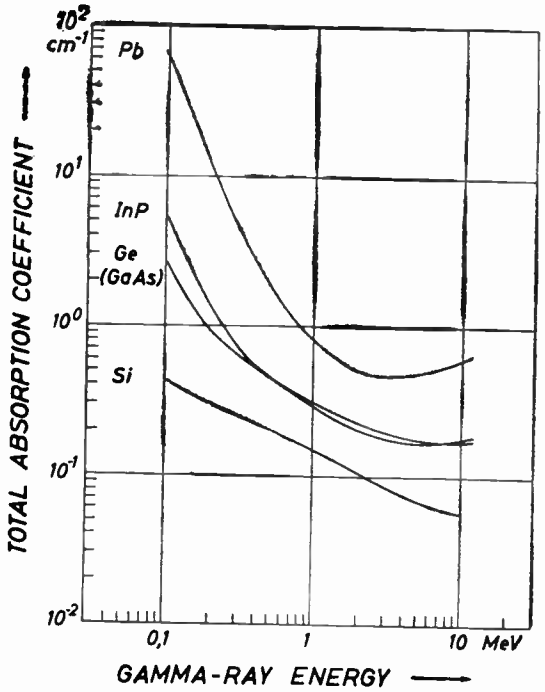


Fig. 2—Total absorption coefficient of γ rays in InP, Ge (GaAs), Si and Pb as a function of the energy of the γ rays.

III. Calculation of the Generation Rate

γ rays exhibit an exponential absorption in matter. The number N_γ of the γ quanta in a homogenous material decreases exponentially with the distance of penetration x, *i.e.*,

$$N_\gamma(x) = N_\gamma(0) \exp(-\mathcal{K}x). \quad (3)$$

The total absorption coefficient $\mathcal{K}$ is composed additively of the absorption coefficient of the three absorption processes: 1) photoelectric effect ($\mathcal{K}_{ph}$), 2) Compton effect ($\mathcal{K}_c$), and 3) pair production ($\mathcal{K}_p$)[5,6]:

$$\mathcal{K} = \mathcal{K}_{ph} + \mathcal{K}_c + \mathcal{K}_p. \quad (4)$$

Fig. 2 shows the total absorption coefficient for γ rays in InP, GaAs (Ge), and Si and, for comparison, in Pb as a function of the energy of the γ rays from 0.1 to 10 mev.[7]

During the decay of a Co^{60} atom, one γ quantum of 1.17 mev and one of 1.33 mev are emitted. In this energy range the absorption in Si, Ge, GaAs, and InP takes place almost exclusively by Compton effect. The two other processes can be practically ignored. (The photoelectric effect is still most strongly observed in InP. However, $\mathcal{K}_{ph}$ amounts only to approximately 5 per cent of $\mathcal{K}_c$.) This makes calculation of the generation rate considerably easier.

In Fig. 2 it can be seen that the absorption coefficient in the relevant energy range of the Co^{60} radiation is less

[5] W. Heitler, "The Quantum Theory of Radiation," Oxford University Press, New York, N. Y.; 1944.
[6] E. G. Segre, "Experimental Nuclear Physics," John Wiley and Sons, Inc., New York, N. Y.; 1953.
[7] Calculated according to Landolt-Bornstein, "Physikalische-Chemische Tabellen," Springer-Verlag, Berlin, vol. 1; 1952. See, pt. 5, pp. 351–360.

than 1 cm^{-1} for all semiconductors examined, *i.e.*, the mean free path of the γ rays is a few centimeters long. Over the range of a few millimeters—this is large compared with the diffusion length—the absorption in the semiconductor can therefore be regarded with sufficient accuracy as constant. Consequently, $dN_\gamma = N_\gamma \mathcal{K} L$ γ quanta per unit time are absorbed in a layer of thickness L, where N_γ is the number of quanta per unit time incident on the layer.

Fig. 3 illustrates further the processes connected with the absorption. (In the following we restrict ourselves to a discussion of the Compton effect which is alone important for calculating the generation rate.) For each γ quantum absorbed, one Compton electron is obtained which is preferentially emitted in the forward direction at a γ energy of 1.2–1.3 mev. The energy distribution and distribution in angle of the Compton electrons are schematically represented in Fig. 3 (the length of the arrows should be taken as a measure of the energy of the electrons emitted in this direction). The Compton electrons produce electron-hole pairs along their path and, on an average, $\overline{E}_c/\epsilon$ pairs are obtained if $\overline{E}_c$ is the average energy of a Compton electron and ϵ the average energy required to produce one electron-hole pair.

One further point should be noted here. In order to calculate the diffusion length from the short-circuit current, it is necessary to know the generation rate g within a volume on both sides of the p-n junction which is, roughly speaking, limited by the diffusion lengths (shaded part in Fig. 3). The range of the Compton electrons now lies in the order of 1 mm, *i.e.*, it is generally much greater than the diffusion length. Consequently, the Compton electrons produced within this shaded part lose only a fraction of their energy in this volume. Therefore, they do not generate $\overline{E}_c/\epsilon$ electron-hole pairs there but considerably fewer. The calculation is easily carried through if the distance of the shaded part from the surface is greater than the maximum range of the Compton electrons. Then the same number of electrons enter this region as leave it and the number of electron-hole pairs produced therein per unit time is equal to the number of Compton electrons produced in this region, multiplied by the number of electron-hole pairs generated on an average by one Compton electron. It is assumed that the lateral dimension of the p-n junction is very large compared with the range of the Compton electrons. So the electrons leaving the region at the edge and the electrons entering the region from the edge can be neglected. Thus the generation rate is given by

$$g = N_\gamma \mathcal{K} \frac{\overline{E}_c}{\epsilon} \tag{5}$$

and the short-circuit current

$$I_k = eN_\gamma \mathcal{K} \frac{\overline{E}_c}{\epsilon} L. \tag{6}$$

N_γ is the number of γ-quanta incident per unit area and unit time. It can be determined from the activity of the γ-ray source and the distance. For the Compton effect, $\mathcal{K} = 0.13$ cm^{-1} in silicon, $\mathcal{K} = 0.27$ cm^{-1} in gallium arsenide and germanium, and $\mathcal{K} = 0.24$ cm^{-1} in indium phosphide. The average energy of the Compton electrons is 0.59 mev,[5] and the average energy required to produce one electron-hole pair in silicon is $\epsilon = 3.6$ ev;[8] no great error will be made by taking $\epsilon = 4$ ev for GaAs and InP.[9]

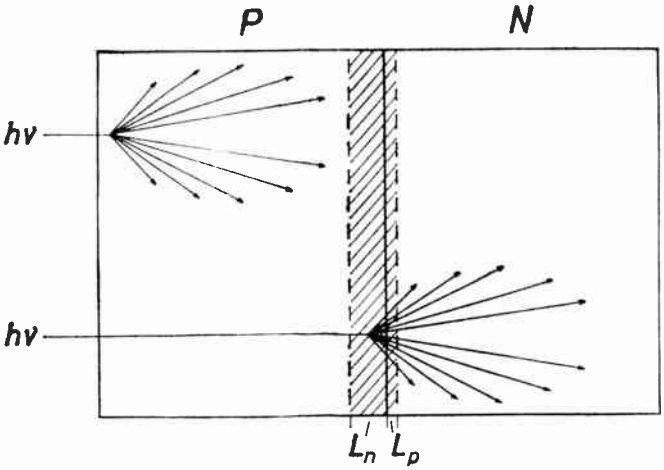


Fig. 3—Schematic representation of the absorption of γ rays by Compton effect.

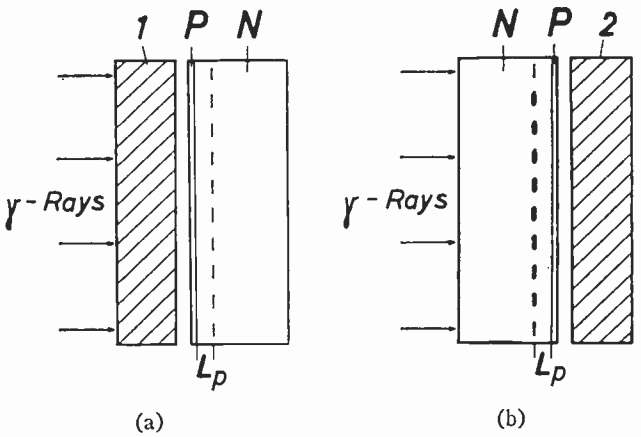


Fig. 4—Irradiation of a p-n junction by γ rays: (a) p-n junction near the irradiated surface. (b) p-n junction on the reverse side of the crystal.

Since, in most practical measurements, the p-n junction lies very near the one surface or electrode (cf. Section II), we must discuss how the generation rate g is modified by this. Two cases should be considered: 1) the p-n junction lies on the side of the crystal on which the irradiation is incident [Fig. 4 (a)], 2) the p-n junction lies on the reverse side of the crystal [Fig. 4 (b)]. In both cases the thickness of the semiconductor crystal should be large compared with the range of the Compton electrons.

The generation rate can be determined in both cases from (5) if a slice of the same material is placed in front of or behind the crystal (1 or 2 in Fig. 4), and if the

[8] K. G. McKay and K. B. McAfee, "Electron multiplication in silicon and germanium," *Phys. Rev.*, vol. 91, pp. 1079–1084; September 1, 1953.

[9] Based on measurement made by H. Pfister, "Electron irradiation of p-n junctions," *Z. Naturforschg*, vol. 12a, pp. 217–222; March, 1957.

ϵ in GaAs has a highest value of 6.3 ev. However, this value is surely too high.

thickness of this slice is greater than the range of the Compton electrons.

In case 1), it makes little difference if slice 1 is replaced by a piece of another material: the range of electrons in matter is inversely proportional to the density of the material.[10] If the density of slice 1 is greater, the range of the Compton electrons in this layer is consequently smaller. On the other hand, the absorption coefficient for the Compton effect is proportional to $\rho(z/A)$ (ρ = density, z = atomic number, A = atomic weight).[5] So with a greater density, correspondingly more Compton electrons per unit volume are produced. The number of Compton electrons incident on the surface of the semiconductor crystal therefore remains almost unchanged. There is only a slight dependence on z/A. When the atomic number increases, the number of Compton electrons decreases slightly. With a very great atomic number (with Au, Pb, *etc.*), the number of the incident electrons once more increases since photoelectrons as well as Compton electrons are now produced in the layer.

In case 2), the backscattering of the electrons entering slice 2 from the semiconductor becomes noticeable.[7] The percentage of backscattered electrons increases with the atomic number. If the atomic number of slice 2 is considerably greater than the atomic number of the semiconductor, the short-circuit current is expected to be higher than in the case where the atomic numbers are the same.

Measurement on silicon p-n junctions with slices of aluminum, brass, indium, and lead gave the following result. With an arrangement according to Fig. 4(a) and the same γ-ray density, the short-circuit current measured changes qualitatively in the manner discussed above. The subsequent increase for higher atomic numbers is slight. When using lead, the short-circuit current lies somewhat below the short-circuit current when aluminum is used. In all, the short-circuit current changes only by 10–12 per cent.

Investigations with GaAs p-n junctions gave a similar behavior of the short-circuit current as function of the atomic number. Here the short-circuit current is changed by a total of 20–22 per cent. This can be explained by the fact that, with small diffusion lengths, the electrons striking the semiconductor from outside contribute relatively more to the short-circuit current than is the case with large diffusion lengths. The diffusion length in GaAs is only a few microns, compared with 100–400 μ in the silicon measured.

Measurements of silicon p-n junction in an arrangement according to Fig. 4(b) showed a steady rise in the short-circuit current with the atomic number of the material of slice 2. The changes in the short-circuit current amounted to 30–40 per cent (from $z=13$ to $z=82$).

The influence of the electrodes can be described in the same manner. If the atomic number of the electrode material deviates widely from the atomic number of the semiconductor material, it is best to make the measurements in an arrangement according to Fig. 4(a), that is, with the p-n junction near the irradiated surface. The error will then be within tolerable limits.

[10] Landolt-Bornstein, *op. cit.*, pp. 325–351.

IV. Experimental Arrangement

From (6) a short-circuit current of approximately 5×10^{-12} A/cm^2 is obtained with a diffusion length of 1μ and 10^7 γ quanta per cm^2-sec. One curie of Co^{60} emits $2\times3.7\times10^{10}$ quanta per sec. Consequently 10^7 quanta per cm^2-sec are obtained in a distance of about 25 cm. A cobalt source of this magnitude is adequate for most measurements. The experimental arrangement is very simple. All that is needed to measure the short-circuit currents is a sensitive galvanometer or, still better, a dc amplifier.

For the measurements described in Section V; a Co^{60} source of approximately 300 millicuries was used. The samples were contained in a thin, light-tight aluminum box. The distance between the γ-ray source and the samples could be varied between 10 cm and 100 cm. The short-circuit current and, in some cases, also the open-circuit voltage, were measured with a Leeds and Northrup dc $\mu\mu$a amplifier or dc μv amplifier and a series-connected Speedomax.

V. Results

Measurements were carried out on alloyed and diffused silicon p-n junctions[11] and on diffused GaAs and InP p-n junctions. The alloyed silicon p-n junctions were a few microns below the one electrode and the silicon slices were about 1.5 mm thick. Diffusion lengths of the electrons between 350 and 450 μ were measured in one group of p-n junctions. In a second group, which had been manufactured by a different process, the diffusion lengths of the electrons were between 130 and 170 μ. Both groups correspond to a series of rectifiers in which Herlet has determined the diffusion length by another method.[12]

For comparison, the lifetime in the group with the smaller diffusion lengths was measured by the method of Lederhandler and Giacoletto.[1] The deviations of these lifetimes from the values determined by the measurements of the short-circuit current amounted to a maximum of 20 per cent.

In diffused silicon p-n junctions diffusion lengths of the holes between 75 μ and 110 μ were measured. The p material at the surface was approximately 1 μ thick.

[11] The author would like to express his thanks to Mr. Patalong of the Pretzfeld Lab. of the Siemens-Schuckerwerke, and Dr. Wiesner of the Halbleiterfabrik, Munchen of the Siemens, and Halske AG for the silicon p-n junction put at his disposal.

[12] A. Herlet, "Determination of the diffusion length L and the inversion density n_i from the forward characteristics of alloyed silicon area rectifiers," *Z. Angew Phys.*, vol. 9, pp. 155–158; April, 1957. Herlet measured the diffusion length as a function of the concentration of injected carriers at high injection levels, whereas our method gives the value of the diffusion length at very small deviations from the equilibrium density of the carriers.

Contrary to the silicon p-n junctions, the p-n junctions in GaAs and InP were situated more in the interior of the material. In each case only the sum of the diffusion lengths could be obtained from the short-circuit current. Because of the great mobility ratio, it is, however, possible to equate the sum of the diffusion lengths approximately with the diffusion length of the electrons without incurring too great an error. The maximum diffusion lengths measured and the lifetimes determined therefrom are summarized in Table I.

TABLE I

	L_{max}	τ_{max}
GaAs	8μ	9×10^{-9} sec
InP	130μ	2×10^{-6} sec

The accuracy of the measurements is determined by the degree of accuracy by which the values in (5) are known and by the magnitude of any disturbance due to the electrodes. $\overline{E}_e$ and $\mathcal{K}$ are known with sufficient accuracy. For ϵ the possible deviation is 10 per cent in the case of Si.[8] With GaAs and InP the deviation may be greater. The error in determining N_γ is approximately 10 per cent (inaccuracy in the measurement of the activity of the cobalt source). These errors in determining ϵ and N_γ give a systematic deviation in the results. Errors due to the influence of the electrodes and due to unfavorable position of the p-n junction can be kept small by using a careful experimental arrangement (cf. Sections II and III).

Formation of Junction Structures by Solid-State Diffusion*

F. M. SMITS†

Summary—**The diffusion of group III and group V impurities into germanium and silicon is reviewed. Observed and possible variations of the diffusion coefficient with concentration are discussed, followed by a summary of the diffusion coefficients and of solutions to the diffusion equation. Finally, methods for the evaluation of diffused layers and diffusion techniques are described.**

Introduction

THE formation of p-type and n-type regions within a single crystal of semiconductive material is the basis for the fabrication of semiconductor devices. This paper is concerned with the application of solid-state diffusion techniques as a method of junction preparation. Since most studies of solid-state diffusion in semiconductors have been performed during the past decade, only a few review articles cover the subject.[1–5] So far, most published studies of impurity diffusion in semiconductors have been concerned with diffusion in germanium and silicon while the studies in other semiconductors, such as III-V compounds, are still in early stages. In silicon and germanium, the elements of group III and group V are the common acceptor- and donor-type impurities which are used in the preparation of junction structures. This discussion is restricted therefore to the diffusion of these elements in silicon and germanium.

The diffusion coefficient is the basic parameter in a diffusion process and is discussed first. Then, solutions to the diffusion equation are given which correspond to configurations encountered in practical applications, followed by a discussion of methods for the evaluation of diffused layers. Finally, typical diffusion processes are reviewed.

Theory

Differential Equation of Diffusion

Diffusion processes generally are described in terms of a diffusion coefficient D, which is the factor of proportionality relating the flow density $\vec{F}$ of diffusing atoms with the concentration gradient $\vec{\nabla}N$ of the atoms,

$$\vec{F} = -D\vec{\nabla}N. \tag{1}$$

The negative sign expresses the fact that diffusion takes place down the concentration gradient. This definition

* Original manuscript received by the IRE, March 27, 1958.
† Bell Telephone Labs., Inc., Murray Hill, N. J.

[1] N. B. Hannay, "Recent Advances in Silicon" in "Progress in Semiconductors," Heywood and Co., London, Eng., vol. 1, p. 3; 1956.

[2] G. C. Dacey and C. D. Thurmond, "P-N junctions in silicon and germanium: principles, metallurgy, and applications," *Metallurgical Rev.*, vol. 2, pp. 157–193; June, 1957.

[3] C. S. Fuller, "Diffusion Techniques" in "Transistor Technology," D. Van Nostrand Co., Inc., New York, N. Y., vol. 3, ch. 3A, in press.

[4] H. Reiss and C. S. Fuller, "Diffusion and Precipitation in Germanium and Silicon," in "Semiconductors," Reinhold Publishing Corp., New York, N. Y., ch. 7, in preparation.

[5] J. N. Hobstetter, "Equilibrium Diffusion and Imperfections in Semiconductors" in "Progress in Metal Physics," Pergamon Press, Ltd., London, Eng., vol. 7, in preparation.

of the diffusion coefficient dates back to Fick (1855)[6] and the form of (1) can be justified by more fundamental treatments of the diffusion process.

In the general case, the diffusion coefficient is a tensor, since the rate of diffusion can be anisotropic. However, in a cubic lattice, symmetry requires an isotropic rate of diffusion and D in such cases must be a scalar. This fact simplifies the treatment of diffusion processes in silicon and germanium. Since we are generally interested in plane parallel junction structures, we are able to restrict ourselves to the simplest case of diffusion flow in one dimension.

Furthermore, we are mostly interested in single crystal semiconductors, so that complications, such as grain boundary diffusion, do not enter. The density of the diffusing impurities generally is small enough that we can neglect changes in dimensions which occur during diffusion. Accordingly, we can use the crystal lattice as a frame of reference.

The application of the continuity equation to (1) leads to what is known as Fick's Second Law:

$$\frac{\partial N}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial N}{\partial x}\right). \tag{2}$$

Only, if the diffusion coefficient is a constant, (2) simplifies to

$$\frac{\partial N}{\partial t} = D\frac{\partial^2 N}{\partial x^2}. \tag{3}$$

Generally, the assumption of a constant D has been made for the case of impurity diffusion in semiconductors and the diffused distributions are analyzed in terms of solutions to (3). These solutions are readily available, while solutions to (2) in most cases, require numerical approximations. Before applying (3) to the diffusion of group III and group V elements in silicon and germanium, we wish therefore to examine the validity of the assumption of a constant D.

Diffusion Coefficients

By general considerations, for an ideal solution the diffusion coefficient can be shown to be related to the microscopic mobility G (*i.e.*, the velocity attained by an atom under unit applied force) by the relation (valid for a single mechanism operating)[7]

$$D_i = kTG \tag{4}$$

where k is Boltzmann's constant and T is the absolute temperature.

The subscript i indicates that no force fields are affecting the motion of the atoms. Departure from ideality (activity coefficient not unity) can be attributed to force fields and will modify the diffusion coefficient. Certainly, a change in the microscopic mobility G also affects the diffusion coefficient. Both sources of a variation in D can be encountered in the diffusion of impurities in semiconductors.

Effects of "Built-In" Fields: For the diffusion of group III and group V acceptors and donors, only effects of electrical fields appear to cause departures from an ideal solution; ion pairing[8] and compound formation[9,10] should be negligible for practical cases.[11] For a generalized treatment of diffusion, the effects of electric fields can be included in the chemical potential and the particle flow can be analyzed in terms of the gradient of the chemical potential. For the present case, such a treatment is identical to considering the force acting on an ionized impurity atom in an electric field E.[12] This force is given by $\pm qE$, with q as the electronic charge. We therefore can express the flow of impurity atoms by (we assume complete ionization):

$$\begin{aligned} F &= -kTG\frac{\partial N}{\partial x} \pm GqEN \\ &= -D_i\frac{\partial N}{\partial x} \pm \mu EN \end{aligned} \tag{5}$$

where $\mu = Gq$, the more familiar field mobility (velocity attained in a unit field).

The form of (5) is identical to the usual description of the flow of mobile carriers as a sum of a diffusion current and a field current. For the present case of impurity diffusion it is more convenient to identify (5) with (1), thus introducing a variable diffusion coefficient given by

$$D = D_i\left(1 \pm \frac{qEN}{kT\partial N/\partial x}\right). \tag{6}$$

Electric fields arise in an extrinsic semiconductor with a nonuniform impurity concentration due to the rapidly diffusing electrons and holes which, compared to the impurity atoms, come to equilibrium instantaneously. If, therefore, the density of diffusing impurity atoms approaches or exceeds the density of intrinsic carriers at the temperature of diffusion, (6) has to be considered. In most cases, E can be found from the neutrality condition. For the special case of only one diffusing impurity, (6) takes the form[13]

[6] A. Fick, "Ueber diffusion," *Ann. Phys.*, vol. 94, pp. 59–86; 1855.

[7] A. Einstein, "Ueber die von der molekularkinetischen Theorie der waerme geforderte Bewegung von in ruhenden Fluessigkeiten suspendierten Teilchen," *Ann. Physik*, vol. 17, pp. 549–560; July, 1905.
A. D. Le Claire, "Diffusion of Metals in Metals" in "Progress in Metal Physics," Pergamon Press, Ltd., London, Eng., vol. 1, p. 306; 1949.
L. S. Darken, "Diffusion, mobility and their interrelation through free energy in binary metallic systems," *Amer. Inst. Mining Metallurgical Engs.*, vol. 175, pp. 184–201; 1948.

[8] H. Reiss, C. S. Fuller, and F. J. Morin, "Chemical interactions among defects in germanium and silicon," *Bell Sys. Tech. J.*, vol. 35, pp. 535–636; March, 1956.

[9] H. Reiss, C. S. Fuller, and A. J. Pietruskiewicz, "Solubility of lithium in doped and undoped silicon, evidence for compound formation," *J. Chem. Phys.*, vol. 25, pp. 650–655; October, 1956.

[10] S. Zaromb, "An analysis of diffusion in semiconductors," *IBM J. Res. Dev.*, vol. 1, pp. 57–61; January, 1957.

[11] H. Reiss, private communication.
H. L. Frisch, private communication.

[12] H. J. Oel, "Diffusion von Ionen und Elektronen," *Z. physik. Chem., Frankfort*, vol. 10, pp. 165–183; February, 1957.

[13] Derived by F. X. Hassinon.

$$D = D_i\left(1 + \frac{N}{\sqrt{(2n_i)^2 + N^2}}\right) \quad (7)$$

with n_i the density of intrinsic carriers. This form holds equally for donor- and acceptor-type impurities since the fields are always in such a direction as to retard the mobile carriers. This implies that these fields always aid the oppositely charged impurities. The effect can account therefore for a change of D by a factor of two. Howard observed a variation of D with concentration for phosphorus diffusion in silicon, which could be fitted by (7).[14] He evaluated his results in terms of (3) which means that the diffusion coefficients quoted are some average values.

Mobility Variations: Variations in the microscopic mobility G will result in a variation of D_i. Such mobility variations will be closely related to the particular diffusion mechanism.

Experiments by Valenta and Ramasastry showed an effect of heavy doping on the self-diffusion in germanium.[15] Their experiments were performed in order to determine the nature of the self-diffusion mechanism, in particular, if it occurs by the vacancy mechanism. For a vacancy mechanism the mobility term should be proportional to the density of vacancies, and it should be possible to change the mobility term by changing the vacancy density.[16] A method for varying the vacancy density is suggested by Reiss, Fuller, and Pietruszkiewicz[9] if one assumes that vacancies act as acceptors. Reiss and co-workers conducted experiments in which they were able to confirm Reiss' prediction[17] that the solubility of donors and acceptors can be influenced by the position of the Fermi level. The position of the Fermi level in turn can be controlled by donors and acceptors already present in the crystal. In particular, Reiss showed that, to a first approximation, the solubility should be proportional to the fraction of neutral impurities. Consequently, the solubility of an acceptor, in comparison to intrinsic material, should be increased in n-type material and decreased in p-type material. Since in germanium vacancies are found to act as acceptors, the self-diffusion in germanium, if proceeding by the vacancy mechanism, should be enhanced in n-type material and retarded in p-type material. This, indeed, is the finding of Valenta and Ramasastry.

Since other possible diffusion mechanisms would not show such a behavior, these experiments must be considered as proof that the self-diffusion in germanium proceeds by the vacancy mechanism. By analogy, one can assume that the self-diffusion in silicon and the diffusion of group III—group V elements in silicon and germanium proceeds by the vacancy mechanism also. Therefore, this could result in a similar dependence of the mobility term for impurity diffusion.

There is a distinct difference between self-diffusion and impurity diffusion, since a diffusing impurity itself is a doping agent and thus can change the location of the Fermi level. The diffusion coefficient therefore becomes dependent on concentration once the impurity concentration approaches or exceeds the density of carriers for the intrinsic semiconductor at the temperature of diffusion. With vacancies acting as acceptors, the effect would result in a decrease of the diffusion coefficient of an acceptor and an increase for a donor. However, the exact dependence might be more complicated.

So far, no conclusive experimental evidence for such an effect on the diffusion of impurities has been reported, and further experimental studies are necessary. In particular, the electrical behavior of vacancies in silicon is still uncertain.

The previously mentioned results of Howard do not cover a range wide enough to ascertain a constant D (equal to $2D_i$) for high concentrations. The presence of a vacancy effect in this case could result in an increase of D above the value $2D_i$. It would be interesting to obtain similar results on the diffusion of acceptors where the field effect and the vacancy effect would tend to affect D in opposite directions.

Recently, Karstensen found an effect of lattice defects on the rate of diffusion.[18] He reports a preferential diffusion of *Sb* along small angle boundaries in germanium. In particular, he finds that the increased rate of diffusion occurs only in the direction of the lines of dislocations formed by the small angle boundaries; no increase could be observed in a direction perpendicular to the line of dislocations.

Values of the Diffusion Coefficients: From the foregoing we conclude that in a good single crystal the diffusion coefficients of group III and group V elements vary only if the impurity density exceeds the intrinsic density of carriers at the temperature of diffusion. These densities are generally fairly high. In many cases, therefore, (3) is strictly valid. For higher concentrations the use of (3) may introduce errors.

The diffusion coefficients reported in the literature are based on experiments evaluated with solutions to (3). They generally have a temperature dependence of the form

$$D = D_0 \exp -(\Delta Q/kT) \quad (8)$$

where ΔQ is an activation energy.

Table I and Table II give the values for D_0 and ΔQ for the group III and group V elements in germanium and silicon respectively.

[14] B. T. Howard, "Phosphorus Diffusion in Silicon," presented at the Fall Meeting of the Electrochemical Society, Buffalo, N.Y., October, 1957, and *J. Electrochem. Soc.*, to be published.

[15] M. W. Valenta and C. Ramasastry, "Effect of heavy doping on self-diffusion of germanium," *Phys. Rev.*, vol. 106, pp. 73–75; April 1, 1957.

[16] R. L. Longini and R. F. Greene, "Ionization interaction between impurities in semiconductors," *Phys. Rev.*, vol. 102, pp. 992–999; May, 1956.

[17] H. Reiss, "Chemical effects due to the ionization of impurities in semiconductors," *J. Chem. Phys.*, vol. 21, pp. 1209–1217; July, 1953.

[18] F. Karstensen, "Preferential diffusion of antimony along small angle boundaries in germanium and the dependence of this effect on the direction of the dislocation-lines in the boundary," *J. Elect. and Control*, vol. 3, pp. 305–307; September, 1957.

TABLE I
DIFFUSION COEFFICIENTS IN GERMANIUM

Element	D_0(cm²/sec)	ΔQ(eV)	References
B	1.8 10^3	4.6	19
Ga	35	3.1	19
In	0.15	2.6	22, 19, 21
Tl	0.09	2.7*	23
P	3.6	2.5	19
As	3.6	2.4	22, 19, 21, 20
Sb	1.2	2.3	22, 19, 21
Bi	4.7	2.4	23
Ge	7.8	2.98	24

* Estimated value.

TABLE II
DIFFUSION COEFFICIENTS IN SILICON

Element	D_0(cm²/sec)	ΔQ(eV)	References
B	14	3.7	25, 26
Al	5.5	3.4	25
Ga	4.0	3.5	25
In, Tl	40	4.0	25
P	1500*	4.4	14, 25
As	0.55	3.6	25
Sb	10	4.0	25, 26
Bi	2200	4.7	25

* For intrinsic silicon.

Solutions to Diffusion Equation

The form of actual impurity distributions can be calculated from the differential equation subject to certain initial conditions and certain boundary conditions. The initial conditions are specified by the spatial dependence of the impurity concentration at time zero, while the boundary conditions specify conditions for the concentration or the flow of impurities at certain points or planes. These boundary conditions may have a time dependence. The differential equation describing a diffusion process is mathematically identical to the differential equation describing heat conduction. A great variety of solutions are available in the literature.[27–32] We discuss certain solutions which are applicable to practical cases; in particular, only boundary conditions which are constant in time.

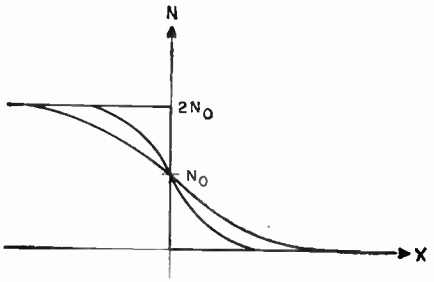


Fig. 1—Diffusion from a concentration step.

Diffusion from a Concentration Step and from a Constant Surface Concentration: Only the initial condition has to be considered for the diffusion from a concentration step in an otherwise infinite solid. Such an initial condition is encountered in grown-diffused and diffused-meltback techniques. The initial condition can be specified by (Fig. 1):

$$N = 2N_0 \qquad \text{for } x < 0,$$

$$N = 0 \qquad \text{for } x > 0.$$

The diffusion from such an initial condition is described by

$$N = N_0 \operatorname{erfc}(x/2\sqrt{Dt}) \tag{9}$$

with

$$\operatorname{erfc} y = \frac{2}{\sqrt{\pi}} \int_y^\infty e^{-\xi^2} d\xi.$$

At the point $x=0$, the concentration at all times has the constant value $N(0)=N_0$. The same equation describes therefore the diffusion from a constant surface concentration N_0 into a semi-infinite solid bounded at $x=0$.

Outward Diffusion: Such a reduction of a problem involving a semi-infinite solid to a problem of an infinite solid is typical for cases involving constant surface concentrations. Hereby, we can use purely mathematical terms, like negative concentrations. As an example, consider an infinite solid with a step in concentration from $-N_0$ to $+N_0$ at $x=0$ (Fig. 2). For such an initial condition the diffused distribution is described by

$$N = N_0[1 - \operatorname{erfc}(x/2\sqrt{Dt})] \equiv N_0 \operatorname{erf}(x/2\sqrt{Dt}). \tag{10}$$

For this case the concentration accordingly remains 0 at $x=0$ for all times. For $x>0$, the distribution describes the outward diffusion from an initially uniform concentration with 0 surface concentration at all times.

[19] W. C. Dunlap, Jr., "Diffusion of impurities in germanium," *Phys. Rev.*, vol. 94, pp. 1531–1540; June 1, 1954.
[20] K. M. McAfee, W. Shockley, and M. Sparks, "Measurement of diffusion in semiconductors by a capacitance method," *Phys. Rev.*, vol. 86, pp. 137–138; April, 1952.
[21] C. S. Fuller, "Diffusion of donor and acceptor elements into germanium," *Phys. Rev.*, vol. 86, pp. 136–137; April, 1952.
[22] W. Boesenberg, "Diffusion von Antimon, Arsen und Indium in festem Germanium," *Z. Naturforsch.*, no. 10a, pp. 285–291; 1955.
[23] C. S. Fuller, private communication.
[24] H. Letaw, Jr., W. M. Portenoy, and L. Slifkin, "Self-diffusion in germanium," *Phys. Rev.*, vol. 102, pp. 636–639; May, 1956.
[25] C. S. Fuller and J. A. Ditzenberger, "Diffusion of donors and acceptor elements in silicon," *J. Appl. Phys.*, vol. 27, pp. 544–553; May, 1956.
[26] W. C. Dunlap, Jr., H. V. Bohm, and H. P. Mahon, Jr., "Diffusion of impurities in silicon" (abstract), *Phys. Rev.*, vol. 96, p. 822; November, 1954.
[27] H. S. Carslaw and J. C. Jaeger, "Conduction of Heat in Solids," Clarendon Press, Oxford, Eng.; 1947.
[28] R. M. Barrer, "Diffusion in and through Solids," Cambridge University Press, Cambridge, Eng.; 1951.
[29] W. Jost, "Diffusion in Solids, Liquids and Gases," Academic Press, New York, N. Y.; 1952.
[30] W. Seith, "Diffusion in Metallen," Springer-Verlag, Berlin, Ger.; 1955.
[31] J. Crank, "The Mathematics of Diffusion," Clarendon Press, Oxford, Eng.; 1956.
[32] W. Jost, "Diffusion," D. Steinkopf, Darmstadt, Ger.; 1957.

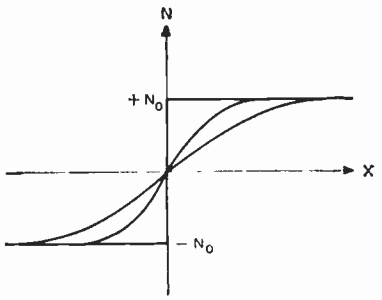


Fig. 2—Out diffusion.

In semiconductors the difference between donor concentration and acceptor concentration, $N_D - N_A$, determines the electrical behavior of the material. Therefore, we like to plot donor concentrations as positive and acceptor concentrations as negative. Now negative concentrations have a physical meaning and we can interpret the purely formal use of negative concentrations. The diffusion of an acceptor from a constant surface concentration is described by an initial condition of $N = -2N_0$ for $x < 0$ and $N = 0$ for $x > 0$. Since the differential equation is linear we can superpose solutions. Accordingly, the case of outward diffusion also can be described by specifying an initial condition of constant concentration $N = +N_0$ for $x \lesseqgtr 0$ superposed on a constant concentration $N = -2N_0$ for $x < 0$ (Fig. 3). The concentration $-2N_0$ alone describes an acceptor diffusing in the $+x$ direction from a surface concentration $N(0) = N_0$, while the constant concentration $+N_0$ alone describes a nondiffusing donor. The sum of both is described by (10).

We conclude therefore that the outward diffusion of a donor results in a distribution of electrically active impurities which is equivalent to the diffusion of an acceptor into a sample uniformly doped with a nondiffusing donor. Such a picture is helpful in considering, for example, the effect of outward diffusion of the original doping on the location of a *p-n* junction formed by the inward diffusion of another impurity.

Planar Source: The condition of constant surface concentration requires that the diffusion is the only rate determining factor with no additional rate limitation at the surface. The other extreme is a completely impermeable surface for which case the flow across $x = 0$ has to vanish for all times. To obtain the diffused distribution with such a boundary condition at the surface, the diffusing material has to be introduced into the solid prior to diffusion. The simplest configuration for such a case is a planar source at $x = 0$. Again, this case is readily analyzed by considering an infinite solid with a planar source at $x = 0$. This source is described by the total number of atoms per cm². If this number is $2\overline{N}_0$, then the diffused distribution is described by (Fig. 4):

$$N = \frac{\overline{N}_0}{\sqrt{\pi Dt}} \exp(x^2/4Dt). \qquad (11)$$

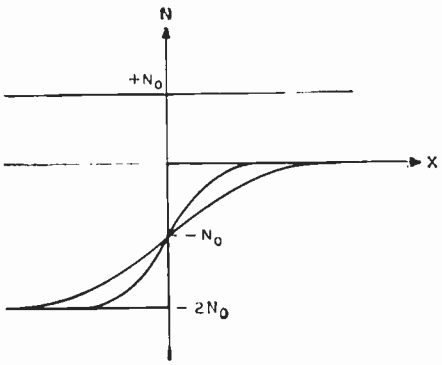


Fig. 3—Superposition equivalent to out diffusion.

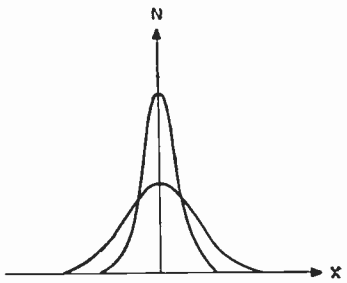


Fig. 4—Diffusion from a planar source.

This function is known as the Gaussian distribution. Clearly, the gradient at $x = 0$ vanishes for all times. This means that there is no flow across $x = 0$. Half of the material of the planar source can be found on the left, the other half on the right. Eq. (11) therefore describes the diffusion of impurities for a planar source of a sheet density $\overline{N}_0$ into a semi-infinite solid with an impermeable surface. Clearly, the surface concentration for such a distribution becomes a function of the time and can be described by

$$N(0) = \frac{\overline{N}_0}{\sqrt{\pi Dt}}. \qquad (12)$$

Rate Limitations at the Surface: For the diffusion into a surface, we considered two extremes thus far: 1) for the case of constant surface concentration, no rate limitation at the surface, and 2) for the Gaussian distribution, a completely impermeable surface. Clearly, if the surface would be impermeable for the case of constant surface concentration, no diffusion would be possible. On the other hand, for the Gaussian distribution with no rate limitation at the surface, the material would be lost instantaneously. With a finite rate limitation both cases would result in a diffused distribution but both would be modified. For the general case, we therefore have to consider a rate limitation at the surface which may be described as a potential barrier which an impurity atom must surmount in order to enter or to leave the crystal.

The appropriate boundary condition for diffusion with a rate limitation at the surface is equivalent to the so-called radiation boundary condition encountered in heat

conduction problems. For this reason the solutions for similar heat flow problems are applicable with a proper change in variables.[33]

After a diffusion time $t \to \infty$, the impurity concentration reaches a constant equilibrium value N_e which is determined by the external phase. Under actual conditions, the surface concentration therefore approaches this equilibrium value N_e. The most reasonable assumption one can make (and this is the only assumption that has been treated mathematically) is that the flow across the surface is proportional to the difference between the actual surface concentration and the equilibrium concentration. This flow must produce an equal diffusion flow within the solid which yields as a boundary condition

$$[N_e - N(0)]K = -D\partial N/\partial x|_{x=0}. \tag{13}$$

The constant K thus introduced describes the rate limitation at the surface. Clearly, $K = \infty$ requires that the surface concentration is always equal to the equilibrium concentration, which is the case of constant surface concentration treated previously. The case $K = 0$ means that there is no flow across the surface, the condition we required previously for the Gaussian distribution.

To obtain the distribution of an impurity in the solid for the equivalent cases treated before, we have to solve the diffusion equation subject to the boundary condition given in (13). The solutions for similar heat flow problems are applicable with a proper change in variables. The solutions are more easily expressed in terms of the parameters

$$y = x/2\sqrt{Dt}, \qquad z = \frac{K}{D}\sqrt{Dt}. \tag{14}$$

The more important solutions are as follows:

Case I (solution equivalent to the case of constant surface concentration) N_e = constant, 0 initial concentration in the solid:

$$N = N_e \exp(-y^2)\,(\exp(y^2)\operatorname{erfc}(y) - \exp[(y+z)^2]\operatorname{erfc}(y+z)) = f_1(y, z). \tag{15}$$

Case II (diffusion out of the solid) equilibrium concentration $N_e = 0$, uniform initial concentration N_0 in the solid:

$$N = N_0(1 - f_1(y, z)) = N_0 f_2(y, z). \tag{16}$$

Case III (solution equivalent to the Gaussian distribution) equilibrium concentration $N_e = 0$, planar source of sheet density $\bar{N}_0$ initially in the solid at the surface:

$$N = \frac{\bar{N}_0}{\sqrt{\pi Dt}}\exp(-y^2)(1 - z\sqrt{\pi}\exp[(y+z)^2] \times \operatorname{erfc}(y+z)) = \frac{N_0}{\sqrt{\pi Dt}} f_3(y, z). \tag{17}$$

[33] F. M. Smits and R. C. Miller, "Rate limitation at the surface for impurity diffusion in semiconductors," *Phys. Rev.*, vol. 104, pp. 1242–1245; December, 1956.

Certainly, the last function approaches the Gaussian distribution for $K \to 0 (z \to 0)$. With a finite K the distribution has a maximum. The distribution described in Case I approaches the erfc distribution for $z \to \infty$ ($K \to \infty$ or $t \to \infty$), while for the other extreme, $z \to 0$, the distribution is given by

$$N = N_e 2z \int_y^\infty \operatorname{erfc} \xi\, d\xi. \tag{18}$$

The last solution corresponds to a constant flux $N_e K$ across the surface. The surface concentration (which is a function of time) of such a distribution is given by $N(0, t) = N_e 2z/\sqrt{\pi}$.

Evaporating Surface: Diffusion may occur while the sample surface is evaporating.[34] Due to the evaporation, the surface of the sample is constantly receding with the velocity v. Under such a condition a diffused distribution reaches a steady state after a certain time. Under steady state, the surface concentration certainly will be independent of time even in the presence of a rate limitation at the surface, in which case the surface concentration would not correspond to the equilibrium concentration. By measuring depth from the surface of the sample, the steady-state distribution has the form

$$N = N(0) \exp -\left(\frac{v}{D}\, x\right). \tag{19}$$

Steady state is effectively reached for

$$t > \frac{D}{v^2}. \tag{20}$$

It is instructive to compare the impurity distributions described by (9), (11), (18), and (19). Fig. 5 gives such a comparison showing $N/N(0)$ against distance with the distance normalized in such a way that all distributions coincide for $N/N(0) = 10^{-3}$. The similarity between the case of no rate limitation (9) and the case of an extreme rate limitation (18) is particularly striking. This demonstrates that the *shape* of an impurity distribution for the case of a constant external phase is not significantly affected by a rate limitation at the surface. This finding is significant for the experimental evaluation of a diffused distribution.

Evaluation of Diffused Layers

Analysis of Profiles

The standard techniques involving radioactive isotopes[35] have been successfully applied to the study of impurity diffusion in semiconductors. These techniques are useful in finding the exact profile of an impurity dis-

[34] F. M. Smits, R. C. Miller, and R. L. Batdorf, "Surface Effects on the Diffusion of Impurities in Semiconductors," presented at the International Symp. on Semiconductors and Phosphors, Garmisch-Partenkirchen, Germany, 1956. Also to be published by F. Viweg, Braunschweig, Ger.

[35] See, *e.g.*, R. E. Hoffman, "Tracer and Other Techniques of Diffusion Measurements" in "Atom Movements," American Society for Metals, Cleveland, Ohio, p. 51; 1951.

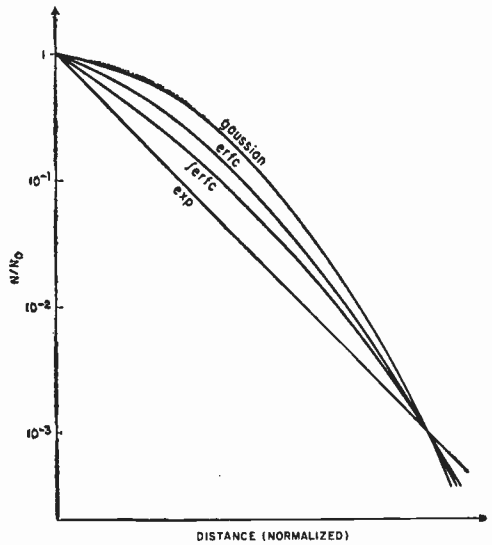


Fig. 5—Comparison between basic distributions.

tribution by lapping thin layers off a diffused sample and counting the activity of the impurities in the removed material.

The electrical effects of the impurities also are convenient in studying the distribution of the impurities. In particular, the sheet conductivity of a material is a measure of the sheet density of current carriers and, therefore, a measure of the sheet density of impurities. The difference in sheet conductivity before and after the removal of a layer, divided by the thickness of the removed layer, gives the average conductivity of the removed portion. The profile of an impurity distribution can be deduced easily from such measurements. However, the sheet conductivities have to be measured very accurately since the difference of two measurements is the number to be determined.

Measurement of Layer Thicknesses

The evaluation of a diffused layer is considerably simplified if one knows the functional relationship describing the diffused distribution. In most cases, two independent parameters are then sufficient to describe the exact profile.

For the case of an impurity diffused into material doped with an impurity of opposite conductivity type, an inversion layer results with a *p-n* junction occurring where $N_D = N_A$. The position of such a *p-n* junction is readily obtained by cross sectioning the sample. Such cross sectioning may be done perpendicularly to the surface and after locating the *p-n* junction, gives the true layer thickness. Since the layer depth frequently is very small, a better accuracy is obtained by cross sectioning at an angle and thus enlarging the thickness to be measured. From this, the true layer thickness is obtainable by either geometrical relationships, or one can employ optical interference techniques,[36] in which case the depth is obtainable in terms of the wave length of monochromatic light.

Various techniques are employed to locate *p-n* junctions. The thermoelectric voltage, occurring between hot and cold contacts to a sample, changes sign across the junction; similarly, the characteristic of a point contact rectifier changes sign at the junction. Also, one may probe the potential distribution across a reverse-biased junction, or one can electroplate only the *p*-type side of the junction by applying a proper bias to the *n*-type side. The photoelectric effect can also be used.

Preferential etching between *p*-type and *n*-type material delineates the junction. This technique, however, is only practical for a direct cross section. For surface-diffused samples, it requires that the samples be embedded in, for example, a plastic resin. For the case of silicon surface-diffused layers, a staining technique is preferable.[25] For this, the cross section is wet with a small drop of concentrated HF acid containing a trace of HNO_3 (0.1–0.5 per cent by volume). Under the proper conditions, the *p*-type regions turn dark (presumably an oxidation), thus sharply defining the *p-n* junction.

Measurement of Sheet Resistivities

A second important parameter for a diffused inversion layer is its sheet conductivity. For surface layers, the four-point probe technique[37] is suitable for the evaluation of the sheet conductivity or its inverse, the sheet resistivity ρ_s. Since, in most practical cases, the layer thickness is much thinner than the point spacing, the logarithmic potential distribution applies. For a linear four-point probe with equal point spacing, the sheet resistivity on an infinite sheet is obtained as

$$\rho_s = \frac{V}{I} \times \frac{\pi}{\ln 2} = \frac{V}{I} \times 4.5324 \cdots \qquad (21)$$

where V is the voltage between the inner points and I is the current flowing through the outer points. For finite geometries the method of images can be applied to obtain correction factors. For circular samples one image dipole is necessary, while for rectangular structures a two-dimensional, infinite array of dipoles has to be considered. The evaluation of this problem leads to an expression for the sheet resistivity of the form

$$\rho_s = \frac{V}{I} \cdot C. \qquad (22)$$

The correction factor C for various sample sizes is available in the literature.[38]

Surface Concentrations

The surface concentration of a diffused impurity distribution can be obtained from the layer thickness and

[36] W. L. Bond and F. M. Smits, "The use of an interference microscope for measurement of extremely thin layers," *Bell Sys. Tech. J.*, vol. 35, pp. 1209–1221; September, 1956.

[37] L. B. Valdes, "Resistivity measurements on germanium for transistors," PROC. IRE, vol. 42, pp. 420–427; February, 1954.

[38] F. M. Smits, "Measurement of sheet resistivities with the four point probe," *Bell Sys. Tech. J.*, vol. 37, pp. 711–718; May, 1958.

the sheet resistivity. In general, an impurity distribution can be expressed as

$$N = N_0 \cdot f(\lambda \cdot x)$$

whereby $\lambda = 1/2\sqrt{Dt}$ for the nonsteady-state distributions discussed before and $\lambda = v/D$ for the steady-state distribution (19). With N_b the original doping of the material, the sheet resistivity is given by

$$\begin{aligned} 1/\rho_s &= \int_0^a q\mu(N + N_b) \cdot (N - N_b)dx \\ &\approx N_0 \int_0^a q\mu(N) f(\lambda x)dx - N_b \int_0^a q\mu(N)dx \end{aligned} \tag{23}$$

where a is the layer thickness. The latter function can be brought into the form

$$\rho_s \cdot a = F(N_0, N_b). \tag{24}$$

In other words, the average resistivity of a diffused layer for a given type of distribution is only a function of the surface concentration and the impurity concentration in the parent material. For silicon the function $F(N_0, N_b)$ has been evaluated for the Gaussian distribution, the erfc distribution, and the exponential distribution.[39] A rate limitation at the surface makes the impurity distribution deviate from the erfc distribution. However, it has been shown that even for the extreme case—for the purpose of evaluating the surface concentration—the deviation in the shape of the impurity distribution is small enough to be negligible for most practical cases. As an additional parameter one can determine the gradient of an impurity distribution in a *p-n* junction by measuring the small-signal ac capacitance as a function of the reverse bias.[20,40] The simultaneous diffusion into two samples of different resistivities has also been employed to measure two points of an impurity distribution.[21]

Diffusion Techniques

The magnitude of the diffusion coefficients makes the diffusion of group III—group V impurities into silicon and germanium an ideal technique for the fabrication of diffused layers in a range of thickness from less than 10^{-4} cm to 10^{-2} cm or more. In designing a diffusion process, emphasis has to be given to the reproducibility of the results. A higher or lesser degree of control is required for each device design. For example, a diode requiring only a single diffusion usually does not require the same control on the diffusion parameters as is necessary for a double-diffused transistor. Accordingly, one may select a different diffusion process for a diode application than one would for a transistor application.

The diffusion coefficient, the time of diffusion, the initial conditions, and the boundary conditions determine the diffused distribution. As discussed before, the diffusion coefficient generally has an exponential temperature dependence. The control of the diffusion coefficient requires therefore a good control of the diffusion temperature. The dependence of the diffusion coefficient on the concentration of impurities does not represent an independent variable and need only be considered in the analysis of particular impurity distributions. The necessity for controlling the diffusion coefficient and the diffusion time is common to all diffusion techniques. A convenient classification of the various techniques can be made by considering the initial conditions and the boundary conditions involved in any particular process.

Diffusion of Impurities in Solution in the Solid

The simplest configuration is the diffusion from a concentration step in an infinite solid. A solid can be considered infinite if the length over which the diffusion is carried out is small compared to the thickness of the sample. A diffusion process involving such a principle reduces the control problem (besides the temperature and time control) to the control of the initial condition established prior to diffusion.

Diffused Meltback: A typical process for such a configuration is the diffused-meltback process.[41–43] The process requires the simultaneous diffusion of at least two elements to give *n-p-n* or *p-n-p* structures. As an example we discuss this technique as applied to the fabrication of *n-p-n* structures in silicon which requires a single crystal of silicon intentionally doped with both donor and acceptor impurities. The doping concentrations are such that the crystal grows uniformly *n*-type with a low resistivity. Bars of such material are partially remelted and solidified again. Since the impurities are less soluble in the solid silicon than in the liquid, the impurities segregate as the melted portion freezes, resulting in a sharp drop in impurity concentrations to very low values determined by the initial concentrations and the segregation coefficients of the impurities. By proper choice of the impurities and their concentrations in the crystal, the conductivity in the regrown region will remain *n*-type. The region of low concentration extends only over a relatively short distance and increases again to higher values because, due to desegregation, the impurity concentration in the liquid phase increases during the regrowth resulting in an impurity

[39] G. Backenstoss, "Evaluation of surface concentration of diffused layers in silicon," *Bell Sys. Tech. J.*, vol. 37, pp. 699–709; May, 1958.

[40] K. Lehovec, K. Schoeni, and R. Zuleeg, "Evaporation of impurities from semiconductors," *J. Appl. Phys.*, vol. 28, pp. 420–423; April, 1957.

[41] K. Lehovec and A. Levitas, "Fabrication of multiple junctions in semiconductors by surface melt and diffusion in the solid state," *J. Appl. Phys.*, vol. 26, pp. 106–109; January, 1957.

[42] I. A. Lesk and R. E. Gonzalez, "Germanium and silicon transistor structures by the diffused-meltback process employing two or three impurities," IRE Trans. on Electron Devices, vol. ED-5, No. 3; July, 1958.

[43] A. B. Philips and A. M. Intrator, "A new high frequency *n-p-n* silicon transistor," 1957 IRE National Convention Record, pt. 3, pp. 3–14.

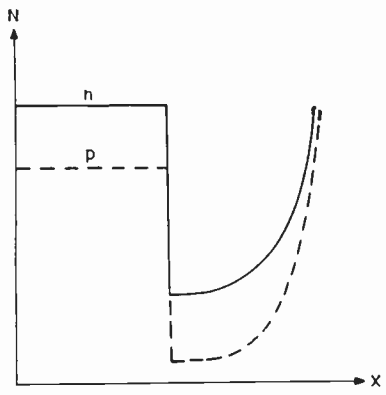


Fig. 6—Diffused-Meltback initial condition.

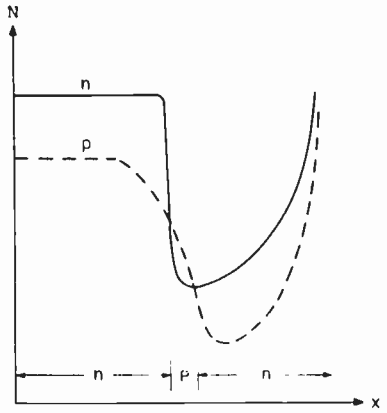


Fig. 7—Diffused-Meltback final impurity distribution.

distribution as shown in Fig. 6. This meltback procedure results therefore in a concentration step for both the donor and the acceptor impurity exactly at the plane to which the meltback was carried. This impurity distribution represents the initial condition for the diffusion. Both the donor impurity and the acceptor impurity are diffused and form distributions described by (9), superposed on constant concentrations corresponding to the concentrations in the regrown region. If the acceptor impurity has a significantly higher diffusion constant than the donor impurity, a p-type region is formed between the region of high impurity concentration and the region of low impurity concentration, resulting in a n-p-n structure (Fig. 7). The increase in net impurity concentration during the regrowth is a very desirable feature since it reduces the series resistance without reducing the breakdown voltage of the adjacent p-n junction.

The limitations of such a technique are given by simultaneous requirements on the segregation coefficients and the diffusion constants of the impurities. Since, in silicon, the acceptor impurities diffuse faster than the donor impurities, it is possible to produce n-p-n structures as discussed above. In germanium the opposite is true and p-n-p structures can be made. For best results with germanium, three impurities are required for practical reasons.[42]

Grown-Diffused: A similar technique is the grown-diffused process.[44,45] In this process the step of impurity concentration required as an initial condition is produced during the crystal growth itself. After growing the collector region, growth is stopped, the base and emitter producing impurities are added, and growth is resumed. During the growth of the emitter region, the base producing impurities diffuse into the collector region and thus produce a narrow base region. The thickness of the base region depends upon the relative doping levels, the impurities used, and the growth rate and time taken to grow the emitter region. In germanium, p-n-p structures can be produced by this procedure using gallium and arsenic as the impurities. In silicon, n-p-n structures are possible with the combinations arsenic-aluminum and arsenic-boron as doping impurities.[44]

Out-Diffusion: In both previous techniques the control of the boundary conditions during the actual diffusion cycle is completely eliminated and the control of the concentrations in the diffused distributions is entirely reduced to procedures preceding the actual diffusion step (mainly the crystal growing). Contrary to this, all diffusions from the surface require control of boundary conditions at the surface. The case of outward diffusion of impurities from a crystal, uniformly doped with impurities, takes a special place. As discussed previously, outward diffusion of one impurity is equivalent to inward diffusion of an impurity of opposite conductivity type into a semiconductor doped with a nondiffusing impurity. In the absence of a rate limitation at the surface, the equivalent inward diffusion has a surface concentration equal to the concentration of the impurity in the solid prior to diffusion. A rate limitation at the surface modifies this picture and the original doping level now corresponds to the equilibrium concentration N_e. We see therefore that the outward diffusion of an impurity from a previously doped crystal represents a case in which one has only to consider the boundary condition at the interface while the concentrations are determined by the initial conditions and one can effectively diffuse an "acceptor" with the diffusion coefficient of a donor.

The magnitude of the rate constant K, as defined before, is particularly important for an out-diffusion technique. The rate constant has been studied for antimony diffusion in germanium by a radio tracer technique,[46] by measurements of punch-through volt-

[44] B. Cornelison and W. A. Adcock, "Transistors by the grown-diffused technique," 1957 IRE WESCON Convention Record, pt. 3, pp. 22–27.

[45] W. C. Brower and C. E. Earhart, "70 MC Silicon Transistor," presented at the Third Annual Electron Devices Meeting, Washington, D. C.; October 31–November 1, 1957.

[46] R. C. Miller and F. M. Smits, "Diffusion of antimony out of germanium and some properties of the antimony-germanium system," *Phys. Rev.*, vol. 107, pp. 65–70; July, 1957.

ages,[47] and by a technique using the capacitance of a rectifying metal contact.[40] Essentially the same temperature dependence was found for the diffusion constant D and the rate constant K;[40,46] therefore, the quotient D/K is independent of temperature. The length D/K has a physical significance. It follows from (16) that for $z=(K/D)\sqrt{Dt}=1$ the surface concentration has decreased to approximately half of the original concentration. This condition is met for a diffusion length $\sqrt{Dt}=D/K$. Accordingly, the thinnest usable depletion layers are of the order D/K in thickness. The values reported for this length are 5×10^{-4} cm,[46] 1×10^{-4} cm,[47] and 5×10^{-5} cm[40] respectively. The first value was obtained between 800°C and 900°C, the latter values were measured at 700°C.

The discrepancies might well be within experimental error, and they might also reflect a slight temperature dependence of D/K. For practical applications, outward diffusion will have to be carried out under the very best vacuum conditions possible.

Diffusion from a Surface Phase

A great variety of diffused distributions is obtainable by the inward diffusion of impurities from the surface. For this the impurities have to be introduced from some external phase which determines the concentration of the impurities in the solid. The application of such procedures requires not only a control of the conditions at the interface, but also a control of the external phase. For most applications, starting material, doped with only one impurity, is required.

Alloy Source: Most diffusion constants were measured in a system in which the impurity was introduced from a liquid phase on the surface of the semiconductor. For this, the semiconductor specimen is sealed into a quartz tube together with the impurity to be diffused. Upon heating, the impurity forms a liquid alloy with the semiconductor. At the interface there is a transfer of impurity into the solid semiconductor. The concentration in the solid is related to the concentration in the liquid by the segregation coefficient. It is safe to assume that the rate at which the impurity crosses the interface is fast compared to the rate at which the impurity diffuses in the solid phase. This means that the concentration in the solid at the interface is constant in time and the distribution can be described by (9). Actually, during the diffusion, material is removed from the liquidus phase into the solid, thus depleting the impurity concentration in the liquidus. This should result in a solidification of liquidus at the interface. However, the segregation coefficients frequently are so small that the amount diffusing into the solid causes only a negligible depletion of the liquidus.

In such a system the surface concentration of the impurity is determined only by the thermodynamic properties of the system impurity-semiconductor and is equal to the solid solubility of the impurity at any given temperature. The disadvantage in the application of such techniques lies mainly in the fact that one is dealing with an alloy at the surface. Therefore, the location of the junction is determined by an alloy depth plus a diffused depth which might result in irregularities in the junction depth.

Such irregularities may not be significant for diode applications,[48] or if one can use the diffused distribution resulting from the simultaneous diffusion of two elements from a common alloy at the surface.[49] In such a case one of the impurities need not diffuse appreciably if the regrown region has the corresponding conductivity type.[50,51]

Reaction Phase: In a system involving more than the components semiconductor and impurity, a reaction phase, which also can be liquid, might be formed at the surface. Such a phase, for example, might be composed of an oxide originally present on the semiconductor (SiO_2 in the case of silicon) and the donor or acceptor oxide employed in the diffusion. At a fixed temperature an equilibrium concentration of donors or acceptors can be established in the reaction phase, which concentration in turn gives rise to a fixed surface concentration in the semiconductor. The rate of transfer of impurity from such a reaction phase into the semiconductor is believed to be fairly rapid. Thus the impurity distribution corresponds to the case of constant surface concentration. It should be pointed out that the surface concentration obtained from such a reaction phase cannot exceed the solid solubility which is obtained for the case of an alloy source. Frequently, the exact nature of such a reaction phase is complicated and not well understood. For example, the surface concentration obtained for aluminum diffusion into silicon in a quartz tube[25] is believed to yield a surface concentration which is limited by a reaction since the surface concentration of aluminum obtained under such conditions is approximately two orders of magnitude less than the surface concentration one obtains in a true two-component silicon-aluminum system.[52] The latter system certainly

[47] J. Halpern and R. H. Redicker, "Out diffusion as a technique for the production of diodes and transistors," this issue, p. 1068.

[48] J. S. Saby and W. C. Dunlap, Jr., "Impurity diffusion and space-charge layers in 'fused-impurity' *p-n* junctions," *Phys. Rev.*, vol. 90, pp. 630–632; May, 1953.

[49] M. Tanenbaum and D. E. Thomas "Diffused emitter and base silicon transistors," *Bell Sys. Tech. J.*, vol. 35, pp. 1–22; January, 1956.

[50] J. R. A. Beale, "Alloy-diffusion: a process for making diffused-base junction transistors," *Proc. Phys. Soc. B, London*, vol. 70, pp. 1087–1089; November, 1957.

[51] R. S. Schwartz and B. N. Slade, "A High Speed *PNP* Alloy-Diffused Drift Transistor for Switching Application," presented at the Third Annual Electron Devices Meeting, Washington, D. C.; October 31–November 1, 1957.

[52] R. C. Miller and A. Savage, "The diffusion of aluminum in single crystal silicon," *J. Appl. Phys.*, vol. 27, pp. 1430–1432; December, 1956.

requires extreme precautions with regard to cleanliness. It has not been employed for practical applications.

In silicon such a reaction phase has been obtained by a deposit from the vapor phase and also by the application of compounds on the surface.[53]

Vapor Source

A certain partial pressure of impurity vapor will be established over an impurity-semiconductor alloy. Establishing such a vapor pressure over the solid semiconductor is sufficient to produce an equilibrium concentration N_e corresponding to the solid solubility. However, if one reduces the vapor pressure to lower values, it corresponds to equilibrium concentrations below the solid solubility. In the ideal case, the equilibrium concentration is proportional to the vapor pressure. Obviously, with a vapor pressure below the one corresponding to the solid solubility, the formation of a liquid alloy is prevented. The control of the vapor pressure is a convenient means to control the surface concentrations over a wide range of concentrations. The vapor phase does not necessarily have to be composed of the vapor of the impurity element; it may be composed of a chemical compound of the impurity, such as an oxide. Particularly for the latter case, a reaction phase is frequently formed at the surface of the semiconductor. Generally, one will have to consider rate limitations at the surface for the case of a vapor source. These rate limitations might be due to a limit in mass transport in the vapor phase or they might be due to limitations in the reactions at the interface.

Vapor Pressure Control by Dilution: A convenient way to produce a controlled vapor pressure corresponding to a certain equilibrium concentration of impurity in a semiconductor is to heat semiconductor material doped with the desired equilibrium concentration together with undoped pieces of semiconductor. If the system is not completely closed, the loss of impurity vapor will give surface concentrations which are somewhat less than the concentration in the source material. Rate limitations at the surface tend to reduce the surface concentration even further. Diffused base germanium transistors have been produced by such a technique.[54]

Two Temperature Systems—Vacuum: As an alternate technique the impurity diffusion may be performed in a vacuum furnace having two temperature zones as the one shown in Fig. 8 described by Kestenbaum and Ditrick.[55] In this system the temperature of the first

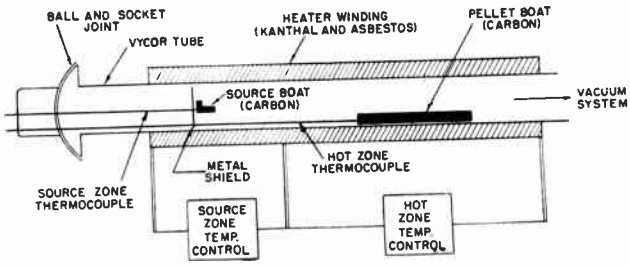


Fig. 8—Vacuum diffusion system for germanium.

zone is such that an arsenic vapor pressure of approximately 10^{-3} mm Hg is obtained from a source of pure arsenic. A metal shield in this zone prevents undesirable condensation of the arsenic vapor in the colder portion of the furnace tube. The diffusion specimen is located in a second zone which is maintained at a higher temperature which is adjusted to control the diffusion process. In such a system the vapor pressure determining the surface concentration is controllable by the pellet temperature and the geometry. In the system shown here a surface concentration of 5×10^{17} cm^{-3} was obtained for 200°C in the first zone and 785°C in the second zone.

The same basic principle of controlling the surface concentrations can be applied for silicon. However, in silicon the rate of evaporation is comparable to the rate of diffusion for group III and group V elements and one can obtain a steady-state distribution as discussed before (19).

Fig. 9 shows a diffusion system which is used for the simultaneous diffusion of phosphorus and gallium into *n*-type silicon for the fabrication of *n-p-n* structures.[56] In this system the surface concentrations are controlled by the temperature of the gallium source and the phosphorus source. The phosphorus source is heated by an external heater, while the temperature of the gallium is controlled by the position of the source in the extension tube, utilizing the temperature gradient which exists along this tube when the system is heated to the diffusion temperature.

The important parameter of the steady-state distribution is the quantity D/v which has been found to have an activation energy of only -0.95 ev.[56] This means that the steady-state layer thickness decreases approximately 40 per cent for 100°C increase in temperature. This slight temperature dependence of the diffused distribution relaxes the requirements for temperature control.

Values for D/v obtained at 1300°C for phosphorus and gallium are 1×10^{-4} cm and 1.4×10^{-4} cm respectively. The fact that, for the steady-state distribution, transients in the beginning of the process are unimportant gives such a process an inherent advantage for diffusions requiring low surface concentrations. How-

[53] C. J. Frosch, "Silicon Diffusion Technology" in "Transistor Technology," D. Van Nostrand Co., Inc., New York, N. Y., vol. 3, ch. 3B, in press.

[54] C. A. Lee, "A high-frequency diffused base germanium transistor," *Bell Sys. Tech. J.*, vol. 35, pp. 23–34; January, 1956.

[55] A. L. Kestenbaum and N. H. Ditrick, "Design, construction, and high-frequency performance of drift transistors," *RCA Rev.*, vol. 18, pp. 12–23; March, 1957.

[56] R. L. Batdorf and F. M. Smits, "The diffusion of impurities into evaporating silicon," to be published.

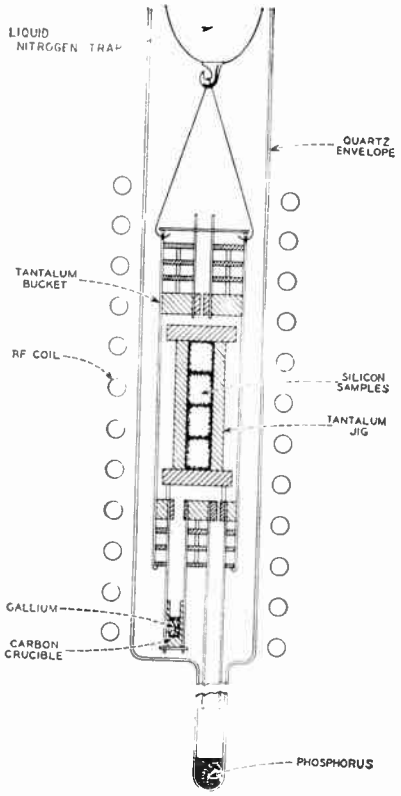


Fig. 9—Vacuum diffusion system for silicon.

ever, the layer thicknesses cannot be changed readily. Certainly, by choosing different impurity elements, the layer thicknesses in steady state will be proportional to the respective diffusion coefficients. It also is feasible to reduce the rate of evaporation by a proper geometrical arrangement. This would lead to an increase in the steady state layer thickness. As a matter of fact, increased layer thicknesses have been observed on samples whose surfaces are exposed to the direct evaporation from other pieces of silicon.

Two Temperature Systems—Gaseous Atmosphere: Diffusion systems under atmospheric pressure have been published only as applied to silicon. A typical system described by Frosch and Derick[57] is shown in Fig. 10. It consists essentially of a fused silicon tube extending through two controlled temperature zones. The first temperature zone serves to regulate the rate of evaporation of an impurity placed therein. The silicon samples are located in the second temperature zone. The positive temperature gradient between the two zones prevents the redeposition of the impurity vapor before reaching the silicon samples. The vapor from an impurity heated in the first temperature zone is carried by a gas past the silicon heated in the second zone. In such a system the source material may consist of the element of the impurity or a properly chosen chemical compound. Alternatively, one might introduce the impurity

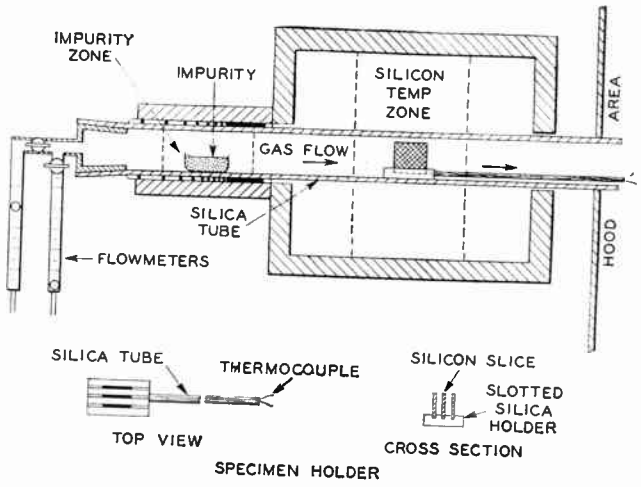


Fig. 10—Gas flow diffusion system for silicon.

into the gas flow in the form of a compound which is volatile at room temperature.[58,59]

In such a system the carrier gas is likely to react with the silicon surface which can result in an undesirable erosion of the silicon surfaces. Frosch and Derick found that the erosion is avoided by the use of an oxidizing atmosphere.[57]

The continuous nonvolatile SiO_2 layer formed on the silicon protects the underlying silicon and prevents its evaporation and complete oxidation. The oxide layer itself is easily removed by washing in HF. Oxygen, water vapor, and carbon dioxide have been reported as oxidizing agents in the carrier gas.

In addition to surface protection, an SiO_2 surface layer introduces a rate limitation at the surface for the diffusion of some donors and acceptors into silicon. Most likely, the rate limitation will not follow the simple law discussed previously for which case we showed that even with a very strong rate limitation, for the case of diffusion from a constant external phase, the impurity distribution is very close to an erfc distribution. Accordingly, it is reasonable to assume that for the same conditions an impurity distribution under a masking SiO_2 layer can be considered also as an erfc distribution (for the purpose of evaluating the surface concentration).

The rate limitation introduced by an SiO_2 layer can vary over a wide range. It depends on the impurity element and the compound in which it is used. Also, the carrier gas atmosphere and the temperature of diffusion have a significant effect. Conditions of an extreme rate limitation which lead to complete masking in the diffusion are very desirable for device applications since partial removal of an oxide layer grown prior to diffusion makes it possible to obtain intricate diffused patterns. A complete masking has been reported for

[57] C. J. Frosch and L. Derick, "Surface protection and selective masking during diffusion in silicon," *J. Electrochem. Soc.*, vol. 104, pp. 547–552; September, 1957.

[58] H. S. Veloric and K. D. Smith, "Silicon diffused junction avalanche diodes," *J. Electrochem. Soc.*, vol. 104, pp. 222–226; April, 1957.

[59] R. J. Andres and E. L. Steele, "A medium power silicon rectifier," 1957 IRE WESCON Convention Record, pt. 3, pp. 73–79.

arsenic, antimony, and boron as the diffusants. For phosphorus, a reduction in surface concentration by more than two orders of magnitudes under the oxide layer has been observed.

Prediffusion: Basically, the surface concentration can be controlled by the temperature of the source material. Also, the composition of the carrier gas can have a strong effect on the surface concentrations. This fact must be due to certain chemical equilibria in the gaseous phase. Certainly the higher surface concentrations are more readily controllable. To obtain lower surface concentrations, a two-step process might be employed.[57] In a prediffusion which is carried out at a relatively low temperature, thin layers with a high surface concentration are produced. After removing all deposits on the silicon surface by washing the samples in HF, the main diffusion is carried out in an oxidizing atmosphere at high temperatures. During the prediffusion an essentially planar source is produced in the surface of the solid. If the oxide layer grown in the main diffusion has masking properties, no impurities are lost during this step. When the final layer thickness is deep compared to the thickness obtained in the prediffusion, the impurity concentration essentially follows a Gaussian distribution. In such a process the total number of atoms in the diffused layer is controlled by the prediffusion and this number may be varied by changing the temperature and times for the prediffusion. Subsequently, any given prediffused layer can be subjected to a variety of diffusion cycles to produce a complete family of layers having characteristics suitable for both emitter and base regions of transistors. It is evident that the masking properties of the SiO_2 layer during the main diffusion are essential in such a procedure.

A special system for phosphorus diffusion into silicon by this technique has been reported by Howard,[14] who achieved a control of the sheet resistivity and the junction depth of a diffused layer to within ± 5 per cent. This corresponds to a control in the surface concentration to within ± 10 per cent. However, these control limits gradually increase, once the sheet resistivity of the diffused layer exceeds 100 Ω. On account of the high reproducibility achieved in this system, it was possible to observe the increase in diffusion coefficients with the concentration mentioned before.

The prediffused technique has an additional inherent advantage, since the removal of the thin prediffused layer requires only a very slight etch of the silicon. If such an etch is carried out with local masking of the silicon surface (*e.g.*, by photoengraving[60]) the diffused layer can be limited to defined regions. It is important to realize that the depths to be etched can be much smaller than the depths of the final diffused layer.

All of the diffusion techniques described here are subject to a variety of variations and modifications.[61] By the successive application of several diffusion steps, it is certainly possible to obtain almost any structure required for device applications. Particularly, the possibility of localizing the diffusions allows, for example, transistor structures in which the base region extends to the surface and which are therefore easy to contact.

Lifetime Effects[62]

A previous drawback for diffusion techniques has been the degradation of the carrier lifetimes as a consequence of the heat treatment required for the diffusion. The smaller base-layer thicknesses obtainable by diffusion methods make these effects less important. Also, methods have been discovered to reduce the magnitudes of these effects. In particular, it has been found that the degradation is mainly due to the introduction of rapidly diffusing impurities, for example, gold in silicon or copper in germanium. Since these impurities are mobile at relatively low temperatures where the group III and group V impurities are essentially immobile, these undesirable impurities are often removable subsequent to the main diffusion. For this purpose the sample can be heated in the presence of a liquidus layer for which the distribution coefficient for the undesirable impurity is considerably less than unity and which therefore acts as a getter for the undesirable impurity. Slow cooling also has been reported as a method to increase the minority carrier lifetime subsequent to the heat treatment. If the fast diffusing, undesirable impurity has a solid solubility which increases with temperature, slow cooling results in a precipitation of these impurities at, for example, dislocations. As a precautionary method it is advisable to always cool the silicon samples slowly from the diffusion temperature.

ACKNOWLEDGMENT

The author thanks C. S. Fuller, H. Reiss, and F. L. Frisch for helpful discussions. He is grateful to A. L. Kestenbaum who made Fig. 8 available and to C. L. Frosch who made Fig. 10 available. Receipt of preprints from J. Halpern and R. H. Redicker, and from I. A. Lesk and R. E. Gonzalez, also is acknowledged.

[60] J. Andrus and W. L. Bond, "Photoengraving in transistor fabrication," *Recent News Abstracts of Electrochem. Soc. Semiconductor Symp.*, Washington, D. C.; 1957, and in "Transistor Technology," D. Van Nostrand and Co., Inc., New York, N. Y., vol. 3, ch. 5E, in press.

[61] R. W. Aldrich and M. Waldner, "Medium power silicon transistors," *Proc. Natl. Conf. on Aeronautical Electronics*, Dayton, Ohio, pp. 353–356; 1957.

F. B. Maynard and E. L. Steele, "Very high frequency *pnp* switching transistor," *Proc. Natl. Electronics Conf.*, Chicago, Ill., pp. 246–253; 1957.

H. Nelson, "The Preparation of Semiconductor Devices by Lapping and Diffusion Techniques," presented at the Third Annual Electron Devices Meeting, Washington, D. C.; October 31–November 1, 1957.

C. Orman and L. White, "Diffused Emitter and Collector *PNP* and *NPN* Silicon Medium Power Transistors," presented at the Third Annual Electron Devices Meeting, Washington, D. C.; October 31–November 1, 1957.

H. S. Veloric and M. B. Prince, "High-voltage conductivity-modulated silicon rectifier," *Bell Sys. Tech. J.*, vol. 36, pp. 975–1004; July, 1957.

E. A. Wolff, Jr., "Diffused 50-watt silicon power transistors," 1957 IRE WESCON CONVENTION RECORD, pt. 3, pp. 40–47.

[62] For references, see G. Bemski, "Recombination in semiconductors," this issue, p. 990.

The Preparation of Semiconductor Devices by Lapping and Diffusion Techniques*

H. NELSON†

Summary—A new approach to the fabrication of semiconductor devices has been investigated. This approach allows the processing of large semiconductor wafers to a point where they can be diced into numerous and identical devices. Lapping, instead of etching, is employed for all shaping of the units, and the high degree of precision built into the lapping apparatus is passed on to all of the devices prepared. The approach is applicable to the fabrication of a great variety of semiconductor devices. Unipolar, photo-unipolar, as well as bipolar transistors and negative resistance devices, have been fabricated. These devices exhibit superior electrical characteristics. Silicon power transistors have current transfer ratios in the 20 to 40 range and power gains from 30 to 40 db. Silicon unipolar transistors have transconductances in the neighborhood of 500 μmhos and input resistance of about 100 megohms. Silicon photo-unipolar transistors show a dc photo response of 2 to 20 a per lm.

Introduction

TRANSISTOR-fabrication procedures have been described[1,2] in which impurity diffusion is employed to achieve control of doping and dimensions. These procedures, however, involve the use of troublesome etching and masking practices to delimit the diffused regions. A different approach has been investigated that seems to be more amenable to mass production. This approach, which involves lapping operations, allows the precision processing of large semiconductor wafers to the point where they can be diced into numerous finished semiconductor units. The approach is simple and precise, and is applicable to the fabrication of a great variety of semiconductor devices. Unipolar and photo-unipolar, as well as bipolar transistors and negative resistance devices, have been fabricated. This paper describes these devices as well as the details of the new fabrication procedure.

Control of Transistor Geometry by Lapping

The use of the lapping process to confine a desired impurity doping to a specific region of a device structure can be illustrated by a description of its application in the preparation of silicon-power transistors. Large-area silicon wafers are used as starting material. As a first step in their processing, these wafers are provided with emitter grooves in a lapping apparatus of the type shown in the photograph of Fig. 1. The wafers are attached to the lap base by means of a thin film of wax. The runners at the sides of the lap base provide for automatic stoppage of the lapping when the desired groove depth is reached. After grooving, the wafers are exposed to a phosphorus diffusion process which leads to the generation of a phosphorus-containing glass film on the surface of the wafer and to the establishment of a *p-n* junction below this surface. This *p-n* junction and the doped regions are shown in the cross-section sketch of Fig. 2(a). The diffused regions in the lands of the grooved surface are removed by a second lapping step. The silicon wafers are then subjected to a diffusion step wherein the lands of the grooved surface are doped with boron and the *n* regions with additional phosphorus.

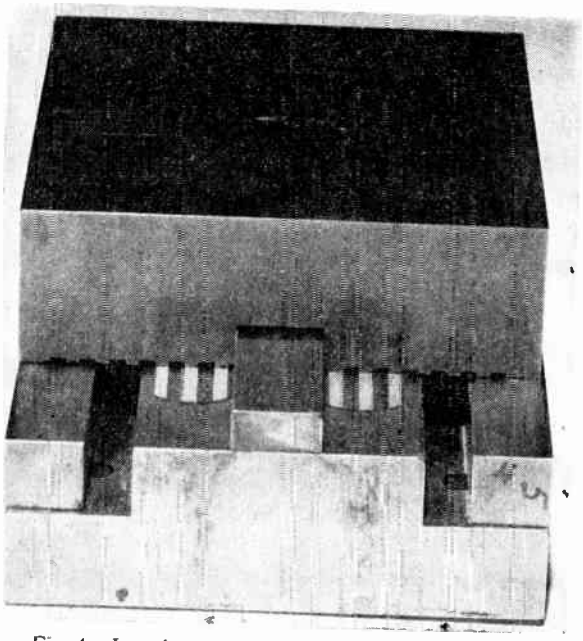


Fig. 1—Lapping apparatus for the preparation of semiconductor devices.

After this final diffusion step, the wafers are plated with nickel and subjected to further lapping to isolate emitter from base regions. For this process, a lap head is used which is an exact duplicate of the one used for cutting the original grooves, except that the groove-cutting teeth are a few mils wider. A set of runners is used to provide a lap depth just short of the original groove bottom. Fig. 2(b) shows the cross section of the wafers after the final lapping. Fig. 2(c) shows one of the transistors diced from this wafer.

An obvious advantage of the above technique is that it lends itself to the production of large lots of identical transistors. The same high degree of precision, originally built into the lapping apparatus, is passed on to all of the transistors prepared. Base-width variations from unit to unit, as well as from point to point over large emitter

* Original manuscript received by the IRE, February 26, 1958.
† RCA Labs., Princeton, N. J.
[1] M. Tanenbaum and D. E. Thomas, "Diffused emitter and base silicon transistors," *Bell Sys. Tech. J.*, vol. 35, pp. 1–22; January, 1956.
[2] C. A. Lee, "A high-frequency diffused base germanium transistor," *Bell Sys. Tech. J.*, vol. 35, pp. 23–24; January, 1956.

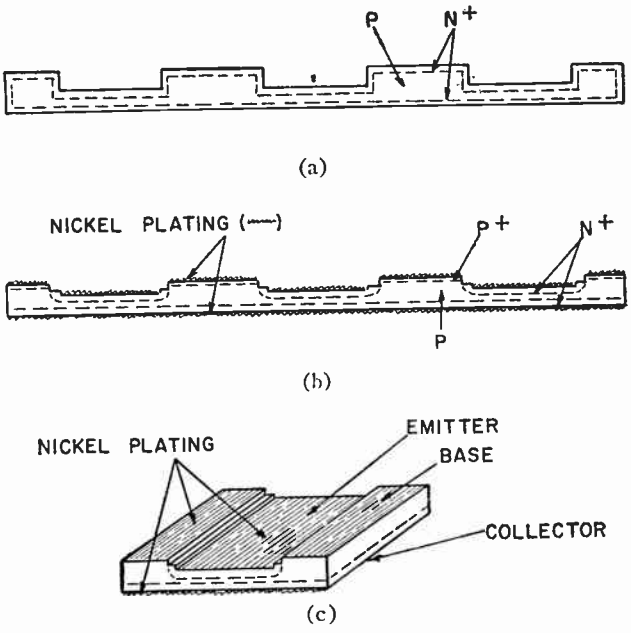


Fig. 2—The preparation of diffused emitter and collector transistors. (a) Wafer cross section after first diffusion. (b) Wafer cross section after final lapping. (c) Transistor diced from processed wafer.

areas, may accordingly be effectively minimized. The lapping process, also, leads to a separation of the emitter from the base that is uniform from unit to unit as well as from point to point along the emitter periphery.

The lapping-diffusion technique may be applied advantageously, not only to the preparation of bipolar transistors, but to the manufacture of semiconductor devices in general. Essentially, it is a technique that provides precisely controlled confinement of a desired impurity doping to a specific region of a semiconductor structure. The technique appears to be particularly attractive for the manufacture of semiconductor devices in which achievement of this objective is essential.

Transistors Prepared by Diffusion and Lapping

Diffused Emitter and Collector Transistors

Several lots of this type of transistor have been prepared by the lapping-diffusion cycle described in the preceding section. *P*-type silicon in the 2 to 6 ohm-cm range has been used as starting material for *n-p-n* power transistors. Large-area wafers, of dimensions $1 \times \frac{1}{2}$ inch, of this material are lapped to a thickness of 10 mils and are provided with emitter grooves 80 mils wide and 6 mils deep. After grooving, and after removal of wax and loose particles, the wafers are cleaned by a two-minute immersion in boiling, concentrated nitric acid and by a thorough rinsing in deionized water. The cleaning procedure is followed by the application of a "phosphate glass" film to the surface of the wafers by a phosphorus pentoxide treatment[3] at 1200°C. Diffusion of phosphorus is then carried out at 1300°C for four hours. The diffused regions in the lands of the grooved surface are then removed in a second step of lapping. For this step, a set of runners is used which provides for automatic stoppage of the lapping at an 8-mil thickness of the wafers. This lapping is followed by cleaning of the wafers and by a second diffusion step in which the wafers are exposed to boron-trichloride treatment[3] at 1200°C and to diffusion at 1300°C for one hour. After the diffusion, the wafers are cleaned by rinsing in hydrofluoric acid, boiling in concentrated nitric acid, and by rinsing in deionized water. They are then plated by an "electroless nickel" process[4] and subjected to further lapping to isolate emitter from base regions. For this lapping step, a lap-head is used which is an exact duplicate of the one used for the original cutting of the emitter grooves, except that the groove cutting teeth are 86 mils instead of 80 mils wide. A set of runners is used to provide for a lap depth of 3 mils and for the final geometry of the wafers which is shown in Fig. 2(b). The wafers are then diced into transistor units, one of which is shown in Fig. 2(c). The mounting of the units is facilitated by the presence of the nickel coating on the electrode surfaces, since leads and cooling fins can readily be soldered onto this coating. After mounting, the elimination of interelectrode leakage is readily accomplished by a light etch in CP4.

Transistors prepared in the above manner have been subjected to cross-section studies and to such electrical tests as would serve to evaluate the degree of success attained in the precise control of amount and location of doping impurities. The dimensions of the emitter, base, and collector regions can be determined from the cross sections. The net acceptor concentration in the starting material is known and also, therefore, the phosphorus concentration at the *p-n* junction planes. Since the diffusion coefficient of phosphorus[5] and the time and temperature of the diffusion are known, the phosphorus concentration can be calculated for any parallel plane between the *p-n* junctions and the surface. Penetration of diffused boron in the base-lead region is calculable on the basis of known values of the diffusion coefficient of boron and its concentration at the surface of the wafer. This concentration is assumed equal to that obtained at the surface during the diffusion process and approximately equal also to the value of 1×10^{21} atoms per cm^3 obtained by Fuller[5] for the processing conditions employed in diffusing the cross-sectioned specimen.

Emitter efficiency and base resistance of the transistors have been estimated on the basis of the calculations described above. For the lot of transistors, of which the cross-sectioned unit of Fig. 3 is representative, the

[3] C. J. Frosch, "Surface protection and selective masking during diffusion in silicon," *J. Electrochem. Soc.*, vol. 104, pp. 547–552; September, 1957.

[4] M. V. Sullivan and J. H. Eigler, "Electroless nickel plating for making contacts to silicon," *J. Electrochem. Soc.*, vol. 104, pp. 226–229; April, 1957.

[5] C. S. Fuller and J. A. Ditzenberger, "Diffusion of donor and acceptor elements in silicon," *J. Appl. Phys.*, vol. 27, pp. 544–553; May, 1956.

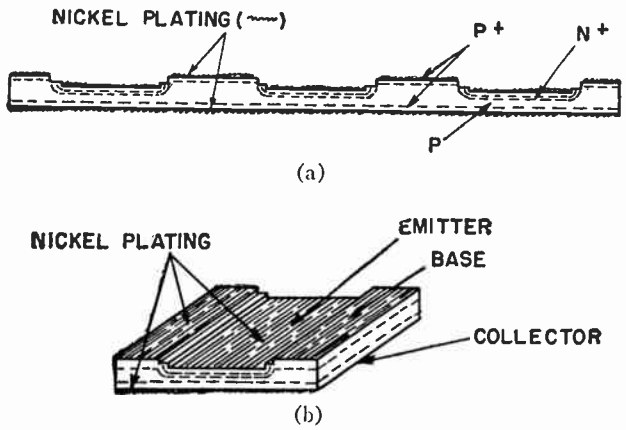


Fig. 3—The preparation of diffused emitter and base transistors. (a) Wafer cross section after final lapping. (b) Transistor diced from processed wafer.

emitter efficiency was estimated to be roughly comparable to that of alloyed emitters. The average density of donors in the emitter region within 0.5 mil of the p-n junction is about 10^{18} per cm^3 while the density of acceptors in the base region is in the neighborhood of 10^{15} per cm^3. It is assumed, therefore, that the current transfer ratio of the transistors is primarily determined by the base width and effective lifetime of the injected carriers. Table I shows measured values of these parameters for five samples of the transistor lot in question. The relatively small spread in current transfer ratio, α_{cb}, and effective lifetime,[6] τ_e, is indicative of the uniformity of product achieved. The emitter area of these units is 0.04 cm^2 and the base width 0.75 mil (from cross-section measurements). The α_{cb} of these units is limited by the high recombination rate of injected carriers, rather than by low-injection efficiency. When the surface-recombination rate is lowered by a sodium-dichromate treatment,[7] higher values of α_{cb} and τ_e are obtained. Values of 28 and 1.5 μsec, for instance, were obtained for unit No. 5 after sodium dichromate treatment.

The data in Table II show results obtained for p-n-p silicon transistors prepared by the same diffusion-lapping cycle employed for the fabrication of the n-p-n units. The p-n-p transistors differ from the n-p-n ones in that 1.5 to 3 ohm-cm n-type silicon was used as starting material and boron, instead of phosphorus, was used for diffusing the emitter and the collector regions. The base width is approximately 0.6 mil.

Diffused Emitter and Base Transistors

Transistors in which the emitter and base, instead of emitter and collector, are diffused have also been made. These have been prepared through a cycle of lapping, cleaning, and diffusion processes similar to those described above. For the preparation of n-p-n power transistors, 9-mil thick wafers of 2 to 8 ohm-cm n-type silicon are provided with 40-mil-wide and 2-mil-deep grooves. These wafers are first subjected to boron diffusion for the generation of a p-n junction 2 mils below the surface. They are then relapped to remove a 1.5-mil-thick layer from the ungrooved surface and a 0.5-mil-thick layer from the bottom of the grooves. The relapping is followed by an application of "phosphate glass" and further lapping to remove this "glass" from the lands of the grooved surface of the wafers. The wafers are then introduced into the diffusion furnace, exposed to boron tricholoride gas at 1200°C, and afterwards brought to a temperature of 1300°C. Diffusion is then carried out at 1300°C for about 4 hours. After removal from the diffusion furnace, the wafers are cleaned and plated with "electroless nickel." Then follows a final lapping step in which the emitter regions are isolated from the base regions by removal of nickel and degenerate silicon from the groove edges. A lap head with 44-mil, groove-cutting teeth is used for this operation along with a set of runners which provides for a 0.7-mil groove penetration. Fig. 3(a) shows the cross section of one of the wafers after this final lapping and Fig. 3(b) shows one of the transistor units obtained after the dicing of the wafer.

Transistors prepared in this manner have been subjected to cross-section studies wherein the dimensions of emitter and base regions have been determined. The concentration of diffused impurities in these regions can be roughly calculated on the basis of diffusion time and

TABLE I

ELECTRICAL CHARACTERISTICS OF DIFFUSED EMITTER AND COLLECTOR n-p-n SILICON TRANSISTORS

Transistor	4-kc Current Transfer Ratio (α_{cb})	4-kc Power Gain db	Effective Lifetime (τ_e) μsec
No. 1	11.5	34	0.85
No. 2	13.0	33.5	0.95
No. 3	11.0	33.5	0.9
No. 4	12.5	34.0	0.95
No. 5	12.0	35.0	1.0

All values of gain measured at $V_c=6$ v, $I_e=50$ ma. Power gain measured with resistive input and conjugate-matched output.

TABLE II

ELECTRICAL CHARACTERISTICS OF DIFFUSED EMITTER AND COLLECTOR p-n-p SILICON TRANSISTORS

Transistor	4-kc Current Transfer Ratio (α_{cb})	4-kc Power Gain db	Effective Lifetime (τ_e) μsec
No. 1	35	36	2.1
No. 2	28	34	1.2
No. 3	16	29	1.1
No. 4	27	32	1.1

All values of gain measured at $V_0=6$v, $I_e=50$ ma. Power gain measured with resistive input and conjugate-matched output.

[6] L. R. Lederhandler and L. J. Giacoletto, "Measurement of minor ty carrier lifetime and surface effects in junction devices," PROC. IRE, vol. 34, pp. 477–483; April, 1955.

[7] A. R. Moore and H. Nelson, "Surface treatment of silicon for low recombination velocity," *RCA Rev.*, vol. 17, pp. 5–12; March, 1956.

TABLE III

ELECTRICAL CHARACTERISTICS OF DIFFUSED EMITTER AND BASE *n-p-n* SILICON TRANSISTORS

Transistor	1 α_{cb}	2 P. G. db	3 α_{cb}*	4 P. G. db*	5 BV_{CBO}	6 BV_{EBO}
No. 1	11.5	39.5	18.5	38	12 v	90 v
No. 2	10.0	39.0	17.0	40	13.5	140
No. 3	12.5	40.0	18.0	41	16.0	110
No. 4	10.5	37.0	15.5	39	17.0	120

* After sodium dichromate surface treatment.
All values of gain measured at $V_c = 6$ v, $I_e = 50$ ma. Power gain measured with resistive input and conjugate-matched output.

temperature data, and the knowledge of diffusion coefficients and surface concentrations referred to earlier. For one lot of transistors the average net concentration of donors in the emitter region within 0.5 mil of the emitter junctions was calculated to be about 10^{18} per cm^3. The average acceptor concentration in the base region, on the other hand, was calculated to be about 5×10^{16} per cm^3. It may be expected, therefore, that the current-transfer ratio of these transistors will be limited by recombination loss of injected carriers rather than by low-injection efficiency. This is borne out by the data shown in Table III. The current transfer ratio, α_{cb}, and the power gain obtained for four representative samples of these transistors after a light CP4 etch are shown in columns 1 and 2, while the values of the same parameters after sodium dichromate treatment are shown in columns 3 and 4. Columns 5 and 6 show emitter and collector breakdown voltages before the dichromate treatment.

The effective lifetime of minority carriers injected into the base region could not be measured with available equipment (<0.1 μsec). Values as high as 1.0 μsec were obtained, however, for holes injected from the base into the collector region.

Diffused Transistor with Negative Collector Resistance

Transistors with negative collector resistance may be prepared by a diffusion-lapping cycle similar to that employed for the preparation of diffused emitter and collector transistors. The procedure used for the fabrication of the latter type of transistors is modified to give the final cross section of the processed wafer shown in Fig. 4(a), or a final shape of the diced transistor unit as shown in Fig. 4(b). This transistor structure is different in that the region of the base, B_1, to which the base lead is attached, is separated from the regions of the base, B_2, between the emitter-collector junctions, by a third region, B_3, which is so thin that its resistance is substantially affected by the penetration of the depletion layer associated with the collector junction. When the degree of depletion-layer modulation is suitably related to other pertinent electrical characteristics of the transistor, a negative collector resistance is obtained. An increase in the collector bias causes a sufficient decrease in the conductance of the region B_3 to

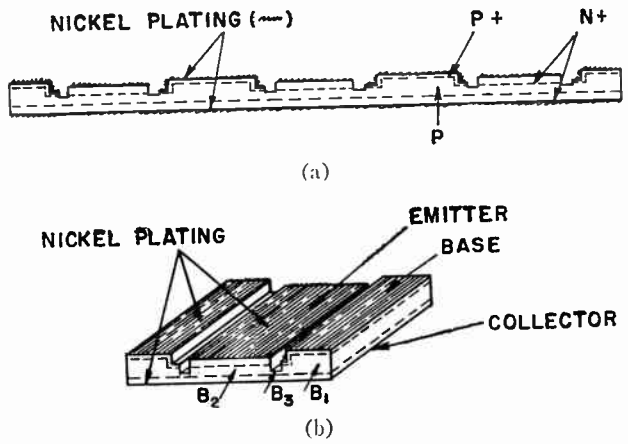


Fig. 4—The preparation of transistors with negative collector resistance. (a) Wafer cross section after final lapping. (b) Transistor diced from processed wafer.

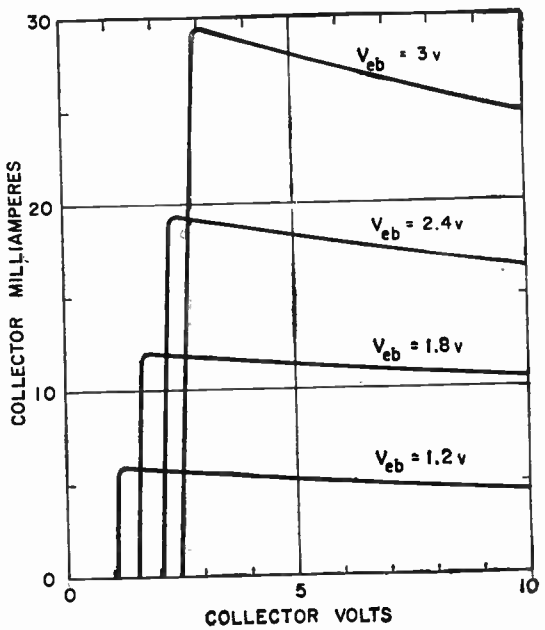


Fig. 5—Output characteristics of silicon transistor with negative collector resistance—load resistance = 1000 ohms.

cause lowered forward bias and lowered injection at the emitter junction.

Measurements taken on several cross sections of this type of transistor show that the thickness of the B_3 section of the base can be precisely controlled by the diffusion-lapping technique. One lot of devices prepared showed a channel thickness range before etching from 0.7 to 1.0 mil. These particular transistors, which were prepared from 8 ohm-cm *p*-type silicon, exhibited collector current-collector voltage characteristics as shown by the family of curves in Fig. 5.

Diffused Unipolar and Photo-Unipolar Transistors

To fabricate unipolar transistors,[8] a cycle of diffusion and lapping is employed which leads to a final cross

[8] W. Shockley, "A unipolar field-effect transistor," PROC. IRE, vol. 40, pp. 1365–1376; November, 1952.

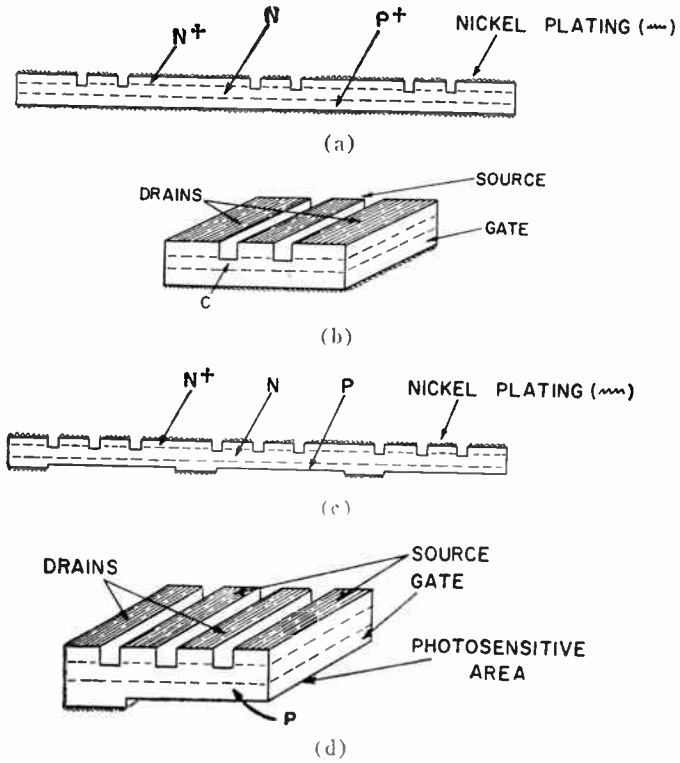


Fig. 6—The preparation of unipolar devices. (a) Si wafer processed for unipolar transistor preparation. (b) Unipolar transistor diced from processed wafer. (c) Si wafer processed for photo-unipolar transistor preparation. (d) Photo-unipolar transistor.

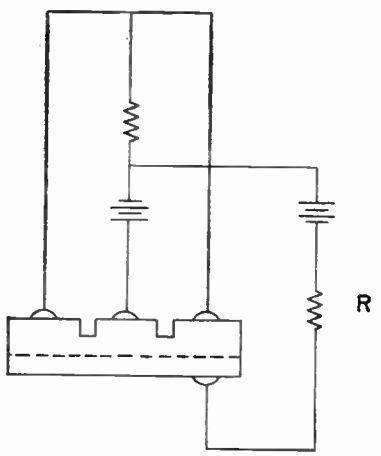


Fig. 7—Unipolar transistor operated as combined photocell and amplifier.

section of the wafer as shown in Fig. 6(a). These wafers are diced into transistor units as shown in Fig. 6(b). The thickness of the channel, C, is susceptible to precise control by etching after mounting the transistor.

When a unipolar transistor of the type shown in Fig. 6(b) is operated in a circuit as shown in Fig. 7, it will respond to light impinging upon the channel surfaces. Drain-current flow will increase as a result of increased channel conductivity because of hole-electron pair generation and also, to a greater extent, as a result of a lowered gate bias caused by an increased current flow through the resistor R. Unipolar devices especially designed for this type of operation have been prepared. These have been termed "photo-unipolar" transistors. To fabricate these devices a lapping-diffusion cycle is employed which produces a final wafer cross section as shown in Fig. 6(c). The wafers are diced into photo-unipolar units as shown in Fig. 6(d). These units are provided with a large photosensitive area which serves as a source of photogenerated hole-electron pairs.

Dimensional measurements from a number of cross sections of unipolar transistors have shown a high degree of channel uniformity from unit to unit. This uniformity is also evidenced by the electrical properties of the units. When the transistors have been prepared from 21 ohm-cm, n-type silicon, they have shown transconductances, g_m, in the neighborhood of 500 μmhos. At a gate bias of -4 v and a drain bias of 10 v, one representative unit shows a g_m of 460 μmhos, a drain resistance of 3200 ohms, and a gate-to-channel capacity of 96 $\mu\mu$f. This unit was prepared from 21 to 30 ohm-cm n-type silicon and had been provided with channels 80 mils long, 4 to 6 mils wide, and 0.5 to 0.7 mil thick.

The unipolar silicon transistors show improved electrical characteristics after sodium dichromate treatment. The curves of Fig. 8 show the I_d as a function of E_g before and after this treatment for a typical unit. The increase in channel resistance is in conformity with the view that application of sodium dichromate film leads to the extraction of electrons from the underlying silicon.[7] The detailed reasons for the increase in g_m are not known.

The curves of Fig. 9 show I_d as a function of V_g for another unipolar transistor at room and liquid nitrogen temperature. The greatly increased g_m observed at the liquid nitrogen temperature is a consequence of the increased mobility of the electrons at the lower temperature. As expected, the reverse gate current was found to be extremely low at liquid nitrogen temperature. In fact it was lower than 10^{-12} a, which was the limit of the sensitivity of the instrument used for the measurements.

Photo-unipolar transistors have been prepared from 21 to 30 ohm-cm n-type silicon and they have been provided with a p-n junction through boron diffusion. The three channels are each 180 mils long and they are connected in parallel. The channels are approximately 5 mils wide and 0.6 mil thick. The section P of the boron-doped region, which serves as the photosensitive element of the device, is approximately 0.070×0.180 inch in area and 0.001 inch thick.

Measurements on a representative unit showed a gate-to-channel capacity of 140 $\mu\mu$f, a g_m of 500 μmhos, and an I_d of 7×10^{-3} a, at a gate bias of -4 v and a drain bias of 10 v. An amplified photo-response of 3 a per lm was obtained when a 2-megohm resistor was connected in series with the gate bias.

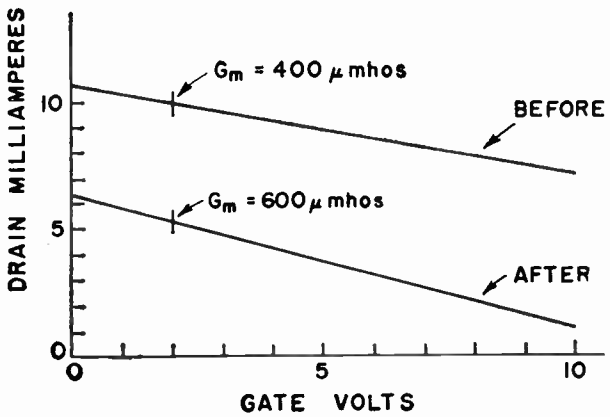


Fig. 8—Drain current as a function of gate bias before and after application of sodium dichromate—$V_d = 10$ v.

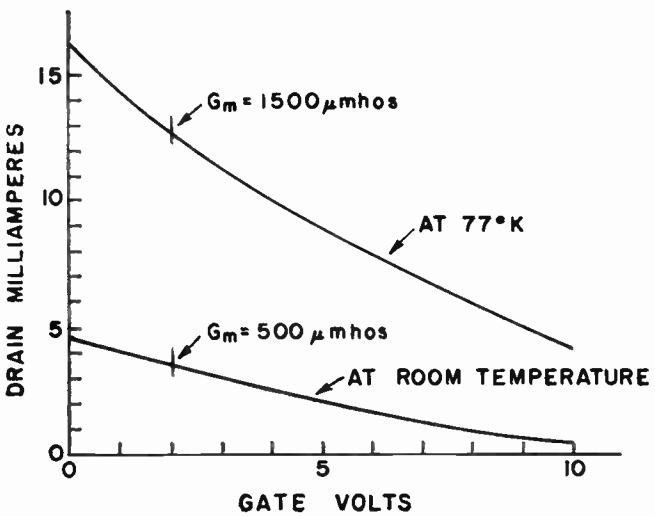


Fig. 9—Drain current as a function of gate bias at room temperature and at 77°K.

Discussion of Experimental Results

The results obtained in the course of this investigation indicate that a diffusion-lapping technique may be applied advantageously to the preparation of a great variety of silicon devices. Some of the advantages expected from the application of the impurity-diffusion process have been realized. Large-area, diffusion-doped regions of uniform thickness have been obtained with relative ease. When phosphorus and boron have been used as diffusants, emitter-impurity concentrations have been obtained which are compatible with high-injection efficiency in bipolar transistors. Although the lifetime of the silicon deteriorates as a result of its exposure to the high-diffusion temperatures, this deterioration does not preclude the attainment of "high" transfer ratios at thin base widths (<0.8 mil). As in the case of silicon transistors prepared by the alloy process, high surface, rather than high body, recombination appears as the obstacle to be overcome for the attainment of high current multiplication. Steps taken to minimize emitter and collector junctions separation at the emitter periphery of diffused emitter and collector transistors have, accordingly, been found effective in reducing loss of minority carriers in the base.

As indicated by the experimental results, "high" transfer ratios may be obtained with diffused emitter and base, as well as with diffused emitter and collector transistors. The fact that very thin and uniform base widths may be obtained readily in the former transistors favors the attainment of high-current gains in these although the effective lifetime in the base is low (presumably because of high impurity concentration in the base near the emitter junction). With regard to other electrical characteristics, it follows from a consideration of the nature of the processing, that the steps involved in the preparation of diffused emitter and base transistors favor the attainment of high-frequency response and high-collector breakdown. The steps involved in the preparation of diffused emitter and collector transistors on the other hand are more compatible with the attainment of high-emitter breakdown and low-collector lead resistance.

The ready preparation of unipolar and photo-unipolar devices with useful electrical characteristics is a good demonstration of the versatility of the diffusion-lapping technique. The above experimental data are included to illustrate this fact, rather than to show what can be ultimately achieved.

It is worth noting that unipolar, unlike bipolar silicon transistors, show a general improvement in electrical characteristics with decreasing temperature (an optimum is reached at about 100°K—where the mobility peaks). The virtual absence of reverse gate current at low temperatures suggests applications for silicon-unipolar devices where the amplification of extremely minute currents or light signals is desired.

Relatively few transistors with negative collector resistance have been prepared. The experimental data presented, therefore, represent only preliminary attempts in this direction. The results obtained do, however, indicate that devices possessing predeterminable, stable, negative resistance at low applied voltages can be prepared by the diffusion-lapping technique.

Conclusions

By means of impurity diffusion and lapping procedures it has been possible to prepare a variety of silicon devices in which the shape and the location of doped regions are subject to precise control. In bipolar transistors suitable for power applications, thin and uniform base widths compatible with high-current transfer ratios have been obtained. In unipolar transistors, channel geometries compatible with high power gain have also been attained.

Two novel semiconductor devices have also been prepared: a photo-unipolar transistor with a dc output of 2 to 20 a per lm and a silicon transistor with a predeterminable, stable, and negative resistance.

Outdiffusion as a Technique for the Production of Diodes and Transistors*

J. HALPERN†, MEMBER, IRE, AND R. H. REDIKER†, ASSOCIATE MEMBER, IRE

Summary—**The outdiffusion process, which consists of the extraction of impurities from a semiconductor wafer by heating it to an elevated temperature in a high vacuum, is shown to be a practical method for the production of diodes and transistors. The usefulness of outdiffusion as a technique for device fabrication depends on how easily impurities can be evaporated from the crystal surface. The surface-evaporation velocity K which characterizes the ease of removal of impurities has been determined for the evaporation of antimony out of germanium at 700°C and is $(1.5 \pm 0.5) \times 10^{-3}$ cm/sec. This value is large enough to indicate that it is definitely feasible to make high-frequency devices by outdiffusion. Narrow base germanium computer diodes have been fabricated that have a forward drop of 0.11 volt at 1 ma and that switch at speeds up to 5 mc. The operation of these graded base diodes is analyzed. Germanium n-p-n graded base transistors have also been fabricated which have grounded-emitter current gains, β, of over 100 and alpha frequency cutoffs, $f_{\alpha_{eq}}$, of above 200 mc.**

I. Introduction

THE diffusion of impurities out of semiconductors has been described in the literature.[1–4] The process consists of the extraction of impurities from a crystal by causing them to diffuse to the surface from where they are evaporated into a vacuum. In this paper we will describe the application of this technique to the production of p-n junctions and consequently to the fabrication of diodes and transistors.

The usefulness of outdiffusion as a practical technique for the fabrication of devices depends on how easily the impurities can be removed from the crystal. The surface-evaporation velocity K is a measure of the ease with which the impurities can be removed. This quantity was determined from punch-through voltage measurements on the p-n junctions which were produced by the outdiffusion process.

The value of the surface-evaporation velocity which we have obtained is large enough to permit the fabrication of diodes and transistors from compensated[5] n-type germanium using practicable heating times and temperatures. Since it has been possible to grow germanium with predetermined concentrations of both p- and n-type impurities, the precise impurity profile after outdiffusion is known once K is known, and hence it has been possible to design diodes and transistors to desired electrical specifications. Narrow base diodes with high-speed capability and with very low forward drop have been fabricated by this technique. These devices have graded base regions. This gives an extra degree of freedom in their fabrication and permits a higher optimization for the combination of junction capacitance, avalanche breakdown, and punch-through voltage than for a uniformly doped base region. N-p-n graded base transistors have also been fabricated using the outdiffusion process and have shown, in addition to all the advantages of the graded base structure, extremely high grounded-emitter current gains.

II. Semiconductor Junctions by Outdiffusion

Junctions can be produced by outdiffusion if the semiconductor material with which one starts contains both donor and acceptor impurities (*i.e.*, it is compensated) and the impurity having the larger diffusion constant is present in the greater concentration. In this case the boundary conditions at the semiconductor surface must be such as to permit sufficient evaporation of the faster diffusing impurity so that a surface layer of conductivity different from that of the bulk can be produced. It is also important that the evaporation rate of the semiconductor material itself be smaller than that of the impurities. For germanium at most practical temperatures this last condition is always fulfilled. If all these conditions are satisfied it is then possible to produce p-n junctions by heating the material to an elevated temperature in high vacuum.

The rate at which an impurity evaporates from the surface can be assumed to be proportional to the deviation of its surface concentration from equilibrium. The boundary condition at the surface is then

$$K[C(0, t) - C_{eq}] = D \left.\frac{dC}{dx}\right|_{x=0}, \qquad (1)$$

where the left-hand side of (1) is associated with evaporation of impurities from the surface and the right-hand side with diffusion of impurities from the bulk to the surface. The quantity D is the diffusion constant, $C(x, t)$ is the impurity concentration at a depth x from the surface at a time t, and C_{eq} is the equilibrium concentration at the surface. All the outdiffusion runs were performed in vacua of better than 10^{-6} mm of Hg in

* Original manuscript received by the IRE, February 26, 1958; revised manuscript received, March 24, 1958. The research reported in this document was supported jointly by the Army, Navy, and Air Force, under contract with the Mass. Inst. Tech. Preliminary reports of this work were presented at the IRE-PGED Meeting, Washington, D. C., October 25–26, 1956, and at the IRE-AIEE Semiconductor Devices Res. Conf., Univ. of Colorado, Boulder, Colo., July, 1957.

† Lincoln Laboratory, M.I.T., Lexington, Mass.

[1] B. Serin, "Heat treatment of semiconductors and contact rectification," *Phys. Rev.*, vol. 69, pp. 357–362; April 1 and 15, 1946.

[2] F. M. Smits and R. C. Miller, "Rate limitation at the surface for impurity diffusion in semiconductors," *Phys. Rev.*, vol. 104, pp. 1242–1245; December, 1956.

[3] K. Lehovic, K. Schoeni, and R. Zuleeg, "Evaporation of impurities from semiconductors," *J. Appl. Phys.*, vol. 28, pp. 420–423; April, 1957.

[4] R. C. Miller and F. M. Smits, "Diffusion of antimony out of germanium and some properties of the antimony-germanium system," *Phys. Rev.*, vol. 107, pp. 65–70; July, 1957.

[5] Compensated germanium is material containing both donor and acceptor impurity elements, the type being that of the impurity element of greater concentration.

which case C_{eq} can be assumed to be zero. The quantity K in (1), which is called the surface-evaporation velocity, is a measure of the ease with which an impurity atom can be removed from the surface of the semiconductor and a knowledge of it is necessary for the controlled utilization of the outdiffusion process.

If a compensated semiconductor is initially uniformly doped with N_{ao} acceptors/cm^3 and N_{do} donors/cm^3 the net acceptor concentration as a function of distance at a time t can be determined from the solution of the diffusion equation subject to the boundary condition given by (1) and is:[6]

$$N_a - N_d = N_{ao}\left\{\operatorname{erf}\left[\frac{x}{2\sqrt{D_a t}}\right] + \exp\left[\frac{K_a x + K_a^2 t}{D_a}\right]\operatorname{erfc}\left[\left(\frac{x}{2\sqrt{D_a t}}\right) + \left(\frac{K_a}{D_a}\sqrt{D_a t}\right)\right]\right\} - N_{do}\left\{\operatorname{erf}\left[\frac{x}{2\sqrt{D_d t}}\right] + \exp\left[\frac{K_d x + K_d^2 t}{D_d}\right]\operatorname{erfc}\left[\left(\frac{x}{2\sqrt{D_d t}}\right) + \left(\frac{K_d}{D_d}\sqrt{D_d t}\right)\right]\right\}, \tag{2}$$

where the subscripts a and d refer to acceptors and donors, respectively.[7]

If the surface-evaporation velocities and diffusion constants are known, the precise impurity profile can be determined from (2). This is unlike the indiffusion case where both the concentration of the impurity atoms in the gaseous phase and the partition function must be known. An impurity profile calculated from (2) is shown in Fig. 1 for an outdiffusion run which was performed at 700°C for 24 hours using a germanium wafer initially doped with 7.4×10^{15} atoms/cm^3 of indium and 10.4×10^{15} atoms/cm^3 of antimony.[8] Fig. 1 has been plotted using the appropriate values for the diffusion constants for antimony and indium at 700°C (1.6×10^{-12} and 9×10^{-15} cm^2/sec, respectively) and different values of surface-evaporation velocity. The lower branch of the curve for $K_{Sb}=10^{-8}$ in the figure has been plotted assuming a value of $K_{\text{Indium}}=\infty$. The other curves have been plotted assuming $K_{\text{Indium}}=0$ (*i.e.*, N_a remains constant throughout the material after outdiffusion). In our work the value of K_{Indium} is immaterial since that region where the impurity profile is affected is relatively small because of the small diffusion constant for indium. Furthermore, in the fabrication of transistors or diodes this region is either indiffused or removed (see Section IV-C and Section III-E). By relating experimental knowledge regarding the impurity profile to the theoretical curves in Fig. 1 we have been able to obtain a value for the pertinent surface evaporation velocity, K, for antimony.

The experimental quantity which has been used in the determination of the surface-evaporation velocity is the reverse junction voltage at which the space-charge region of the junction penetrates through the p-type outdiffused skin. This reverse voltage is the well-known punch-through voltage. The sample whose theoretical impurity densities are plotted in Fig. 1 was then indiffused at 600°C for 4.5 hours from a surface concentration of 5×10^{18} atoms/cc of arsenic. The indiffused n-p junction was found using sectioning and etching techniques to be at a depth of 1.6 microns from the surface.[9] The theoretical impurity profile for the sample after the indiffusion is shown in Fig. 2. There are two advantages to be gained by this additional indiffusion process. Firstly, the effect of the unknown surface-evaporation velocity for indium on the impurity profile has been considerably reduced and therefore the evaporation velocity for antimony can be more accurately determined. Secondly, the punch-through is sharper to the heavily doped n-type indiffused layer than it is to inversion layers or low lifetime regions of unknown depth below the surface. If one makes the valid assumption of complete depletion in the space-charge region of the outdiffused junction when it is reverse biased, the rate of change of electric field with distance in this region is obtained from Poisson's equation:

$$\frac{dE}{dx} = \frac{q}{\epsilon}(N_a - N_d), \tag{3}$$

where q is the electronic charge and ϵ is the dielectric constant in rationalized mks units. This quantity, which we have called *dE/dx punch-through* is shown as ordinate in Fig. 2. By integrating the curves of Fig. 2 (for each of the parametric values of K) from the indiffused junction through the outdiffused junction to the point of total charge neutrality, the field E in the space-charge region at punch-through is obtained as is shown in Fig. 3. The area under each of the curves is the punch-through voltage corresponding to the particular value of K. Performing these graphical integrations yields the result that if K were 10^{-8} cm/sec the punch-through voltage would be 9.4 volts while if K were 2×10^{-8} cm/sec the

[6] The problem is identical in form to that in heat flow: constant initial temperature and radiation at the surface into a medium at zero temperature. See H. G. Carslaw and J. C. Jaeger, "Conduction of Heat in Solids," Oxford Univ. Press, London, Eng., p. 54; 1947.

[7] $$\operatorname{erf} y \equiv \frac{2}{\sqrt{\pi}}\int_0^y \exp(-\xi^2)d\xi \equiv 1 - \operatorname{erfc} y.$$

[8] There should not be more than 5 per cent inaccuracy in these impurity concentrations. The underdoping was determined by growing three p-type indium doped uncompensated crystals under the same conditions that the compensated one was grown. It is assumed that when two types of dopant are put into the melt, no interaction occurs to change the underdoping profile as measured on the control crystals. This seems reasonable when one considers the small amount of impurities put into the melt.

[9] Because the indiffused junction is relatively abrupt its depth from the surface can be accurately measured by sectioning and etching techniques. The outdiffused junction, on the other hand, is very graded.

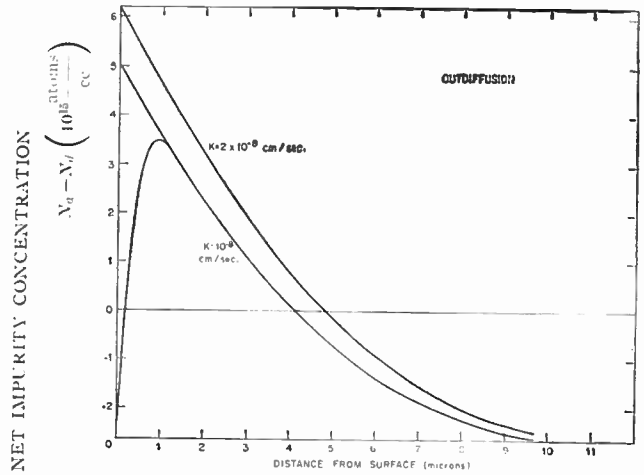


Fig. 1—Impurity profiles calculated for an outdiffusion run of 24 hours at 700°C for two assumed values of surface-evaporation velocity of antimony. The lower branch of the curve for $K_{Sb}=10^{-8}$ has been plotted assuming a value of $K_{Indium}=\infty$. The other curves have been plotted assuming $K_{Indium}=0$.

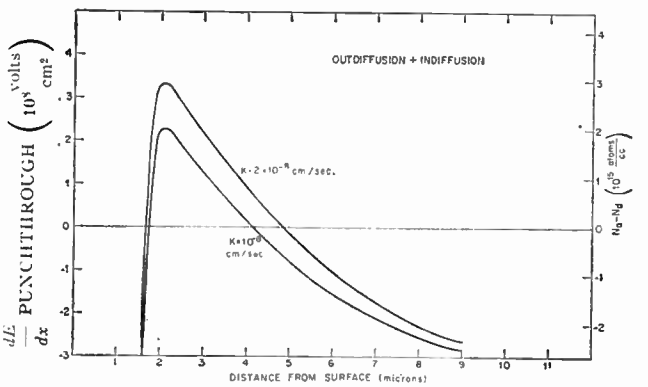


Fig. 2—Theoretical impurity profiles for outdiffused wafer of Fig. 1 after an indiffusion at 600°C for 4.5 hours. From Poisson's equation these can also be given in terms of dE/dx as shown on the left-hand ordinate.

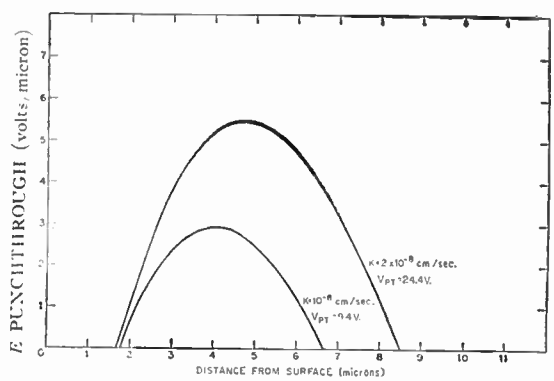


Fig. 3—Punch-through field as a function of distance into the crystal. These curves have been obtained from those of Fig. 2 by a point by point integration. The punch-through voltage corresponding to each of the K values is also given.

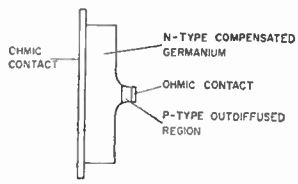


Fig. 4—Cross section of an outdiffused diode (not to scale).

punch-through voltage would be 24.4 volts. The measured punch-through voltage on a number of dice from this sample wafer was 15 ± 0.5 volts.[10] These results indicate a value of the surface-evaporation velocity K for the diffusion of antimony out of germanium at 700°C of $(1.5\pm0.5)\times10^{-8}$ cm/sec. An attempt has been made to include uncertainties in the theory and in the experiment in the indicated probable error. This value of K for antimony is in agreement with the value obtained by us over the past 18 months in other experiments on the punch-through voltage of outdiffused junctions. It is slightly smaller than the value 3×10^{-8} cm/sec reported by Lehovic and collaborators.[2] It is larger than the mean value for K of 6×10^{-9} cm/sec as determined from extrapolating the data of Miller and Smits[4] to 700°C but is within their indicated probable error.

It can be seen from Figs. 2 and 3 that the punch-through voltage is far more sensitive to variations in K than is the base width. Hence this method enables one to determine K more accurately than would a determination of the outdiffused junction depth.

The value we have obtained for the surface-evaporation velocity for antimony, the more rapidly diffusing impurity, is sufficiently large to make feasible the fabrication of diodes and transistors using outdiffusion techniques.[11] These devices will now be described.

III. Outdiffused Diode

A. Impurity Distribution

The outdiffused diode is a narrow base diode produced by the outdiffusion of a compensated n-type germanium wafer. The graded p-type skin so formed serves as the base of the diode. Fig. 4 illustrates the geometry of the outdiffused diode.

The starting material and the outdiffusion cycle used in the fabrication of most diodes give impurity profiles after outdiffusion which are very similar to the curves shown in Fig. 1. The impurity profile corresponding to any K value within our probable error $[K=(1.5\pm0.5)\times10^{-8}$ cm/sec] will fall between the two curves in the

[10] To make sure that the measured breakdown voltage was associated with punch-through and not avalanche, indiffusions to various depths were performed on dice all of which had originally come from the same outdiffused wafer. The breakdown voltage from sample to sample varied as would be expected for punch-through. In addition the breakdown voltage was measured as a function of temperature and the temperature dependence was as expected for punch-through (see O. Garretta, "Variation of the punch-through voltage of a transistor as a function of the temperature," *C. R. Acad. Sci.*, vol. 241, pp. 857–859; October, 1955) and opposite to that expected for avalanche.

[11] It has been pointed out by the reviewer that it is also possible to produce junctions by outdiffusion from compensated material in the case where the rapidly diffusing impurity has a smaller surface-evaporation velocity and a smaller concentration than that of the slower diffusing impurity. It also has been pointed out that if one had impurities with suitable diffusion coefficients and surface-evaporation velocities one could obtain two junctions by outdiffusion alone.

figure. An examination of Fig. 1 indicates that the net impurity density in the p-type base region can be much more closely represented by a linear function of distance than by the exponential approximation which gives a constant built-in field and which is usually assumed in calculations for graded regions. It is assumed that the material near the surface, in which the impurity density may deviate from this linear behavior because of the outdiffusion of the indium, will be absorbed into the ohmic contact when the device is fabricated (see Section III-E).

B. Current-Voltage Relationship

The current-voltage relationship for the outdiffused diode will be derived under the assumption that the net impurity density in the base region is a linear function of distance:

$$N_a - N_d = (x/w)N_o \tag{4}$$

where N_o is the net impurity density at the ohmic contact, w is the base width, and x is here defined as the distance from the junction towards the ohmic contact as is shown in Fig. 5(b). If conductivity modulation at high injection level is neglected, this doping distribution gives rise to a built-in field:

$$E = \frac{kT}{q}\,\frac{N_o}{2n_i w}\,\frac{1}{\sqrt{\left(\frac{N_o x}{2n_i w}\right)^2 + 1}}, \tag{5}$$

where k is the Boltzmann constant, T the absolute temperature, q the electronic charge, and n_i the intrinsic carrier density. This built-in field, which is shown in Fig. 5(c), can be approximated by a constant field

$$E = \frac{kT}{q}\cdot\frac{N_o}{2n_i w}$$

from the junction to the point $\delta = 2n_i w/N_o$ and by a hyperbolic field, $E = kT/qx$, from the point δ to the ohmic contact. Using this approximation it would be necessary to solve the diffusion equation for each of the two regions and to match boundary conditions. However, in the case of the doping gradients that are being used, it turns out from transition region theory[12] that the junction width at zero applied voltage extends to the point $x_m > \delta$ as is shown in Fig. 5(c) and hence one need only solve the diffusion equation:

$$D_n\frac{d^2n}{dx^2} + \mu_n n\frac{dE(x)}{dx} + \mu_n E(x)\frac{dn}{dx} = 0 \tag{6}$$

between the limits of x_m and w, where $E = kT/qx$ and n is the minority carrier concentration in the base. The

[12] The theory for the transition region for exactly this case is given in W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

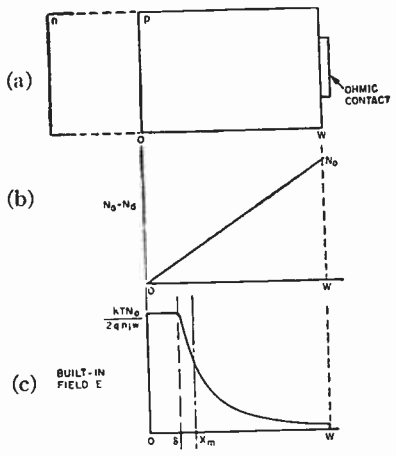


Fig. 5—Representation of the outdiffused diode (a) showing the assumed doping distribution (b) and the built in field (c) which results from this doping distribution. (Figure not drawn to scale.)

lifetime term in the diffusion equation has been neglected because the base width is much smaller than a diffusion length.

The boundary conditions that must be satisfied by the solution to (6) are: first, at $x = w$, the current density is

$$J_n = qs\left(n - \frac{n_i^2}{N_o}\right) = -qD_n\left(\frac{dn}{dx} + \frac{n}{w}\right), \tag{7}$$

where we have characterized the ohmic contact by a contact generation velocity, s, defined by the first equality in (7) as the ratio of the current density to the change in the charge density of minority carriers at the contact. The quantity n_i^2/N_o is the equilibrium value of this minority carrier density. The expression on the right-hand side of the second equality in (7) includes the current due to both diffusion and electric field. Second, at $x = x_m$ the minority carrier density is

$$n = \frac{n_i^2 w}{N_o x_m}\,e^{qV/kT}. \tag{8}$$

It should be noted that the quantity $n_i^2 w/N_o x_m$ in (8) is the equilibrium carrier density at x_m only if $x_m \gg \delta$, which condition implies that the net impurity density at x_m is significantly larger than the intrinsic carrier density. This is also the condition for the validity of the assumed hyperbolic field. For the impurity densities used, this condition is reasonably satisfied. The determination of x_m, the transition-region boundary, is discussed by Shockley[12] and can be determined for the case of complete depletion by solving the two equations relating the junction potential ψ with x_m.

$$\psi_m = \frac{q}{3\epsilon}\,\frac{N_o}{w}\,x_m^3 \quad \text{and} \quad x_m = \frac{n_i w}{N_o}\,e^{q\psi_m/kT}. \tag{9}$$

From the solution of the diffusion equation (6) subject to the boundary conditions (7) and (8) the current-voltage relationship can be obtained:

$$J = \frac{2qD_n n_w}{w}\left[\frac{1}{\frac{2D_n}{sw} + 1 - \frac{x_m^2}{w^2}}\right](e^{qV/kT} - 1), \qquad (10)$$

where $n_w = n_i^2/N_o$ is the equilibrium value of the minority carrier concentration at the ohmic contact. If $x_m^2 \ll w^2$, which is usually the case in practice, the current-voltage relationship can be written as

$$J = qn_w s\left[\frac{1}{1 + \frac{sw}{2D_n}}\right](e^{qV/kT} - 1). \qquad (10a)$$

In the derivation of (10) the minority carrier lifetime has been assumed large, the leakage resistance shunting the junction has been neglected (these assumptions can be fairly well approximated in fabricated diodes), and the hole current flowing in the n-type region has been neglected, since the ratio of this current to the electron current into the p-type skin is small because of the narrow base width. Eq. (10) may be compared with the current-voltage relationship for a uniformly doped narrow base diode, which has been derived under equivalent assumptions.[13]

$$J = qn_w s\left[\frac{1}{1 + \frac{sw}{D}}\right](e^{qV/kT} - 1). \qquad (11)$$

For narrow base diodes, as the reverse voltage is increased the space-charge region eats into a significant portion of the base region and the effective base width w is reduced. For a perfect ohmic contact, defined as $s = \infty$, the diode reverse current will vary as the reciprocal of the effective base width and hence will strongly depend on the reverse voltage; therefore the diode will not saturate. However, if the ohmic contact is not perfect, the reverse current will vary with effective base width as

$$\left(1 + \frac{sw}{2D}\right)^{-1}$$

for the outdiffused diode and as

$$\left(1 + \frac{sw}{D}\right)^{-1}$$

for a uniformly doped narrow base diode. The factor $\frac{1}{2}$ multiplying the term sw/D in the expression for the outdiffused diode reduces the dependence of the reverse current on the effective base width and thus improves the diode saturation characteristic. The physical reason for this improved saturation characteristic is that the field due to the graded doping in the base region helps to collect a greater percentage of the electrons that are generated at the ohmic contact even at larger base thicknesses. Thus as the effective base thickness is reduced and the collection efficiency of the junction increased, there are fewer additional electrons which can be collected.

In the above treatment conductivity-modulation effects which occur at high injection levels have been neglected. Because of its narrow base width the outdiffused diode requires a relatively small amount of carrier injection to produce the concentration gradient necessary for forward current flow. Conductivity modulation is advantageous, however, in that at large forward currents it reduces the built-in field which opposes the flow of this current.

C. Switching Speed

The switching speed of a narrow base diode such as the outdiffused diode is determined by the time it takes to switch the diode from the forward low-impedance state to the reverse high-impedance state. This reverse recovery time consists mainly of the sum of the time required to clear the stored minority-carrier charge from the base region and the time required to charge the junction depletion layer capacitance.

It has been shown[13] that a figure of merit for the minority-carrier storage switching time is the ratio of the forward current to the charge of the minority carriers stored during forward conduction. For linear grading in the base region such as assumed above in the derivation of the J-V relationship this figure of merit is

$$\frac{I_f}{Q} = \frac{2D}{w^2}\left[\frac{1}{\left(1 + \frac{2D}{sw}\right)\ln\frac{w}{x_m} - \frac{1}{2} + \frac{x_m^2}{2w^2}}\right], \qquad (12)$$

while for a uniformly doped base region it is given by

$$\frac{I_f}{Q} = \frac{2D}{w^2}\left[\frac{1}{1 + \frac{2D}{sw}}\right]. \qquad (13)$$

Because the built-in field opposes the flow of minority carriers when the outdiffused diode is biased in the forward direction, more minority carriers must be stored in the base region to produce a given forward current. As can be seen by comparing (12) and (13) for a given base width, the figure of merit for minority-carrier storage for an outdiffused diode is smaller than that for a uniformly doped diode.

Outdiffused diodes whose impurity profiles are similar to that indicated in Fig. 1 and whose base contacts have penetrated about 1.5×10^{-4} cm from the surface, have base widths w of approximately 3×10^{-4} cm and junction widths x_m at zero bias of approximately 0.1×10^{-4} cm. If the ohmic contact is characterized by a generation velocity $s = 5 \times 10^4$ cm/sec, the figure of merit for these diodes as determined from (12) is $I/Q \sim 4.5 \times 10^7$ sec^{-1}, which value is about three times smaller than that for a diode with a uniformly doped base region of identical base width. However, for the outdiffused diode the elec-

[13] R. H. Rediker and D. E. Sawyer, "Very narrow base diode," PROC. IRE, vol. 45, pp. 944–953; July, 1957.

tric field in the base region helps to clear the stored minority carriers when the diode is switched from forward to reverse bias and most of them are removed from the base region at the maximum current permitted by the external circuit. Thus a good approximation to the minority-carrier storage switching time for these outdiffused diodes is 22 mμsec (which is just the reciprocal of the figure of merit) multiplied by the ratio of the forward current to the maximum current permitted by the external circuit to clear the stored charge.[14]

In the outdiffused diode the entire base region is to a good approximation in the active region between the junction and the ohmic contact, as is shown in Fig. 4. Thus in this diode the problem of charge stored in regions of the base which are laterally displaced from the active region can be neglected, and the one-dimensional analysis presented above is applicable.[15]

In some narrow base diode designs the time required to charge the depletion layer capacitance is larger than the time required to clear out the stored minority carrier charge and becomes the major factor in determining the reverse recovery time. Outdiffused diodes have been fabricated in which this is indeed the case. Because of the grading of the net impurity concentration, the resistivity rising as the p-n junction is approached, the outdiffused diode will have a smaller depletion-layer capacitance than an equivalent diode with a uniformly doped base region of identical geometry and punch-through voltage (and therefore the same maximum voltage rating). The ratio of the capacitance of the outdiffused diode to that of this uniformly doped base-region diode (which may for the purposes of this discussion be assumed to be of alloy junction type[16]) is

$$\frac{C_{\text{out}}}{C_{\text{alloy}}} = 0.5\left\{\frac{(V+V_{o_{\text{alloy}}})^{1/2}}{(V+V_{o_{\text{out}}})^{1/3}}\right\}\frac{1}{(V_p)^{1/6}} \tag{14}$$

for any reverse voltage V. $V_{o_{\text{alloy}}}$ and $V_{o_{\text{out}}}$ are approximations to the internal contact potentials of the alloy and outdiffused junctions, respectively,[13,17] and have been neglected in the second fraction of (14) as they are small compared to punch-through voltage V_p. At any appreciable applied reverse voltage the junction contact potentials may be neglected and (14) reduces to

$$\frac{C_{\text{out}}}{C_{\text{alloy}}} = 0.5\left(\frac{V}{V_p}\right)^{1/6}. \tag{15}$$

To the same approximation, the ratio of the charge in the depletion layer of the outdiffused diode to the charge in the depletion layer of the equivalent alloy-type diode is:

$$\frac{Q_{\text{out}}}{Q_{\text{alloy}}} = \frac{3}{8}\left(\frac{V}{V_p}\right)^{1/6}. \tag{16}$$

Thus in switching from a given forward current to a given reverse bias more minority-carrier charge must be removed from the outdiffused diode than from the equivalent diode with a uniformly doped base region, but less charge must be put into the depletion layer.

D. Maximum Reverse Voltage

The two breakdown effects, the lower of which determines the maximum reverse voltage of the diodes discussed, are avalanche and punch-through. To maximize the reverse voltage capability of a uniformly doped base-region diode of any given base width, it is necessary to choose the resistivity such that avalanche breakdown and punch-through occur at the same reverse voltage since if one is increased the other will decrease. The higher resistivity near the junction gives the outdiffused diode a greater avalanche voltage, while the lower resistivity near the ohmic contact causes the rate of penetration of the space-charge region to tend to bog down, giving a greater punch-through voltage. Hence for any given base width the outdiffused diode can always be designed with a greater maximum reverse-voltage rating than can a diode with a uniformly doped base region.

E. Fabrication

Most of the diodes have been fabricated from 0.5 ohm-cm n-type compensated germanium wafers containing both antimony and indium as impurities ($N_d = 10.4\times10^{15}$; $N_a = 7.4\times10^{15}$ atoms/cc). The indium underdoping if uncompensated would be equivalent to 0.5 ohm-cm p-type. After suitable surface treatment the wafers are outdiffused for a period of 24 hours[18] in vacua of better than 10^{-7} mm of Hg at fixed temperatures which have ranged from 700°C to 800°C. After outdiffusion the p-type skin on one face of the wafer is etched or lapped off. Ohmic contact to the n-type bulk is then made by bonding this entire face to a gold-antimony plated tab. Ohmic contact to the p-type skin has been made in a number of ways including sandblasting followed by indium plating, and aluminum evaporation and alloying. These contacts absorb approximately one micron of germanium thus removing from the base of the diode that region where the net impurity distribution is not linear with distance. (Hence the analyses given in Sections III-B and III-C hold.) The ohmic contact is then masked and the area of the p-n junction defined by etching. Diodes have been fabricated having

[14] For analyses of minority carrier storage switching time for diodes with uniformly doped base regions see: R. H. Kingston, "Switching time in junction diodes and junction transistors," PROC. IRE, vol. 42, pp. 829–834; May, 1954, and B. Lax and S. F. Neustadter, "Transient response of a p-n junction," *J. Appl. Phys.*, vol. 25, pp. 1148–1154; September, 1954.

[15] In many other diodes and transistors, however, this charge exists and must be laterally removed. This adds appreciably to the minority carrier storage time and does not permit a one-dimensional analysis to be made for these devices.

[16] The condition for which (14) is derived is that the equivalent diode has a depletion layer which penetrates the base region only. The most common diode for which this condition applies is the alloy junction diode.

[17] D. R. Muss, "Capacitance measurements on alloyed indium-germanium junction diodes," *J. Appl. Phys.*, vol. 26, pp. 1514–1517; December, 1955.

[18] This time was chosen only because of its convenience for laboratory operation.

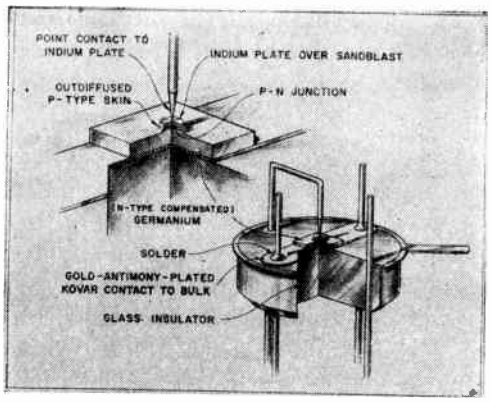


Fig. 6—Artist's representation of a typical outdiffused diode.

junction diameters varying from 0.010 inch to 0.030 inch. An artist's representation of a typical outdiffused diode is shown in Fig. 6.

Computer diodes have been fabricated by the outdiffusion process which meet the electrical specifications shown in Table I which has been adapted from Rediker and Sawyer.[13] It is believed that the outdiffusion process is eminently suited for the production of many unique types of fast diodes, the electrical specifications of just one of which are given in Table I.

TABLE I

ELECTRICAL SPECIFICATIONS FOR THE OUTDIFFUSED DIODE ALL PARAMETERS DEFINED AT (25±1.5)°C

Forward characteristics	$V=0.11$ volt $I>1$ ma $V=0.5$ volt $I>100$ ma
Reverse characteristics	$I_{sat}<25$ μamp $V_{max}>15$ volts (Voltage at which $I>100$ μamp) $r>750$ KΩ ($3\frac{1}{2}$ volts) $C<15$ μμf ($3\frac{1}{2}$ volts)
Reverse recovery time <0.15 μsec	Reverse recovery time is the time for the back resistance to recover to 100 K ($I<85$ μamp) when the test diode is switched from 2 ma forward current to 6 volts reverse bias (initial reverse current 6 ma).

IV. DOUBLE DIFFUSED *n-p-n* TRANSISTOR

A. General Considerations

The *n-p-n* transistor here described is a doubly diffused germanium transistor whose narrow-graded base region is produced by the outdiffusion of a compensated *n*-type germanium wafer and whose emitter is produced by indiffusion. Fig. 7 illustrates the geometry of the doubly diffused transistor.

As in the case of the outdiffused diode the impurity distribution in the base region of the double diffused transistor can best be represented by a net impurity concentration that is linear with distance. This impurity distribution gives rise to a field which aids in the transport of minority carriers, electrons, from emitter to collector. It therefore has the advantages of graded base or "drift" transistors which have been described in the literature[19–22] and have been well publicized. The calculation of Moll and Ross[22] of the frequency response and transit time for a transistor with a linear impurity distribution in the base region is of interest here. They show that, to within their clearly indicated approximations, a transistor with a linear impurity distribution in the base region has a transit time of one-half and a frequency response $2\sqrt{2}$ times that of a transistor with a uniformly doped base region of identical base width.

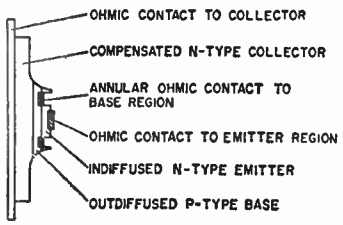


Fig. 7—Cross section of an *n-p-n* graded base transistor (not to scale).

From a fabrication standpoint ultraplanar junctions and extremely fine control of base-layer thickness can be obtained as a consequence of the double diffusion process.

B. Electrical Characterization of Fabricated Transistors

The technique of outdiffusion and then indiffusion is amenable to the production of devices with extremely high emitter-injection efficiencies. This high emitter efficiency may be ascribed to two factors. Firstly, since the emitter is diffused in, there is no liquid phase during its fabrication as there is in the case of evaporated and microalloyed emitters. This greatly reduces the dependence of the shallow emitter region on the surface condition of the germanium and also eliminates the problem of finding, for the *n*-type impurity, a carrier which does not have a detrimental effect on the subsequent cleanup of the emitter junction. Secondly, it is easy to sufficiently dope the emitter region by means of indiffusion of an *n*-type impurity so that the ratio of the total number of charge carriers in the emitter region to the total number in the base region is large. Indicative of this very high emitter efficiency are the values for grounded-emitter current gain β which have been obtained for outdiffused transistors. Curves of β as a function of collector current are shown in Fig. 8 for three "large area" experimental units. These devices have

[19] H. Krömer, "Der drift transistor," *Naturwiss.*, vol. 40, pp. 578–579; November, 1953.
[20] H. Krömer, "Zur theorie des diffusions und des drift transistors," *Arch. eleck. Ubertragung*, vol. 8, pp. 223–228, 363–369, 499–504; 1954.
[21] C. A. Lee, "A high frequency diffused base germanium transistor," *Bell Sys. Tech. J.*, vol. 35, pp. 23–34; January, 1956.
[22] J. L. Moll and I. M. Ross, "Dependence of transistor parameters on the distribution of base layer resistivity," PROC. IRE, vol. 44, pp. 72–78; January, 1956.

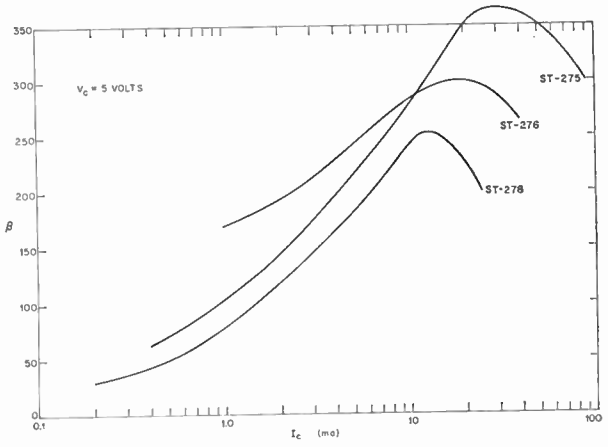


Fig. 8—Grounded-emitter current gain as a function of collector current for three "large area" laboratory *n-p-n* germanium transistors.

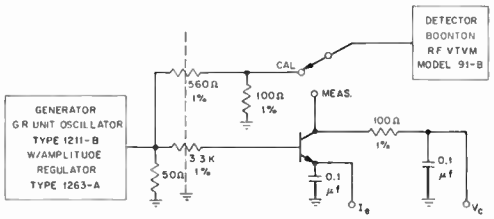


Fig. 9—Circuit for the measurement of gain-bandwidth product. The resistors have been so chosen as to give a gain of 5 in the calibrate position. The oscillator frequency is raised until the meas. and cal. positions give identical readings on the detector. Gain-bandwidth product is then just five times the indicated oscillator frequency.

circular emitter and collector areas, respectively 0.015 inch and 0.025 inch in diameter, and base widths of approximately 2.5×10^{-4} cm. Differences in the curves of Fig. 8 can be explained by modifications in the fabrication technique from unit to unit. For these and similar units, low-frequency values of β measured in the grounded emitter configuration have consistently averaged better than 100 over the range of 2–40 ma collector current and in some units have peaked at as high as 350. At lower currents, of the order of 100 μa, β's on some units have still been found to be as high as 10–15.

Gain-bandwidth product, which is an indication of frequency response, has been measured in the circuit shown in Fig. 9 and has been found to be of the order of 60 mc for these devices (at $V_c=5$ volts and $I_e=5$ ma). The gain-bandwidth product for these transistors is definitely limited by the large collector capacitance, which is of the order of 15 $\mu\mu$f.

Efforts are now being directed towards increasing the frequency response of these devices by reducing the electrode areas. As a step in this direction units have been fabricated which have had collector areas reduced by a factor of three from that of the above described large geometry transistors. These more recently fabricated devices have grounded emitter current gains, β, of over 100 and alpha frequency cutoffs, $f_{\alpha_{co}}$, of above 200 mc. A more complete set of electrical specifications which these units have met is shown in Table II.

TABLE II

ELECTRICAL SPECIFICATIONS FOR THE *n-p-n* GRADED BASE GERMANIUM TRANSISTOR

ALL PARAMETERS DEFINED AT (25±1.5)°C

Gain-bandwidth product (Measured in circuit of Fig. 9)	at 5 volts 5 ma	>100 mc
	10 volts 1.5 ma	> 50 mc
	1.5 volts 10 ma	> 50 mc
β_{min} at 20 ma		> 50
β_{max} at 5 ma		<200
I_{co} at 5 v		< 25 μa
C_e at 5 v		< 4 μμf
$V_{CE_{max}}\{I_{co}$ at $V_{CE_{max}}<2\ I_{co}$ at 5v$\}$		> 15 volts
$V_{EB_{breakdown}}(I_e<50$ μamp)		> 0.2 volt
$V_{sat\ max}$ at $I_e=20$ ma		< 1 volt
$f_{\alpha_{co}}$		>200 mc

These laboratory units which have been made, while showing the feasibility of the use of the outdiffusion process to make high-frequency *n-p-n* germanium transistors, have by no means exhausted the potentialities of the process in this direction.

C. Fabrication

Fabrication procedures for these units are identical with those described for the diode of Section III through the preparation of the outdiffused skin. Emitter regions are then produced by the diffusion in of arsenic. (A typical indiffusion is performed for 4.5 hours at 600°C.) Ohmic contact is made to the collector region by alloying on a gold-antimony plated kovar tab after etching to expose the *n*-type bulk. Ohmic contacts are made to the emitter and base regions with the aid of photolithographic techniques.[23] Masks of appropriate geometry are formed by the exposure of photosensitive lacquer (with which the germanium dice are coated) through photographic transparencies. For the emitter contact gold-antimony is evaporated on and microalloyed in. In view of the inherent difficulties in the production of extremely thin, injecting *n*-type regions by means of evaporation and alloying, or by shallow alloying, one cannot overemphasize the importance of the injecting region being built into the device thus necessitating only an ohmic contact to be made to the emitter. Base contact is made by controlled etching[24] through the emitter region after suitably masking the device and then evaporating on and microalloying in an aluminum layer. The collector area is also defined using photographic masking techniques and then etching. An artist's representation of a typical unit is shown in Fig. 10.

[23] These techniques were discussed in a paper presented by J. W Lathrop and J. R. Nall at the IRE-PGED Meeting, Washington, D. C., November, 1957.

[24] Etching has been controlled to $\pm0.5\times10^{-4}$ cm using an etch composed of 10 cc HF, 10 cc H_2O_2 and 40 cc H_2O.

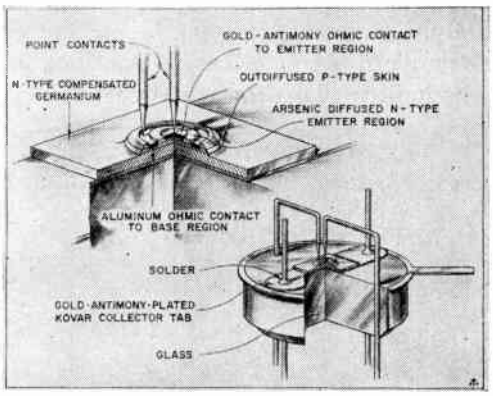


Fig. 10—Artist's representation of a typical *n-p-n* double diffused transistor.

Conclusions

The diffusion of antimony out of germanium has been studied and the use of this process has been shown to be an easy way of producing *p-n* junctions in compensated *n*-type germanium. Outdiffusion has been used to produce high-speed narrow base diodes and high-speed *n-p-n* germanium transistors. While diodes and transistors which have been fabricated using this technique have many desirable and presently unavailable electrical characteristics, the potentialities of the process have been by no means exhausted, either for the production of these or other types of semiconductor devices.

Acknowledgment

The authors wish to express their appreciation to F. M. Sullivan and C. R. Grant for help in fabricating the diodes and transistors and to P. L. Moody and S. A. Kulin for growing the compensated germanium. They are also indebted to E. Chatterton for designing the circuits used to determine the frequency response of the transistors and for taking these measurements. The advice and encouragement of D. T. Stevenson is gratefully acknowledged.

The Evolution of the Theory for the Voltage-Current Characteristic of *P-N* Junctions*

J. L. MOLL, MEMBER, IRE†

Summary—**The rectifying action in semiconductor *p-n* junctions is controlled essentially by the equilibrium densities, diffusion constants, and recombination times of minority carriers. The low-level behavior of germanium junctions at room temperature is adequately described by a theory which is based on the rate of diffusion and recombination of minority carriers on either side of the barrier region. It is necessary to include the effects of carrier recombination and generation in the barrier region to explain the low-level behavior of silicon.**

At high current densities, the junctions depart from the ideal low-level rectifier law because of effects associated with majority carrier modulation. As a consequence of recombination current to the barrier region at low levels and conductivity modulation effects at high levels, the simple $I_s(\exp qV/kT-1)$ behavior is rarely observed in silicon junctions at room temperature.

EVER since the observation of nonohmic behavior in metal-semiconductor contacts, the theoretical explanation of rectification has received a great deal of attention. Most of the theories that were proposed prior to Shockley's definitive article[1] in 1949 contained some features in common with modern theory. However, the theories generally omitted the important role of minority carriers and, for this reason, can be considered of value only to the extent that they helped workers in the field to reach present-day concepts. There is one very notable exception to the preceding statement, *viz.*, a theoretical paper by Davydov in 1938.[2] In this paper Davydov considered most of the essential features of modern low-level rectification theory and specifically pointed out the importance of minority carriers in determining the rectifying action. He also included the concepts of nonequilibrium density and finite lifetime, with perhaps the most notable lapse being the assumption that equilibrium minority and majority carrier density was related to lifetime.

An extremely abbreviated list of pre-1949 contributions would include such names as Wilson and Nordheim who independently proposed the tunnel effect as an explanation of rectification.[3,4] This theory was quickly discarded since it gave the wrong direction of

* Original manuscript received by the IRE, April 4, 1958.
† Bell Telephone Labs., Inc., Murray Hill, N. J.

[1] W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

[2] B. Davydov, "The rectifying action of semiconductors," *Tech. Phys., (USRR)*, vol. 5, pp. 87–95; February, 1938.

[3] A. H. Wilson, "A note on the theory of rectification," *Proc. Roy. Soc.*, vol. 136, p. 487; May, 1932.

[4] L. W. Nordheim, "Xur Theorie der Detector wirkung," *Zeits. f. Phys.*, vol. 75, p. 434; April, 1932.

rectification. However, Esaki has noted rectification by tunneling in extremely thin (*i.e.*, highly doped) junctions.[5]

The possibility of relatively thick barriers between the p and n-type regions was also considered at an early date, and the role of impurity density in determining barrier-layer thickness was recognized.[6,7] If Davydov's paper is disregarded (as was generally done at the time even though it was published in English in 1938), the general view of rectification until 1949 was that n-type and p-type semiconductors were "boxes" of Maxwell-Boltzmann gases of electrons and holes, respectively, with a potential barrier at the p-n boundary. The potential barrier keeps the two gases from spilling across the boundary, and the thickness of the potential barrier is determined from the height of the barrier and the density of ionized impurities. As the applied voltage is changed, escape of particles from the two gases becomes more or less difficult depending on the direction of the applied voltage.[8] Now this view of the rectifying action gives the right qualitative result of an exponential law for rectification and is, in fact, consistent with present theory. However, the theory which has been built up with Shockley's 1949 paper[1] as a basis is more useful for understanding p-n junctions. In Shockley's theory the elementary processes of the escape of the Maxwell-Boltzmann gas have been separated, with the result that the various modifications of the theory which are necessary to fit specific situations are easier for the intuition to grasp. With Shockley's theory it becomes possible, for the first time, to make a quantitative check of the theory in terms of the basic semiconducting properties of the component parts of the junction.[9]

The escape of an electron from the n side to the p side of a junction involves:

1) Surmounting the potential barrier at the junction,
2) Disappearance on the p side by recombination (or removal across a reverse junction in the case of a transistor).

In addition to the process of escape of carriers with recombination, there is, of course, the inverse process of generation of minority carriers which diffuse to the junction where they are swept down the potential barrier. Where the generation is thermal, the process of generation appears as simply a negative recombination or vice versa. If the generation is by means of photons or other particles, it is usually easiest to treat the generation as a separate process. It is not our object here to treat any situations other than the rectification of p-n junctions.

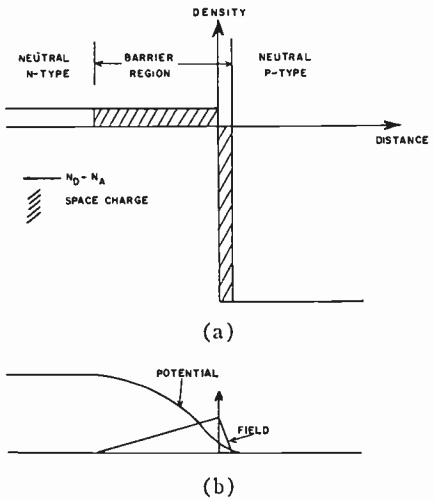


Fig. 1—Idealized planar step junction. (a) Assumed distribution of impurity distribution. It is assumed for purposes of analysis that all of the mobile carriers have been swept out of the barrier region resulting in a space-charge density which is simply the ionized donor or acceptor density. (b) Field distribution and potential in a step junction. The width of the barrier region adjusts itself so that the resulting dipole will sustain the contact potential plus the applied voltage.

Following Shockley's treatment[1] the mathematical model which is used to represent the p-n junction is illustrated with the aid of Fig. 1. The junction and infinitely long pieces of p-type and n-type semiconductor are divided into three separate regions: 1) the neutral n-type region, 2) the neutral p-type region, and 3) the barrier region, or depletion region on both sides of the p-n boundary. It is assumed that the majority carriers have been essentially all swept out of the barrier region, and that a space charge of the ionized donors and acceptors exists. The essentially field-free nature of the neutral region requires that the positive charge per unit area of the donors on the n side of the boundary be balanced exactly by the negative charge per unit area of the acceptors on the p side. Thus, the barrier region extends far enough on either side of the junction to sustain the barrier potential, with the relative extent of the region on the two sides being governed by the requirement of net neutrality. The validity of the assumption of an abrupt transition from neutral material to swept-out material is considered in some detail in Shockley's paper[1] where the assumption is shown to be valid for junctions with steep concentration gradients. For junctions with sufficiently shallow concentration gradients, there is no space-charge region or rectification. For the case of the very shallow gradient, the change in equilibrium carrier concentration with distance is so slow that it is not possible to disturb the equilibrium with applied field, and there is no rectification. The change in equilibrium conditions at the p-n boundary is

[5] L. Esaki, "A new phenomenon in narrow germanium p-n junctions," *Phys. Rev.*, vol. 109, pp. 603–604; January, 1958.
[6] N. E. Mott, "The theory of crystal rectifiers," *Proc. Roy. Soc.*, vol. A171, p. 27; May, 1939.
[7] W. Schottky and E. Spenke, "Zur Quantitativen Durch führung der Rauinladungs-und Randschichttheorie der Kristallgleichrichter," *Wiss. Veroff. Siemenswerke*, vol. 18, p. 225; October, 1939.
[8] See for example, H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 78–79; 1948.
[9] F. S. Goucher, G. L. Pearson, M. Sparks, G. K. Teal, and W. Shockley, "Theory and experiment for a germanium p-n junction," *Phys. Rev.*, vol. 81, pp. 637–638; February, 1951.

essential to the rectifying action. The junctions that are of interest in semiconductor technology generally fall well within the required range of steepness to justify the assumption that the transition between neutral material and material which has had all of the majority carriers swept out is infinitely sharp.

The part of the conduction process which involves surmounting the potential barrier at the junction is most easily understood as a deviation from the thermal equilibrium condition. At thermal equilibrium, the product of electron and hole densities throughout the sample is constant. Thus, we may write[1]

$$pn = n_i^2 \tag{1}$$

where

p = hole density
n = electron density
n_i = intrinsic carrier density.

The intrinsic carrier density is a characteristic of the semiconductor and is a strong function of temperature. The value of n_i^2 for germanium at room temperature is approximately 10^{26}, and doubles with each increase in temperature of 8°C. On the n-type side of the junction, the number of electrons is very nearly equal to the number of excess donors, so that electrons outnumber holes by many orders of magnitude for reasonably large excess donor density. Reciprocal considerations hold for the p-type side. The solid curves in Fig. 2 show the variation of electron and hole density through the junction at equilibrium. The sum of the logarithms of the equilibrium densities is a constant, independent of position.

When a bias is applied to the junction, the equilibrium is disturbed. The p-n product changes in the barrier region by a Boltzmann factor,[1] and becomes

$$pn = n_i^2 \exp(qV/kT) \tag{2}$$

where

V is the applied voltage (p positive with respect to n)
q = electronic charge
k = Boltzmann constant
T = absolute temperature.

The dotted curves in Fig. 2 show the disturbed carrier densities for a forward applied bias. The p-n product is constant through the barrier region and approaches the thermal equilibrium value in the neutral material as the excess carriers diffuse away from the junction and recombine.

The effect of (2) is that when the injected minority carrier density is small compared to the equilibrium majority carrier density, the majority carrier density on either side of the barrier region remains unchanged. However, the minority carrier density at the boundary of the neutral regions and the barrier is increased by the exp (qV/kT) factor. We see then that the effect of a given voltage on a junction is to disturb the carrier

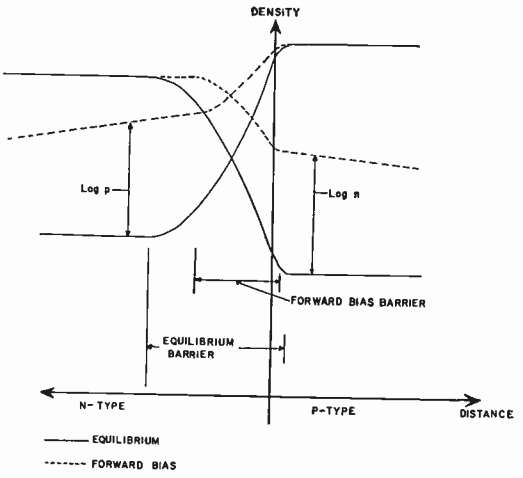


Fig. 2—Carrier density at and near a p-n junction. The solid curves are log n and log p, respectively, for the thermal equilibrium conditions. The sum of the logarithms is constant throughout the specimen. The dashed curves show the densities with a moderate forward bias. The sum of the logarithms is essentially constant in the barrier region implying a constant n-p product. The n-p product decreases in the neutral material from its value in the barrier to the thermal value far from the junction.

densities. The actual current through the junction is limited by the rate of disappearance (or appearance) of the excess (or deficit) minority electrons and holes.

The rate of disappearance of the minority carriers is governed by their rate of diffusion and recombination in the neutral regions. The problem of determining the junction current now becomes the problem of calculating hole and electron densities (and flow) in each of the two neutral regions subject to the boundary conditions of (2) at the boundary of the barrier, and (1) at $x \rightarrow \pm \infty$.

The transport equations for minority carriers are particularly simple for small densities (linear equations) and uniform semiconducting regions (no build-in fields). The result of the calculation of carrier densities for this simple case, interpreted in terms of current, is[1]

$$I = I_s[\exp(qV/kT) - 1)], \tag{3}$$

where

I = junction current amp/cm^2
V = junction voltage—volts
I_s = saturation current $= q[(n_{p0}/\tau_n)L_n + (p_{n0}/\tau_p)L_p]$
τ_n = electron lifetime in p-type semiconductor
τ_p = hole lifetime in n-type semiconductor
n_{p0} = thermal electron density in p-type semiconductor
p_{n0} = thermal hole density in n-type semiconductor
$L_n = \sqrt{D_n\tau_n}$ = electron diffusion length
$L_p = \sqrt{D_p\tau_p}$ = hole diffusion length
D_p = electron diffusion constant
D_p = hole diffusion constant.

We will refer to this result (3) as the ideal rectifier equation since it was derived for the p-n junction with quite drastic simplifying assumptions on the nature of the p-n boundary and the conduction process. In addition, the resulting rectification ratio is the best that is

possible in a simple electronic process. For a forward bias ($V>0$) greater than a few kT/q, the current increases exponentially with voltage. For reverse bias ($V<0$), I saturates at $-I_s$ which corresponds to the rate of generation of minority carriers in the slice which is L_p thick on the n side and L_n thick on the p side of the junction. The solid curve in Fig. 3 is a plot of the ideal rectifier (3). In the case of a reverse bias greater than several kT/q, the boundary condition (2) does not give the correct minority carrier densities.[10] For this case, practically no carriers surmount the barrier, and the density is determined by the rate of flow of holes from the n side and of electrons from the p side across the junction. However, the minority carrier density has become so small that (2) continues to give the correct answer in most practical problems. Eq. (3), which predicts a saturation of current with reverse voltage, continues to hold.

One of the high points in the evolution of junction theory was the development of (3) which predicts the current-voltage characteristic in terms of basic semiconductor properties and its quantitative verification in germanium p-n junctions.[1,9] The simple theory of Shockley[1] has served as a basis for many extensions and modifications. In particular, it was found that the range of current and voltage for which junctions satisfied (3) was limited in the case of germanium to low-level operation and in the case of silicon was nonexistent at room temperature. The departures from (3) are always in the direction of poorer rectification, where rectification is taken as the ratio of forward to reverse current at some low voltage. As based on the semiconducting properties of the two neutral regions and the simple theory (3), more current is conducted in the reverse or blocking direction than I_s, and in the forward or easy conducting direction more voltage than the amount specified in (3) is required. This departure is illustrated in Fig. 3 where voltage is plotted on a linear scale, and current on a logarithmic scale. The slope of the ideal forward characteristic is kT/q volts per e (naperian log base) of current or 2.3 kT/q per decade of current. We will refer to the ideal case as a kT/q slope or exp qV/kT behavior in discussing departures from the simple case. The degraded rectification characteristic in Fig. 3 has a slope steeper than kT/q.

The reasons for the departure from the ideal case are seen most easily as modifications of the simple theory. Our approach will be to consider departures first in the reverse characteristic, then at low forward bias, and finally at high forward bias since this is generally in the direction of increasing complexity. It is convenient to divide the diode V-I characteristic into four distinct regions for the purposes of analysis. These are: 1) high reverse bias, 2) low and medium reverse bias, 3) low forward bias, 4) high forward bias. The boundaries of

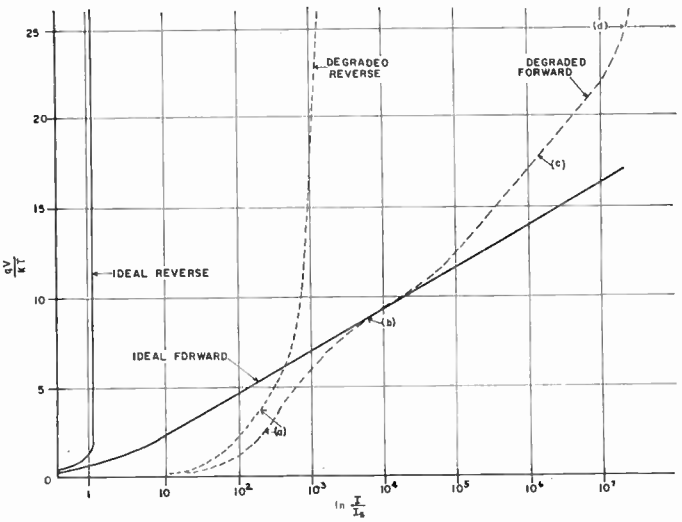


Fig. 3—Rectifier characteristics. Modifications of the ideal theory lead to degradation of the rectifier characteristic. In (a) the conduction mechanism is dominated by the space-charge region carrier recombination and generation. Note the lack of saturation in the reverse characteristic and extra conduction at low forward voltages. As the forward voltage is increased the diffusion current becomes dominant and there may be a tendency towards ideal behavior (b). At high-level injection (c) the diffusion current departs from the ideal law, with an increase in the semilogarithmic slope. Finally the current is limited by series ohmic resistance (d).

these regions are not very sharp but general distinctions can be made.

The problem of high reverse bias is the problem of the mechanism of reverse breakdown and is a subject unto itself. It is not within the scope of this article to consider the breakdown mechanism in detail; however, a short discussion is warranted. In 1953 McKay and McAfee showed that the breakdown in moderately wide junctions occurs through a mechanism of carrier multiplication or avalanche in a manner almost completely described by the Townsend gaseous discharge mechanism.[11–13] In the avalanche process, electrons and holes acquire enough energy from the applied field to cause secondary production of electron-hole pairs. The secondary pairs, in turn, can cause further ionizations. In very thin junctions, the breakdown is believed to be by the Zener mechanism whereby electrons are pulled out of the valence bond by the excessively high fields that occur.[14] In silicon junctions the breakdown which occurs at voltages greater than about 6 v is avalanche breakdown. However, through no fault of Zener, the breakdown has been called "Zener breakdown." A name more appropriate than the erroneous "Zener Diode," as applied to the usual voltage-regulator diode, might be

[10] C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in p-n junction characteristics," Proc. IRE, vol. 45, pp. 1228–1243; September, 1957.

[11] K. G. MaKay and K. B. McAfee, "Electron Multiplication in silicon and germanium," *Phys. Rev.*, vol. 91, p. 1079; September, 1953.

[12] K. G. McKay, "Avalanche breakdown in silicon, "*Phys. Rev.*, vol. 94, pp. 877–884; May, 1954.

[13] S. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, vol. 99, pp. 1234–1241; August, 1955.

[14] A. G. Chynoweth and K. G. McKay, "Internal field emission in silicon p-n junctions," *Phys. Rev.*, vol. 106, pp. 418–426; May, 1957.

simply, "Avalanche Diode" or perhaps "*V-R* Diode" (Voltage Regulator Diode), in analogy to the *V-R* tube which performs the same function in circuits and actually uses the avalanche mechanism. We will take medium reverse bias as being reverse bias sufficiently small that avalanche effects are negligible. This requirement is satisfied for silicon and germanium diodes at voltages roughly less than half the breakdown voltage. (Since the multiplication increases the alpha of a transistor above unity, this restriction to half the breakdown voltage is generally not nearly strict enough for the case of the transistor collector.)

For medium reverse bias, measurements on actual diodes generally result in more current than I_s as given in (3). Part of this current may be a result of surface leakage, inversion layers,[15,16] or simply gross body defects in the junction. The surface leakage as well as inversion layers can be eliminated by proper treatment of the surface while body defects can be eliminated by making the junction so small that there is a statistical chance that all gross defects have been excluded. After all of these precautions have been taken, excess current is still observed in silicon junctions at room temperature and in germanium junctions at lower temperatures. This extra (over the simple theory) current has been attributed to charge generation in the barrier region.[10,17–19] The rate of carrier generation by recombination-generation centers[20] in the barrier region, where almost all of the mobile carriers have been swept away, is approximately

$$G_{\text{barrier}} = n_i/\tau \tag{4}$$

whereas the generation rate in a region where only the minority carriers have been removed is

$$G_{\text{neutral}} = n_{\text{min}}/\tau \tag{5}$$

where n_{min} is the equilibrium density of minority carriers, and τ = lifetime.

The equilibrium minority carrier density is much less than the intrinsic density so that the generation rate in the barrier region is much greater than in the neutral material. The saturation current in (3) varies as the second power of intrinsic carrier density, but the generation current that arises from the generation rate in (4) varies as the first power of intrinsic carrier density. As a result, this generation current is dominant when the intrinsic carrier density is small as is the case for silicon at room temperature or germanium at lower temperatures. The reverse space-charge generation current is proportional to the volume in which generation occurs and is thus proportional to the barrier width. The barrier width generally increases with reverse voltage so that there is no true reverse saturation of current for the case where space-charge generation dominates reverse current. For germanium at room temperature, the intrinsic carrier density is large enough for the saturation current of (3) to be dominant, so that germanium *p-n* junctions obey the simple theory for moderate reverse bias.

If the barrier region generation term is dominant in the case of small reverse bias, this effect will also dominate the conduction mechanism of (3) for a small forward bias.[10] As a result, more current will be conducted at a small forward voltage than is predicted by (3). This current[10] will have a semilogarithmic slope greater than kT/q at biases of several kT/q and explains the departure of silicon *p-n* junctions from the simple theory even at small biases (Fig. 3).

As the forward bias is increased, the diffusion current into the neutral region increases faster than the recombination current in the barrier layer until, finally, the diffusion and recombination of carriers in the neutral region is the essential conduction mechanism. At this point we might expect the diffusion theory that results in (3) to give the correct diode behavior, *i.e.*, a kT/q slope and, in fact, a small region where the *V-I* tends towards exp (qV/kT) behavior has been observed in silicon rectifier diodes.[19] However, at high forward bias so many minority carriers are injected across the junction that the minority carrier density becomes of the same magnitude as the majority carrier density, and (3) must be modified. For this case of high injection, the assumption of charge neutrality outside the barrier region is still satisfied. However, the electric field in the neutral region is no longer negligible. Majority carriers move to the junction to neutralize the minority carriers —thus establishing a gradient of majority carriers. Fig. 4 shows the near equality of majority and minority carriers under conditions of high injection. In this case the *p*-type side is more heavily doped (more extrinsic) than the *n*-type side so that modulation occurs essentially on the *n*-type side only. The gradient of electrons results in a diffusion current of electrons; however, a small unbalance of charge results in a field that is just strong enough to result in an equal and opposite drift current. Exact analysis is difficult since the transport equations for carriers become nonlinear.[21,22] The result-

[15] M. Cutler and H. M. Bath, "Surface leakage current in silicon fused junction diodes," PROC. IRE, vol. 45, pp. 39–43; January, 1957.
[16] W. T. Eriksen, H. Statz, and G. A. DeMars, "Excess surface currents on germanium and silicon diodes," *J. Appl. Phys.*, vol. 28, pp. 133–139; January, 1957.
[17] H. Kleinknecht and K. Seiler, "Einkristalle und ph Schichtkristalle aus Silizium," *Zeits. f. Phyisk*, vol. 139, pp. 599–618; December, 1957.
[18] E. M. Pell and G. M. Roe, "Reverse current and carrier lifetime as a function of temperature in germanium junction diodes," *J. Appl. Phys.*, vol. 26, pp. 658–665; June, 1955.
[19] H. S. Veloric and M. B. Prince, "High voltage conductivity-modulated silicon rectifier, "*Bell Sys. Tech. J.*, vol. 36, pp. 975–1004; July, 1957.
[20] W. Shockley and W. T. Read, Jr., "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835–842; September, 1953.
[21] W. M. Webster, "On the variation of junction-transistor current-amplification factor with emitter current," PROC. IRE, vol. 42, pp. 914–920; June, 1954.
[22] E. S. Rittner, "Extension of the theory of the junction transistor," *Phys. Rev.*, vol. 94, pp. 1161–1171; June, 1954.

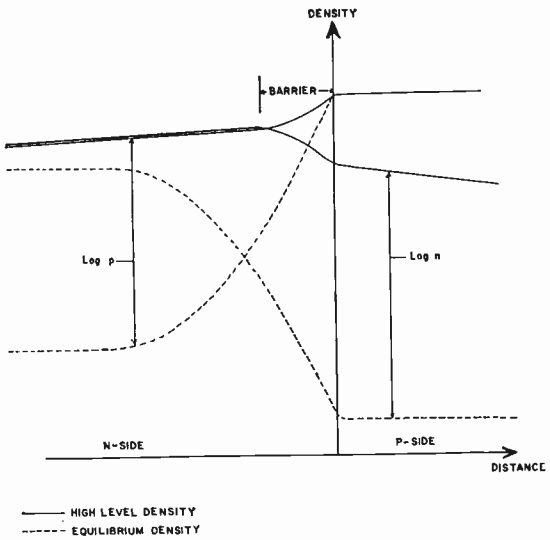


Fig. 4—The solid curves show carrier density near a p-n junction under conditions of high-level injection. The p-n product is constant in the barrier region, and $p \approx n$ in the neutral n-type side near the junction. For reference, equilibrium densities are shown in the dashed lines.

ing minority carrier drift current is in the same direction as the minority carrier diffusion current, thereby increasing minority carrier current.

We must re-examine the boundary condition in the barrier as expressed in (2) for application to the high-level case. The solid curves in Fig. 4 show the carrier densities under high-level injection conditions on one side (n side). The p-n product is constant, as before, through the barrier region. However, the majority carrier density (electrons) at the edge of the barrier on the n side has increased to several times the equilibrium number (conductivity modulated). Eq. (2) can still be used to obtain the p-n product in the barrier if the voltage V is interpreted to include the voltage associated with the field that holds the majority carriers in place as well as the voltage that appears across the barrier region, but excluding ohmic drops.[10] The near equality of majority and minority carriers under high-level injection conditions results in an exp $qV/2kT$ rate of increase of minority carriers. Thus the high-level rate of increase of minority carriers is slower than the low-level rate, and is in the direction of less current than (3). An additional complication analyzing the junction characteristic is the fact that carrier lifetime is a function of carrier density with lifetime increasing with minority carrier density in silicon.

The general approach to the problem of calculating the voltage-current characteristic for the p-n junction with high-level forward bias has been to simplify the problem in some respect until the differential equations can be handled. A junction of particular interest is the PIN structure. In order to reduce series ohmic resistance to a minimum (and still maintain high reverse breakdown), high conductivity contacts are placed near enough to a junction where one side is only weakly extrinsic so that conductivity modulation extends from one contact to the other at high injection. The density of impurities in the weakly extrinsic part of the junction, and even the question of whether the conductivity is p or n-type becomes immaterial in the high injection condition. Thus the three-layer junction which consists of 1) heavily doped p-type, 2) a central layer of lightly doped p or n-type, and 3) heavily doped n-type semiconductor is called the PIN diode. At low biases the PIN diode (as usually constructed) behaves like a junction between heavily doped and lightly doped semiconductors of opposite conductivity types. In 1952 Hall[23] set up a model for the PIN structure in which he assumed that all of the recombination occurs in the central region, with the result that for almost all forward biases, the junction follows exp $qV/2kT$. In this case, equal voltages must be supplied at the $p+I$ and $IN+$ junctions to get carriers into the central intrinsic region.

In 1956 Kleinman[24] considered the same problem with the assumption that recombination in the central $I-$ region was negligible, with the result that the voltage drop across the central region is negligible, and the high-level, forward current-voltage characteristic is determined by the $P+$ and $N+$ end regions. Kleinman concluded that a nonlinear recombination law was necessary to explain the departure of the diode from the ideal theory.

The actual PIN diode has a finite amount of recombination in each of the three regions, and is undoubtedly a compromise between the two idealized models. However, the analytic problem of the PIN diode with finite recombination rates that are functions of density in all three regions is extremely difficult. At the present time we can take the results of Hall[23] and Kleinman[24] to give the qualitative result that the semilogarithmic slope of the forward characteristic should be greater than kT/q.

The p-n step junction has also received considerable attention in the high forward bias range. In this case the high injection range begins when the minority density on one side becomes comparable with the majority carrier density.[10,25] High injection begins on the high resistivity side since the minority carrier density is always greater on this side as compared to the heavily doped side. The calculation of the diffusion current again involves solution of nonlinear differential equations, with the result of various approximations again indicating exp $qV/2kT$ behavior.

A p-n junction that is simple enough for the separate high-level effects to be separately considered is the emitter of a transistor. In this case the carriers that are injected into the base layer disappear in an average time,

[23] R. N. Hall, "Power rectifiers and transistors," Proc. IRE, vol. 40, pp. 1512–1519; November, 1952.
[24] D. A. Kleinman, "The forward characteristic of the PIN diode," *Bell Sys. Tech. J.*, vol. 35, pp. 685–706; May, 1956.
[25] J. S. Saby, "Junction Rectifier Theory," Proc. Rugby Conference, Rugby, Eng.; 1956.

which is the diffusion time across the base layer, and the injected current is given by[26]

$$I = \frac{qpnD_{\min}}{\int_0^w ndx} \tag{6}$$

where

p and n are the carrier densities just inside the base layer at the emitter junction,
$D_{\min}$ is the average diffusion constant for the minority carrier in the base, and
$\int_0^w ndx$ is the integral of majority carrier density the base layer and is the total number of majority carriers per cm² in the base layer.

From (2) and (6) we see that for the one-dimensional transistor case the emitted current is not reduced until the average number of majority carriers in the base layer is increased. It is possible, for example, in diffused emitter transistors for the emitter to be in a high injection state but for the total number of carriers to remain essentially constant. In this case, the high-level injection condition holds, but the emitter diode follows the ideal diode formula. When high injection ($p \approx n$) extends across most of the base layer, the number of majority carriers in the base is proportional to the density at the emitter and the current-voltage characteristic follows an exp ($qV/2kT$) law where V must be interpreted as the voltage across the emitter and includes the contribution to voltage of the field in the neutral region which arises as a result of majority carrier modulation but does not include resistive drops.

A modification of the emitter junction, which is of some interest, is the emitter junction of the diffused base, diffused emitter transistor where the contact is made to the base by alloying through the emitter layer.[27] In this case a relatively small contact is obtained and the problem must be considered as a two-dimensional one. Lateral flow of base current in the relatively high resistivity base causes the emission to be concentrated around the base contact so that the effective area is reduced as the current increases. The result for some cases of idealized geometry is exp ($qV/2kT$) behavior.[28] Emission concentration can occur simultaneously with high-level injection with the result that behavior ranging from exp ($qV/1.5kT$) to exp ($qV/3kT$) has been observed. In silicon alloy transistors made on high resistivity base material, the lateral concentration effect crowds emission to the edge of the emitter region resulting in a semilogarithmic slope greater than kT/q.

A fairly wide range of exp (qV/kT) behavior has been observed in the emitters of diffused base silicon transistors with alloyed emitters. The minority carriers were removed by the collector so that the effective lifetime was the essentially constant diffusion time across the base layer. In addition, the conductivity of the base was high enough that there was no appreciable emission crowding. This type of junction did exhibit exp ($qV/2kT$) behavior at low currents due to space-charge generation of carriers.

Conclusions

The low level p-n junction theory of Shockley, which is based on the statistical escape of particles across a barrier and diffusion into the region beyond the barrier predicts the ideal diode formula, (3), for rectification. This theory adequately explains the low-level operation of p-n junctions that have fairly large saturation currents such as germanium junctions at room temperature.

The low-level operation of junctions made from materials with low saturation currents (such as silicon at room temperature) requires the inclusion of the effects of generation and recombination of electron-hole pairs in the barrier region. In silicon at room temperature and germanium at low temperatures, the generation-recombination dominates the low-level operation with a resulting departure from ideal rectification.

The high-level forward biased junction has a greater voltage drop than predicted by the ideal rectifier formula. This extra voltage arises because of the phenomena associated with conductivity modulation in part of the diode. Exact theory for the conductivity modulated case is difficult to obtain, but solutions for various limiting cases result in a qualitative explanation of the diode characteristics.

As a result of the departure from the ideal law at low biases due to space-charge recombination and generation, ideal rectifier behavior is never observed in silicon junctions at room temperature. The departures from the simple theory at high biases limits the possible range of kT/q behavior to moderate forward bias, and in most cases there is no range of kT/q behavior at all. However, germanium does show ideal behavior at low biases.

The exact formulation of the problem of calculating the V-I of the p-n junction is very difficult when either the low-level effect of barrier region recombination and generation or the high-level effect of conductivity modulation is included. In addition, it is impractical to obtain better than order-of-magnitude guesses as to the magnitudes of some of the semiconductor parameters. As a result, close agreement between theory and experiment is difficult. However, effects that have been discussed are in good qualitative agreement with experiment in both silicon and germanium p-n junctions.

Acknowledgment

The author wishes to acknowledge the helpful discussion and criticism of many of his colleagues.

[26] J. L. Moll and I. M. Ross, "The dependence of transistor parameters in the distribution of bass layer resistivity," Proc. IRE, vol. 44, pp. 72–78; January, 1956.
[27] M. Tanenbaum and D. E. Thomas, "Diffused emitter and base silicon transistors," *Bell Sys. Tech. J.*, vol. 35, pp. 1–22; January, 1956.
[28] F. M. Smits (private communication).

Analog Solution of Space-Charge Regions in Semiconductors*

L. J. GIACOLETTO†, SENIOR MEMBER, IRE

Summary—**Analytical solutions of space-charge regions in semiconductors can be carried out when a one-dimensional symmetry exists. Analog solutions can be used to solve Poisson's equation for more complex geometries. An analog method is described which has been used with some success. This method consists of a rubber membrane with surface mass loading being used to simulate the semiconductor space charge.**

Introduction

FOR a one-dimensional rectangular geometry the depletion of mobile carriers in a semiconductor by a junction biased in the reverse direction is governed by the following Poisson equation[1]

$$\frac{d^2V}{dx^2} = -\frac{\rho}{K\epsilon_0} = \frac{q}{K\epsilon_0}(N_a - N_d), \tag{1}$$

where V is the voltage at distance, x, into the semiconductor, ρ is the space-charge density, K is the relative dielectric constant, ϵ_0 is the permittivity of free space, q is the charge of an electron, and N_a and N_d are the acceptor and donor impurity densities, respectively. The solution of this equation using the boundary conditions,

$$V = 0, \quad \frac{dV}{dx} = 0 \quad \text{at} \quad x = X,$$

is

$$V = \frac{q(N_a - N_d)}{2K\epsilon_0}(x - X)^2, \tag{2}$$

where X, the distance to the edge of the depletion region, is determined with (2) by applying the boundary condition that $V = V_B + V_0$ at $x = 0$.

$$X = \sqrt{\frac{2K\epsilon_0}{q(N_a - N_d)}(V_B + V_0)}. \tag{3}$$

In this equation V_B is the applied bias voltage, and V_0 is an internal contact voltage. The per unit area capacitance of the junction is

$$C/A = \frac{K\epsilon_0}{X}. \tag{4}$$

It is possible to obtain similar solutions of space-charge regions in semiconductors when a one-dimensional symmetry exists.[2,3] Solutions for cylindrical and spherical geometrys are shown in Fig. 1.

For a two-dimensional rectangular geometry the problem is to solve Poisson's equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = \frac{q}{K\epsilon_0}(N_a - N_d), \tag{5}$$

with the boundary condition that $V = 0$, $\partial V/\partial x = 0$, and $\partial V/\partial y = 0$ at the boundary of the depletion. Attempts to solve this problem analytically are complicated by the fact that the boundary condition is not fixed but is, in fact, the part of the solution that is of greatest interest. One method of starting an analytical solution would be to formulate the space charge as a function of voltage such that the function is constant for any finite voltage of a given polarity, changes discontinuously to zero when voltage becomes zero, and then remains zero for any voltage of the opposite polarity.

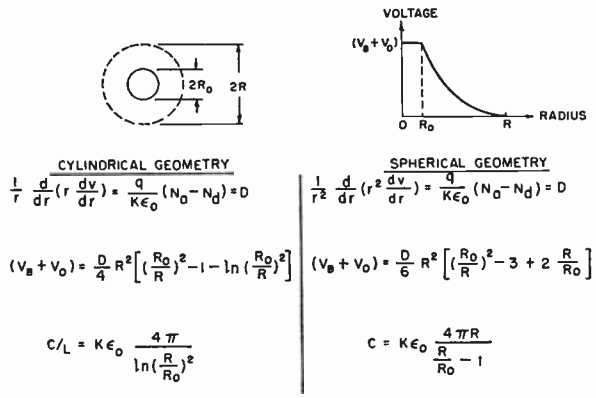


Fig. 1—Voltage and capacitance for space-charge regions in semiconductors with one-dimensional geometry.

Analog Solution

Analog techniques[4] have been used to obtain solutions of the two-dimensional space-charge depletion problem. One method of analog solution is to use a resistance net, as employed for solving Laplace's equation, with constant current of appropriate value injected at each node point to simulate space charge at that point. Current injection is moved progressively

* Original manuscript received by the IRE, February 20, 1958. This investigation was carried out while the author was with RCA Labs., Princeton, N. J.

† Scientific Laboratory, Ford Motor Co., Dearborn, Mich.

[1] Relevant material on the space-charge depletion problem and associated junction capacitance will be found in L. J. Giacoletto, "Junction capacitance and related characteristics using graded impurity semiconductors," IRE TRANSACTIONS ON ELECTRON DEVICES, vol. ED-4, pp. 207–215; July, 1957.

[2] H. L. Armstrong, E. D. Metz, and I. Weiman, "Design theory and experiments for abrupt hemispherical *p-n* junction diodes," IRE TRANS. ON ELECTRON DEVICES, vol. ED-3, pp. 86–92; April, 1956.

[3] H. L. Armstrong, "A theory of voltage breakdown of cylindrical *p-n* junctions, with applications," IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 15–16; January, 1957.

[4] J. H. O. Harries, "The rubber membrane and resistance paper analogies, PROC. IRE, vol. 44, pp. 236–248; February, 1956.

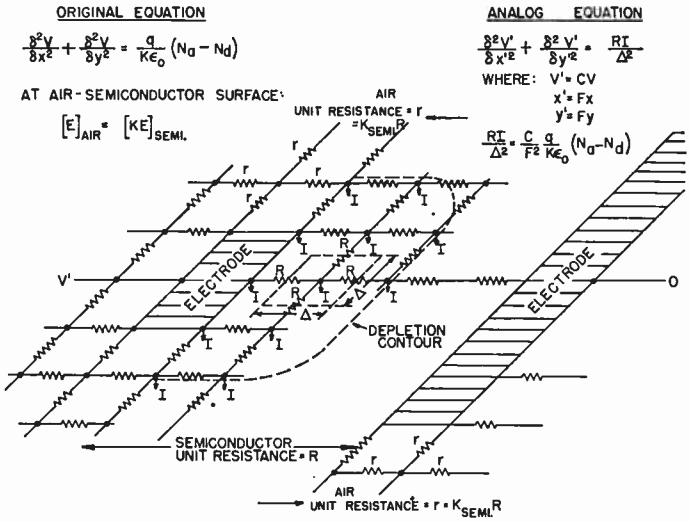


Fig. 2—Resistance net analog of space-charge regions in semiconductors.

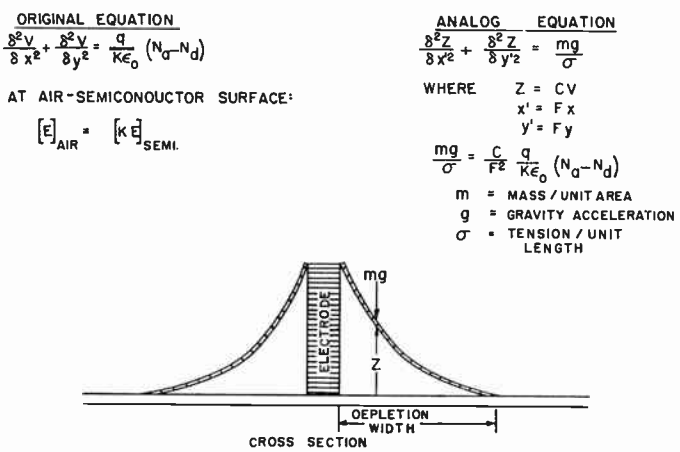


Fig. 3—Rubber membrane analog of space-charge regions in semiconductors.

outward from the simulated junction electrode until the node voltage is zero. This method of solution is shown in Fig. 2. Semiconductor variable impurity density can readily be simulated by injecting different currents at the various node points rather than constant current. The boundary condition at the air-semiconductor surface can readily be accommodated by suitably altering the mesh resistance as shown in Fig. 2. For a typical semiconductor depletion problem, the number of meshes required to give a reasonable accuracy generally will be prohibitively large. Some experiments were attempted using resistance paper in place of the resistance net with current being injected through the paper. The results were not satisfactory due, in large part, to the inappropriate ratio between surface resistivity and transverse paper resistivity.

It is possible to obtain an analog solution of Poisson's equation with the aid of a rubber membrane as has been extensively employed for the solution of Laplace's equation. Mass loading of the membrane as shown in Fig. 3 is used to simulate space charge. The mass loading is moved progressively outward from the simulated electrode until the deflection of the rubber membrane (simulated voltage) is zero. The furthermost extent of the loading determines the depletion contour. In the event that the membrane is on a supporting table, the mass loading can be applied uniformly over the entire membrane, and the depletion contour is determined by the line of initial contact between the membrane and table. The scaling factor between mass and space charge involves the determination of the tension per unit length in the rubber membrane; this is not easily evaluated. A convenient way for determining the scaling factor is to set up the analog solution of a one-dimensional rectangular depletion problem whose solution is known. The simulation of the air-semiconductor boundary condition could be obtained by adjusting the elastic constant of the rubber membrane so that the membrane gradients at the simulated air-semiconductor boundary are in the same proportion as the relative dielectric constants. The adjustment of the membrane elastic constant is rather difficult. If the relative dielectric constant of the semiconductor is large (as in the case for Ge and Si), the assumption can be made that there is no flux extending beyond the semiconductor and therefore that the voltage gradient in the semiconductor at the air-semiconductor boundary is parallel to the boundary. This approximate boundary condition can then be simulated by making the air-semiconductor surface a plane of symmetry.

Typical Solutions

Figs. 4 and 5 illustrate the use of a loaded rubber membrane for analog solutions of two-dimensional semiconductor depletion problems. Fig. 4 illustrates depletion from a surface electrode of a finite length. Sheet-lead squares one centimeter on edge were used to simulate space charge. At places where the membrane deflection gradient was large and at the edge of the depletion, lead squares 1/4 centimeter on edge were used. The membrane was loaded on both sides of the simulated electrode to provide a surface plane of symmetry to accommodate, as mentioned above, the air-semiconductor boundary. The parabolic variation expected for a one-dimensional depletion problem is clearly evident near the center of the electrode. It also is evident that the greatest deflection (voltage) gradient occurs along the surface of the simulated semiconductor rather than inside the semiconductor. This is another reason why surface treatment is so important in obtaining high breakdown voltages for semiconductor junctions.[5]

Fig. 5 illustrates the simulated solution for depletion from two surface electrodes of finite extent. Interaction between electrodes is now apparent; independent depletion takes place until the depletion contours touch after which the membrane height (voltage) increases at the center and lateral depletion occurs in the Y directions.

[5] C. G. B. Garret and W. H. Brattain, "Some experiments on, and a theory of, surface breakdown, *J. Appl. Phys.*, vol. 27, pp. 299–306; March, 1956.

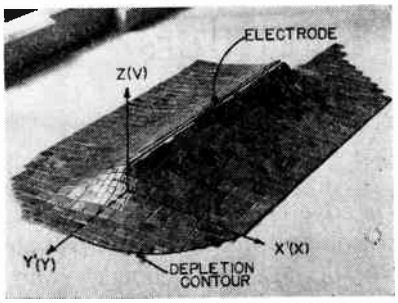


Fig. 4—Rubber membrane analog of space-charge region with depletion occurring from a single surface electrode.

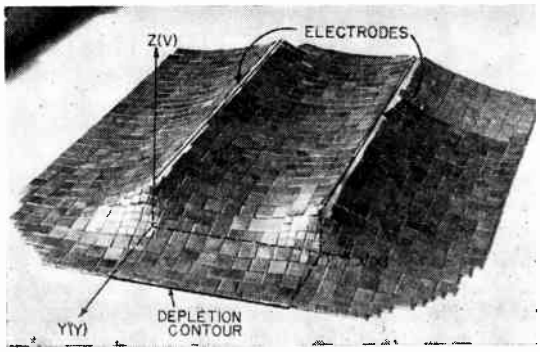


Fig. 5—Rubber membrane analog of space-charge region with depletion occurring from two surface electrodes.

Conclusions

With considerable care quantitative results can be obtained using the loaded rubber membrane. However, it is believed that the real merit of the rubber membrane analog is the use of it as an aid for mentally visualizing complex depletion problems. For such problems, one should first visualize the rubber membrane solution of the Laplace problem associated with the particulae geometry and then visualize the modification of thr membrane height as unit area loading is added starting from the simulated electrodes and moving outward.

Correction

Henry G. Booker, author of "The Use of Radio Stars to Study Irregular Refraction of Radio Waves in the Ionosphere," which appeared on pages 298–314 in the January, 1958 issue of Proceedings, has advised the editors of the following numerical errors.

On page 307, (1) should read

$$\lambda_N^2 = \frac{\pi}{r_e N},$$

and (4) should read

$$\overline{\left(\frac{\Delta n}{n}\right)^2} = \frac{1}{4}\,\frac{\lambda^4}{\lambda_N^4}\,\overline{\left(\frac{\Delta N}{N}\right)^2}.$$

Consequently, on pages 307–310 the expression $\overline{(\Delta N)^2}$ should be replaced everywhere by $\overline{(\Delta N)^2}/(4\pi^2)$.

Germanium and Silicon Rectifiers*

H. W. HENKELS†, SENIOR MEMBER, IRE

Summary—**Two general types of rectifier are singled out for review in this paper. These are the silicon and germanium large area *p-n* junction rectifying cells. A discussion is given of various phenomena which have a bearing upon the volt-ampere characteristics. Methods for fabricating the various kinds of *p-n* junctions into rectifying cells are described and typical volt-ampere characteristics are given. Finally, a brief summary of the kinds of applications these rectifiers have filled is provided with some typical illustrations.**

I. Introduction

SEMICONDUCTOR rectifiers from a variety of materials have been employed in limited applications for very many years. The industry rapidly expanded with the introduction of, first, the copper oxide, and then the selenium rectifier. No sooner had standards for the latter units, termed at times dry plate or metallic rectifiers, been finalized than the germanium alloyed junction device was introduced. The initial impetus given the study of germanium and silicon during World War II, and the acceleration of activities with the discovery of transistor action, have resulted in the germanium device noted, and more recently, in the power silicon rectifier.

This review will be concerned almost exclusively with germanium and silicon rectifiers with readily discernible single-crystal *p-n* junctions. Thus, point contact diodes and their modifications, in which local alloying or diffusion of impurities to form a doped semiconductor phase may result, metal-to-semiconductor rectifiers, and surface barrier rectifiers will not be expressly treated. Parts of the basic theory are applicable to such devices but additional developments are needed in considering the general problem of practical metal contacts to semiconductor surfaces. There exists a quite extensive literature on point contact diodes. Some selected references are given here.[1–14] Recent work on the subject is covered by Uhlir.[15] Early theories of the metal-to-semiconductor contact are found in references to the work of Mott,[16] Schottky and Spenke,[17,18] Davydov,[19] and Jaffe,[20] and are summarized in a book by Henisch.[21] More recent studies are contained in the work of Billig and Landsberg[22] and Landsberg.[23] Much of the difficulty with the treatment of metal-semiconductor contact lies in the problem of the contact potentials of surfaces. Considerable attention has been given to the subject in the literature.[24–29] More recently the development of new experimental techniques (jet-

* Original manuscript received by the IRE, March 3, 1958.

† Westinghouse Electric Corp., Youngwood, Pa.

[1] J. H. Scaff and R. S. Ohl, "Development of silicon crystal rectifiers for microwave radar receivers," *Bell Sys. Tech. J.*, vol. 26, pp. 1–30; January, 1947.

[2] W. H. Brattain and J. Bardeen, "Nature of the forward current in germanium point contacts," *Phys. Rev.*, vol. 74, pp. 231–233; July, 1948.

[3] H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, N. Y.; 1948.

[4] J. Bardeen, "On the theory of the a-c impedance of a contact rectifier," *Bell Sys. Tech. J.*, vol. 28, pp. 428–434; July, 1949.

[5] S. Benzer, "High inverse voltage germanium rectifiers," *J. Appl. Phys.*, vol. 20, pp. 804–815; August, 1949.

[6] V. A. Johnson, R. N. Smith, and H. J. Yearian, "D-C characteristics of silicon and germanium point contact crystal rectifiers. II. The multicontact theory," *J. Appl. Phys.*, vol. 21, pp. 283–289; April, 1950.

[7] H. J. Yearian, "D-C characteristics of silicon and germanium point contact crystal rectifiers," *J. Appl. Phys.*, vol. 21, pp. 214–221; March, 1950.

[8] M. Cutler, "Point contact rectifier theory," IRE Trans. on Electron Devices, vol. ED-4, pp. 201–206; July, 1957.

[9] P. C. Banbury, "Theory of the forward characteristic of injecting point contacts," *Proc. Phys. Soc. B (London)*, vol. 66, pp. 833–840; October, 1953.

[10] J. H. Simpson and H. L. Armstrong, "Reverse characteristics of high inverse voltage point contact germanium rectifiers," *J. Appl. Phys.*, vol. 24, pp. 25–34; January, 1953.

[11] P. M. Tipple and H. K. Henisch, "Thermal effects at point contact diodes," *Proc. Phys. Soc. B (London)*, vol. 66, pp. 826–832; October, 1953.

[12] J. A. Swanson, "Diode theory in the light of hole injection," *J. Appl. Phys.*, vol. 25, pp. 314–323; March, 1954.

[13] M. Cutler, "Forward characteristics of germanium point contact rectifiers," *J. Appl. Phys.*, vol. 26, pp. 949–954; August, 1955.

[14] T. E. Firle, M. E. McMahon, and J. F. Roach, "Recovery time measurements on point contact germanium diodes," Proc. IRE, vol. 42, pp. 603–607; May, 1955.

[15] A. Uhlir, "The potential of semiconductor diodes in high-frequency communications," this issue, p. 1099.

[16] H. F. Mott, "The theory of crystal rectifiers," *Proc. Roy. Soc.*, vol. 171A, p. 27; May, 1939.

[17] W. Schottky and E. Spenke, "Zur Quantitativen Durchfuhrung der Raumladungs und Randschichttheorie der Krystall Gleichrichter," Wissenschaftliche Veroffentlichungen aus dem Siemens-Werken, vol. 18, p. 225; 1939.

[18] W. Schottky, "Simplified and extended theory of barrier layer rectifiers," *Z. Phys.*, vol. 118, pp. 539–592; 1942.

[19] B. Davydov, "On the contact resistance of semiconductors," *J. Phys. USSR*, no. 1, p. 167; 1939.

[20] A. Jaffe, "Electrical resistance of contact between a semiconductor and a metal," *J. Phys. USSR*, vol. 10, pp. 49–60; 1946.

[21] H. K. Henisch, "Metal Rectifiers," Oxford University Press, New York, N. Y.; 1949.

[22] E. Billig and P. T. Landsberg, "Characteristics of compound barrier layer rectifiers," *Proc. Phys. Soc. (London)*, vol. 63A, pp. 101–111; 1950.

[23] P. T. Landsberg, "Further results in the general theory of barrier layer rectifiers," *Proc. Phys. Soc. (London)*, vol. 65B, pp. 397–409; June, 1952.

[24] J. Bardeen, "Surface states and rectification at a metal-to-semiconductor contact," *Phys. Rev.*, vol. 71, pp. 717–727; May, 1947.

[25] W. E. Meyerhoff, "Contact potential difference in silicon rectifiers," *Phys. Rev.*, vol. 71, pp. 727–735; May 15, 1947.

[26] W. H. Brattain and W. Shockley, "Density of surface states on silicon deduced from contact potential measurements," *Phys. Rev.*, vol. 72, p. 345; August, 1947.

W. H. Brattain, "Evidence for surface states on semiconductors from change in contact potential on illumination," *Phys. Rev.*, vol. 72, p. 34; July, 1947.

[27] J. J. Markham and P. H. Miller, Jr., "Effect of surface states on temperature variations of the work function of semiconductors," *Phys. Rev.*, vol. 75, pp. 959–967; March, 1949.

[28] R. J. Rothlein and P. W. Miller, Jr., "Measurements of the variation of work function of silicon with temperature," *Phys. Rev.*, vol. 76, p. 1882; December, 1949.

[29] H. H. Smith, "Temperature dependence of the work function of semiconductors," *Phys. Rev.*, vol. 75, pp. 953–958; March, 1949.

etching and plating) in connection with the surface barrier transistor has awakened renewed interest in the subject.[30–33]

This paper presents the general field of germanium and silicon rectifiers. An attempt will be made to treat the general synthesis of contacts and rectifiers. A section on rectifier preparation is presented. The basic features of these methods, among the possible ones noted, that are presently used commercially, are reviewed. The final device properties, appearance, static characteristics, and typical ratings are discussed in the next section. Finally, a number of important uses are shown to illustrate the wide range of application of these devices which have been available for only a few years.

II. Rectifier Theory

A discussion of germanium and silicon rectifier theory must begin with the treatment of single crystal *p-n* contact theories. These describe simplified structures which are for the most part one-dimensional models. Unfortunately, the correlation of most properties of practical rectifiers with the contact theories requires *a posteriori* evaluation of essential parameters from measured characteristics. The approach neglects problems of structure synthesis and uniqueness. The latter may be of little concern provided consistent theories can indicate qualitatively valid paths to follow in the improvement of rectifiers, but the former problem of synthesis is the one of principal concern to the rectifier industry. In addition, no attempt is made in contact theory to treat the more mundane problems of making electrical connections to the semiconductor, of providing for heat dissipation, and of enclosure to permit the finished units to function under the conditions of interest in applications.

A. Contact Theory

The contact theories applicable to germanium and silicon single crystals are those which developed after the discovery of the nonequilibrium steady-state carrier injection condition in single crystals. Consequently, the early theories already referenced are neglected. A historical development begins with the early work of Shockley,[34] which is more applicable to the germanium contact, and has been followed by important contributions by Shockley and Read,[35] Hall,[36] Pell,[37] and Sah, Noyce, and Shockley[38] to a bulk contact theory involving both diffusion currents, and additional contributions from current generated at traps in the space charge layer. The breakdown conditions involve separate consideration of McKay and others.[39–43] High forward current conditions with consideration of the base contact are noted by Fletcher.[44] Additions to the theories must be made to account for surface leakage effects (see Law,[45] Christensen,[46] McWhorter and Kingston,[47] and Cutler and Bath[48]). Finally, one very important subject—the current through contacts immediately after switching from reverse to forward directions—has been treated in the literature (see Shulman and McMahon,[49] Kingston,[50] and Steele[51]).

B. Junction Synthesis

The first junctions were made by crystal pulling techniques. By altering the composition of a semiconductor melt during the operation of pulling a single crystal, the desired *p-n* junction can be produced. Such junctions are generally more perfect than those made in other ways, and give breakdown voltages more in line with those theoretically predicted. However, except in special cases where high-voltage capabilities or very low leakage characteristics must be used, because of the low current handling capacities, the rectifiers from pulled crystals find restricted application. Consequently, these types will not be discussed further.

[30] W. E. Bradley, "The surface barrier transistor. Part I. Principles of the surface barrier transistor," Proc. IRE, vol. 41, pp. 1702–1706; December, 1953.
[31] F. A. Schwartz and J. F. Walsh, "The surface barrier transistor. Part V. The properties of metal-to-semiconductor contacts," Proc. IRE, vol. 41, pp. 1715–1720; December, 1953.
[32] E. H. Borneman, F. A. Schwartz, and J. J. Stickler, "Rectification properties of metal-semiconductor contacts," *J. Appl. Phys.*, vol. 26, pp. 1021–1028; August, 1955.
[33] E. C. Wurst, Jr. and E. H. Borneman, "Rectification properties of metal-silicon contacts," *J. Appl. Phys.*, vol. 28, pp. 235–240; February, 1957.
[34] W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.
[35] W. Shockley and W. T. Read, Jr., "Statistics of recombinations of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835–842; September, 1952.
[36] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, p. 387; July, 1952.
[37] E. M. Pell, "Reverse current and carrier lifetime as a function of temperature in germanium junction diodes," *J. Appl. Phys.*, vol. 26, pp. 658–665; June, 1955.
[38] C. Sah, R. W. Noyce, and W. Shockley, "Carrier generation and recombination in *p-n* junctions and *p-n* junction characteristics," Proc. IRE, vol. 45, pp. 1228–1243; September, 1957.
[39] K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, vol. 94, pp. 877–884; May, 1954.
[40] K. G. McKay and K. B. McAfee, "Electron multiplication in silicon and germanium," *Phys. Rev.*, vol. 91, pp. 1079–1084; September, 1953.
[41] P. A. Wolff, "Theory of electron multiplication in silicon and germanium," *Phys. Rev.*, vol. 95, pp. 1415–1420; September, 1954.
[42] S. L. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, vol. 99, pp. 1234–1241; August, 1955.
[43] H. S. Veloric, M. B. Prince, and M. J. Eder, "Avalanche breakdown voltage in silicon diffused *p-n* junction as a function of impurity gradient," *J. Appl. Phys.*, vol. 27, pp. 895–899; August, 1956.
[44] N. H.Fletcher, "The high current limit for semiconductor junction devices," Proc. IRE, vol. 45, pp. 862–872; June, 1957.
[45] J. T. Law, "A mechanism for water induced excess reverse dark current on grown germanium *n-p* junctions," Proc. IRE, vol. 42, pp. 1367–1370; September, 1954.
[46] H. Christensen, "Surface conduction channel phenomena in germanium," Proc. IRE, vol. 42, pp. 1371–1376; September, 1954.
[47] A. L. McWhorter and R. H. Kingston, "Channels and excess reverse current in grown germanium *p-n* junction diodes," Proc. IRE, vol. 42, pp. 1376–1380; September, 1954.
[48] M. Cutler and H. M. Bath, "Surface leakage current in silicon fused junction diodes," Proc. IRE, vol. 45, pp. 39–43; January, 1957.
[49] R. G. Shulman and M. E. McMahon, "Recovery currents in germanium *p-n* junction diodes," *J. Appl. Phys.*, vol. 24, pp. 1267–1272; October, 1953.
[50] R. H. Kingston, "Switching time in junction diodes in junction transistors," Proc. IRE, vol. 42, pp. 829–834; May, 1954.
[51] E. L. Steele, "Charge storage in junction diodes," *J. Appl. Phys.*, vol. 25, pp. 916–918; July, 1954.

Of paramount importance to the production of rectifiers is the quality of the semiconductor.

Germanium is commonly zone refined and doped in a horizontal crystal furnace. Crystals may be grown in such a furnace or in the conventional vertical pulling furnace. Silicon is doped and grown in the latter furnace or in zone levitation equipments. Most of the group III and V elements have been employed in producing the desired doping, but As and Sb doping in n-type Ge, Ga doping in p-type Ge, As in n-type silicon, and residual B or Ga doping in p-type silicon are common. The values of breakdown voltages as a function of resistivity given by the theory are achieved only in selected small junctions. In larger units in germanium near-intrinsic doping is commonly employed; in silicon higher ranges as indicated later.

Average lifetimes in the initial ingots of well over 100 μsec in germanium and over 50 μsec in silicon are usual. Recently, crystals with lifetimes to the millisecond region have been employed. Trap densities for the most part cannot be detected directly. Crystals with lineage, aggregations, or dislocations give poor results regardless of lifetime measurements. It is difficult to detect bad effects of twinning in production. Crystals in both the (111) and (100) orientations are used.

Some of the mechanical properties are pertinent to later discussion: the thermal expansion coefficient, the bulk or Young's modulus, and the ultimate stress. In germanium the values are 6×10^{-6} inch/inch°C, 5×10^{6} psi (100), 23×10^{6} psi (111) and 25,000 psi (tension, est.), respectively; in silicon they are 4.1×10^{-6} inch/inch°C, 25×10^{6} psi (111), and 17,000 psi (tension).

1) *Fused Rectifiers:* The first of the two widely used methods for producing p-n junctions is that of fusion. This technique produces a p-n junction by melting an alloy in contact with the semiconductor, dissolving the semiconductor surface to some extent, and redepositing the dissolved semiconductor on the single crystal substrate by cooling the system.

The reverse diffusion current through the p-n junction can be minimized by having a highly doped region of appreciable thickness in contact with a region of opposite low conductivity. The extent of doping of the redeposit is the product of the distribution coefficient, defined as the ratio of the concentration of the impurity in the solid semiconductor to that in the liquid alloy, provided the solubility of the impurity in the solid semiconductor is not exceeded. If the alloy is that of the dissolved semiconductor and a pure metal, the distribution coefficient alone can be obtained directly from the binary phase diagrams (which have been detailed in the high semiconductor region with common group III and V doping agents as one of the constituents). Little information exists on distribution coefficients for other than the simple semiconductor-element binaries. Data on solubilities of other than group III and V elements are limited. Besides the requirements that the acceptor or donor metal or alloy redeposit in the semiconductor to give a highly doped region, there are additional restrictions in the counterelectrode material: 1) it must dissolve sufficient semiconductor to give a redeposit which is reasonably thick at conventional processing temperatures; 2) in general, it must not form diffusable traps in the semiconductor. (In certain cases, however, traps have been added to the counterelectrode material to reduce the charge storage effect noted in Section II.) In the highly doped p-to-n cases, the indium and aluminum element-semiconductor binaries can be and are widely used on germanium and silicon, respectively. In the complementary, highly doped n-to-p cases, the n-doping agents are always carried in a matrix of other metals because of the nonmetallic nature of the pure elements.

Besides the requirements of the alloy or metal concerned with the p-n contact formation itself, there are others associated with the problems of mechanical stability, lead attachment, heat transfer, and encapsulation. They are as follows.

1) It must be liquid at a temperature as far as possible below the melting temperature of the semiconductor or other components of the cell.
2) It must wet the semiconductor and counterelectrode contact metal.
3) It should be solid at all temperatures encountered during processing after the fusion cycle, and solid, mechanically and chemically stable during storage or operation (to required specifications).
4) It should have a low vapor pressure at all temperatures encountered in the fusion cycle.
5) It should have high thermal conductivity.
6) It should not produce excessive strain in the semiconductor.

The first condition arises from the desire to retain the original lifetime of the semiconductor as much as possible. It is not absolutely necessary for the counterelectrode solder to remain solid during processing. Good cells have been made in which the condition was violated, but a solid electrode simplifies post-fusion process controls. The fourth condition is desirable from the standpoint of the ease of post-etching, but again is not essential. The temperature drops in the solders should be low to permit the greatest possible heat dissipation within the cell. The electrical properties depend upon strain to some extent, but it is more important for the semiconductor to be safely below its ultimate elongation in the range of temperatures found in the processing of the basic diode and the application of finished rectifiers.

Condition 6) can be re-expressed quantitatively in terms of the properties of the counterelectrode solder and semiconductor for certain geometries of cell conconstruction,[52] and restated in fundamental terms.

[52] H. W. Henkels, "The fused silicon rectifier," *Commun. and Electronics*, vol. 28, pp. 733–746; January, 1957.

7) The product of thickness, modulus of elasticity, and difference of thermal coefficients of silicon and the alloy must be as small as possible.
8) The solidus temperature of the alloy should be as low as possible consistent with condition 3).

The eight conditions should be considered before noting the doping effect of the alloy since the addition of trace elements will in most cases adequately dope the the redeposited layer without greatly affecting those listed properties except for condition 2) which may be very sensitive to trace impurities.

a) *Germanium contact solders:* In the case of the germanium rectifier, the operating junction temperature (condition 3)) is restricted to about 100°C by the properties of the p-n junction itself. Consequently, low-temperature solders are perfectly satisfactory as long as melting points are safely above this temperature (say 150°C) and processing steps subsequent to the p-n junction fusion operation can be carried out without damage to that junction. It has been noted that junction lifetimes in germanium rectifiers deteriorate badly if the processing temperature exceeds 600°C or so. Thus, a range of solidus and liquidus temperatures from 150°C to 600°C is possible. Conditions 7) and 8) favor a low melting soft solder; condition 5), harder solders with better thermal conductivity. The higher melting possibilities are Ag, Au, and Al. The former appears to affect badly the junction for reasons unknown to the author.

Counterelectrodes on n-type material will be considered first. The aluminum electrode has been employed in emitter contacts of transistors to give good emitter efficiency on low resistivity base material. This situation is not present in rectifiers. Conditions 2) and 6) lead to restricted use despite the good thermal conductivity. Au must be doped to form a p^+ contact. Of the doping possibilities, only Ga is readily carried out and has been used in some small fused wire rectifiers ("gold bonded"). Au-B, Au-Al alloys are brittle and difficult to prepare and fabricate. Practically all p^+n rectifiers presently made employ the low melting alloys despite the low thermal conductivities. (The high thermal resistance may be relatively unimportant in certain designs in comparison with air and liquid film resistances.) Most common of all counterelectrode materials for n-type silicon is pure In. The melting point is satisfactory and operations subsequent to fusion can be carried out without damage to the recrystallized p^+ region. The only drawback is the low thermal conductivity which has been a problem in some cases. The use of In-Ga alloys has been largely restricted to emitter preparation for high injection efficiency. Pb-In alloys with sufficient In have been tried with success but have no particular advantage over pure In.

Counterelectrodes on p-type material employ the same basic materials except for Al. Thus, Au-Sb or Au-As are possibilities. The low melting elements, including In, can be doped to produce n^+ regions. Thus, binaries of Sn, Pb, and In with As and Sb and the ternary Sn, Pb, Sb have been used.

The same mechanical considerations noted for the counterelectrode alloy apply to the base. Electrically, the materials must form either a highly doped region of of the same conductivity type as the base or a metallic recombination contact. Base solder materials to form n^+ regions are the same as those discussed in connection with rectifying junctions on p-type material. One of the most commonly employed materials, especially in large units, is pure Sn which acts as a metallic recombination surface. For ohmic solders on p-type material the range of materials Au-Ga, In, PbIn, PbSnIn are possibilities and have been used.

b) *Silicon contact solders:* The increased fragility of silicon, the more difficult wetting problem, and the higher ambient processing and application temperatures make the choice of electrode materials for silicon rectifiers somewhat more difficult. The same higher melting elements as in the germanium case—Al, Au—are possibilities. In addition, Ag has been employed with success. Aluminum is used in most cases as counterelectrode on n-type material. The contact is not ideal in that the silicon and aluminum eutectic temperature is too high (577°C) and the product noted in condition 7) is too large. The silver-silicon alloy suffers the same deficiencies as Al (eutectic 830°C, brittle redeposit). The conditions in the silver-silicon alloy are bad enough to cause cracking of the silicon despite the presence of backing electrodes. The Au-Si alloy is also brittle but the eutectic temperature is low (370°C). Additions to the pure metals can be made to improve the situation. In lowering the solidus temperature of the Al-Si alloy, Mg and Ga would appear to be of interest (lowering the solid to about 450° and 425°C, respectively). Al-Mg alloys have about the same values of the product in condition 7); the Al-Ga alloys are soft. Sn-Al alloys have a phase that melts at 232°C which restricts processing somewhat. Both Pb and Sn-Al alloys present difficulties in vacuum fusion operations. Additions of Al, In, Ga, or Ge, to Ag will lower the solidus somewhat. The solid solutions with Ga or In may be expected to have lower bulk moduli. Small In or Ga additions to Au dissolve but phase diagrams are not available. The bulk moduli of the Au-Si and Ga or In ternaries should not be adversely affected. Small lead additions to Ag have been used to lower the solidus to the region of 500°C and soften the silicon alloy.

Of the metals mentioned Al, Al-Ga, Al-Sn produce p^+ regions on n material, Ag-Al, Au-Ga, Au-In do not seem to. Au and Ag-Pb alloys with Sb or As have been employed to make n^+ contacts on p material. Au-Sb alloy has been most widely used in commercial rectifiers.

Base contacts on p-type material can be made with Al and the Al-Ga and Al-Sn alloys. On n-type material the Au-Sb, Au-As, AgPbSb, AgPbAs alloys are possibilities, the Sb alloys being more convenient to prepare and use.

Except in the cases of smaller junctions, all the harder alloy solders must be thin foils a few mils thick to prevent undue stress in the silicon. Again in large area junctions the hard solders are usually backed up by metals matching the silicon in thermal expansion coefficient.

c) *Counterelectrode and base electrodes:* The counterelectrode material must satisfy several requirements.

1) It should have close to the same thermal coefficient as the semiconductor over the entire range of temperatures encountered during operation and processing.
2) It should have a bulk modulus which is high if the coefficient in requirement 1) matches well, and low if it does not.
3) It should have high thermal conductivity.
4) It should have fabrication possibilities and be reasonable in cost.
5) It should, if possible, be inert to the action of etches employed in processing the basic diode.
6) It should not contaminate the redeposited silicon.
7) It should be wet by the counterelectrode in the range of fusion temperatures employed.
8) It should be wet by solders used in attaching external leads.

Table I lists properties of a number of interesting electrode materials. Chromium and zirconium are not available in sheet form for general industrial use. The counterelectrode contact and the base contact in small rectifiers where the internal thermal resistance is not important need not have a good thermal conductivity and the Kovar and similar alloys have been used in both germanium and silicon with high-temperature solders. However, difficulty is experienced if an alloy freezes much above the inflection point in the thermal coefficient of expansion. Similar difficulties occur with some of the other Ni, Fe alloys.

TABLE I
PROPERTIES OF COUNTERELECTRODE MATERIALS

	Elastic Modulus, Psi×10⁶	Coefficient of Thermal Expansion Inch Per Inch C×10⁻⁶	Thermal Conductivity, Calories Per Second (cm²) (C Per cm)
Mo	50	5–6	0.35
Mo, Kovar-alloy-clad	50	5–6	0.35
Mo, Ni-clad	50	5–6	0.35
Kovar alloy	19.5	4 to 445 C >5 at high temperature	0.03
Zi	11	5	0.04
W	50	4.3	0.48
Ta	27	6.5	0.13
Invar alloy (36 per cent Ni, Fe)	21	2.7 to 200 C 11.5 to 200–315 C 15.5 to 315–425 C	0.025
Hypernik alloy (47–50 per cent Ni, Fe)	24	5.5 in small range	0.029
Cr	36	6.2	0.16

Of the elemental metals, tungsten best matches silicon, and molybdenum, germanium in thermal coefficient of expansion. Tungsten and molybdenum have been used with both semiconductors. Molybdenum contacts can be readily punched from sheets, and tungsten electrodes are most easily prepared by sintering techniques. Molybdenum can be easily welded or soldered to with soft solders when clad with Kovar alloy, plated with Ni or noble metals, or tinned with Ag, Sn, or In. When the material is used as counterelectrode material, the surface next to the counterelectrode solder is not clad or plated. Tantalum can be easily fabricated and is relatively inert to action of etches employed in processing. The element can be welded, but cannot be wet by soft solders unless clad or plated. It is expensive. The Kovar alloy and similar ones can be readily welded.

2) *Diffused Rectifiers:* The application of diffusion techniques in the production of rectifiers and transistors has grown rapidly in the last few years but attention has been focused on the silicon unit as far as rectifiers are concerned. A *p-n* junction is obtained by diffusing an appropriate impurity into the surface of the base material. In principal, any of the acceptors B, Al, Ga can be diffused into *n*-type material, and the donors P, As, Sb into *p*-type material, but in practice most units are made by diffusion of phosphorus and boron.

The theory is discussed elsewhere in this issue in a paper by Smits.[53]

Two different methods have been developed, one at static reduced pressure, the second at atmospheric pressure in a dynamic gas system. The second method is most widely used for rectifier preparation.

Furnaces with two controlled temperature zones are employed. A quartz furnace tube is connected at one end to a source of different gases, N_2, O_2, H_2, He, water vapor, or combinations. A suitable source of the diffusant, usually the oxide, is located in the zone at lower temperature. The vapors from the heated source are carried by the gases over wafers of silicon in the hot zone of the furnace. Under these open furnace conditions, a SiO_2 film forms on the silicon surface and prevents evaporation of the silicon. The surface concentrations N_s which are required for calculation of junction depths have been determined for a variety of carrier gases and source and silicon temperatures. The diffusion constants vary somewhat with the degree of surface disturbance of the crystal. Quite uniform diffusion results from a lapped surface. With uniform diffusion constants, junction depths still vary somewhat because of change of resistivity of the silicon crystals (especially those of high resistivity for high-voltage units) in the heat treatment at the high temperatures employed. Adequate provision must be made for such effects in rectifier processing. The lifetime of crystals usually degrades badly in the operation and considerable attention is being given to the temperature gradient during heating and cooling, to elimination of stresses on the crystals and the high temperatures, and to annealing procedures.

[53] F. Smits, "Formation of junction structures by solid-state diffusion," this issue, p. 1049.

After the diffusion operation, the oxide layers must be removed in HF before the surfaces of the blanks can be prepared for plating by lapping or chemical roughening.

The silicon surfaces are then plated with Au, Ni, or Cu by electroless techniques. Soft solders are employed in soldering the basic cells to heat sinks and electrodes.

The static reduced pressure method for carrying out the diffusion is not commonly employed for rectifier manufacture and will not be discussed.

III. Rectifier Preparation

A. Germanium Rectifiers

1) *Small Diodes:* Two types of small germanium diodes will be discussed which are analogs of silicon types treated later. The first of these is commonly called the "gold bonded" diode and is similar to the fuzed aluminum wire diode in silicon. Small n-type square dice about 50 mils on a side with thickness about 15 mils are used. These are fused to Dumet pedestals in appropriate jigs in N_2 atmosphere. The Dumet alloy is used because the units are glass encapsulated. An Au-Ga alloy wire a few mils thick is welded to another Dumet wire. The free end is welded to the surface of the germanium with a high current pulse. The glass capsule is sealed. Units made in such a manner have low power ratings (approximately 0.1 watt).

The second type of germanium rectifier is larger and is prepared in much the same manner as some of the large units. A small indium dot comprises the counterelectrode solder, tin the base solder. The latter solder is directly connected to a steel base. Much of the discussion of the larger units given next applies to the smaller device.

2) *Large Power Rectifiers:* The basic cells of the large area germanium rectifiers available are remarkably similar. Although a number of alternate structures are possible as discussed in Section II, the indium n-type germanium-tin system, giving a p^+nr or p^+nn^+ device, is widely used because of its simplicity and the satisfaction of all the requirements noted. This is not to say that very appreciable differences between basic cells do not exist, but rather to point up the extreme importance of the starting germanium materials and fine processing details.

High resistivity n-type germanium of resistivity 20 to 50 ohm-cm is employed. The crystal orientation may be (111) or (100) but the latter may be preferred. The material is zone refined to a high degree and doped to the desired resistivity in a horizontal furnace. The injected carrier lifetimes exceed 100 μsec and may run into the millisecond range. Lineage and aggregations of dislocations must be absent. There is an optimum range of etch pit densities. The crystal must be relatively free of dislocations for best results but if the number is too small difficulties are encountered in containing the indium counterelectrode solder. A value of about $10^3/cm^2$ etch pits represents a good compromise.

The crystal is sawed into slices of diameter up to about one inch. These are then etched or lapped and etched to the desired thickness of 15–30 mils in the common HF, HNO_3 solutions containing iodine, bromine, and/or acetic acid. Thin films of gold may be evaporated onto the surfaces to improve wetting.

Pure indium counterelectrode solder of thickness (3 to 60 mils) appropriate to the fusion process and germanium thickness employed is cleaned by physical, chemical, or electrolytic means. The base solder of Sn or Sn-Sb alloy is likewise carefully cleaned.

The above components are then assembled in appropriate jigs with or without bottom and/or top contacts mentioned previously. These are fused sometimes in inert gas with a preliminary outgassing in vacuum. Temperatures range from 550°C to 600°C. The units are then etched chemically in the mixed acid etches or electrolytically in NaOH solution. In cases, no post-etching at all is employed.

A variety of cases have been employed by different manufacturers. These may be filled with dry air or in cases with resins. Designs are available for both water and forced air cooling.

B. Silicon Rectifiers

Neglecting the very small junction diode types which employ Al junctions with whisker electrodes to the counterelectrode solders in modifications of process 2 below, a total of seven representative methods to yield units covering a broad range of forward and reverse current and voltage ratings will be discussed. Structurally similar devices will be treated in groups although specific dimensions of parts, silicon resistivities, and surface treatments may vary within the group to yield final units with different ratings.

Historically, one of the first small p-n junction rectifiers introduced was the so-called fused-wire diode. There are two distinct types; the first was designed for very low-current (forward current knee 1 volt drop at approximately 10^{-2} amperes) and for low-power operation.

1) *Twin Fused Wire:* The silicon employed is n-type in the form of dice approximately $0.1 \times 0.05 \times 0.05$ inch. These are etched in conventional acid etches. Ten mil wires of Al and Au-Sb alloy are cleaned and the ends are cut square. These are clamped to lead wires through glass or ceramic headers to give a spacing between wires of 0.040–0.050 inch. They are fused to the silicon die located on a heating element. The fusion occurs in inert atmosphere at about 800 or 900°C.

Basic cells are post-etched, immersed in silicone greases, and hermetically sealed in glass or metal cases.

2) *Single Fused Al Wire:* For higher current ratings (forward current knee 1 volt at 1 to 5×10^{-1} ampere), the design above was modified to decrease the forward resistance. There are a number of manufacturers presently producing this type. As in the instance of large power germanium rectifiers, the processes are quite similar.

The starting material is *n*-type silicon of resistivity appropriate to the voltage rating desired. The correlation between resistivity and breakdown voltage is usually quite poor. Ranges commonly employed are indicated in Table II. The silicon is sliced, diced, and etched to give dice in the range of thickness between 7 and 15 mils. Within a given process, the dice thickness variation is the order of ± 0.5 mil.

TABLE II

Resistivity	Voltage for Reasonable Yield
1 ohm-cm	>25 volts
10 ohm-cm	>200 volts
100 ohm-cm	>1000 volts

All types employ an aluminum wire for counterelectrode. The diameters vary from 10 to 20 mils depending upon the current ratings desired.

The base solder to the silicon is usually Au-Sb foil 1 to 3 mils thick which is employed with Mo or Kovar type alloys. If Au foil is not used, the silicon must be plated to permit subsequent attachment to a base contact. Electroless nickel plating procedures have been used.

The preparation of the basic cell with leads attached is carried out in one- or two-step processes:

One-step

Process 1—Simultaneous fusion of Al wire to silicon and of the latter to the base support of Kovar alloy with Au-Sb alloy.

Two-step

Process 2—Prefusion of silicon with Au-Sb alloy to base support of Mo (clad or plated).
Process 3—Fusion of Al wire to silicon, soft solder of plated silicon to base support.

The base support can be a member which is subsequently soft soldered to a case base or can constitute the case base. The processes have different advantages and disadvantages, some of which are tied up with machine design and production capacities. Processes 1 and 3 take place in graphite boats in continuous belt furnaces. Inert gases N_2, A or N_2, H_2 mixtures are employed. Process 2 is carried out piece by piece in a wire fusion jig. In all cases, the temperature and temperature gradient must be carefully regulated so that the silicon surface is somewhat over 600°C.

The aluminum-silicon junctions are in some processes etched, in others not. The units are then welded or soldered into a variety of cases.

All these types have a common problem at the aluminum-silicon junction: an appreciable thickness of aluminum is in contact with silicon. The expansion coefficients do not match which may lead to cracking of the silicon.

3) *Tin-Aluminum Alloy Types:* The silicon type and resistivity and its processing is similar in all cases. However, the choice of counterelectrode and base solders as well as of lead attachment methods vary.

As noted in Section II, a tin-aluminum alloy can be chosen to provide *p*-type doping. This has been done by some manufacturers. The resulting alloy is soft compared with the aluminum-silicon eutectic. Solder penetration on fusion is decreased considerably. The base solder is Sn which is used to solder the wafer to a steel support or directly to steel or copper studs.

4) *Inert Nail Type:* To retain processing advantages of a single step fusion for both base and counterelectrode contacts and to avoid the difficulties encountered with the aluminum wire junction, one producer has employed a structure with inert Ta nail for the counterelectrode contact and aluminum foil for the solder. The thickness of the aluminum is so reduced compared with that of the nail of low expansion coefficient that the average coefficient of the system is near that of silicon and excessive stress is avoided. The base solder is an AgPbSb alloy. The base support is a metal of low thermal expansion coefficient (Mo) which may constitute the case base as in process 1 in Section III-B, 2. The SiAgPbSb alloy that results at the base after fusion has a lower bulk modulus than the Si-Au-Sb alloy and was chosen to further reduce stress in the silicon. Encapsulation of the unit is carried out in a series of welding operations.

5) *Silicon Sandwich Type:* Another unique method of reducing stress in silicon has been employed in producing a line of rectifiers. This uses a support wafer of silicon to reduce the stress in the element. Thus, a sandwich of silicon results with relatively unstrained material near the junction.

6) *Diffused Types:* The diffused rectifier process is such that the basic steps are carried out before the area of the junction need be determined. Consequently, all types from the very low-current, high-voltage (100 ma, 1500 volts) to the high current ones (100 amperes) can be discussed at once.

Although a number of possible methods exist, most presently available units are produced by the one to be reviewed. An n^+pp^+ structure results. *P*-type silicon of resistivity range similar to that given for the fused-wire rectifier is employed. Slices are cut and lapped to a thickness depending upon the inverse voltage to be supported but generally to about the same value or perhaps a mil or two smaller than in the fused cases. After cleaning, the slices are exposed to vapors from a P_2O_5 source held at a temperature about 200°C. The vapors are transported in an oxygen carrier past silicon slices located in the hot zone (approximately 1200°C) of a two-zone furnace. Other carrier gas combinations can be employed in different sequences. The objective is to obtain a uniform diffusion of phosphorus through a phosphosilicate glass formed on the silicon. This prediffusion is continued for about one-half hour. The P_2O_5 source is then removed and diffusion in the silicon continued for several hours.

The phosphorus is then lapped from one side of the slices. These are loaded into another diffusion furnace in which vapors from boric acid held at about 250°C are carried in oxygen past the silicon slices in a hot zone at 1200–1300°C. Again, diffusion is carried out for several hours.

A new wrinkle in the diffusion of the two elements to form the same basic structure has been introduced.[54] In this method, which has been employed to produce rectifiers with 5–10 ampere ratings[55] in a very simple and inexpensive manner, P_2O_5 and a boron compound, each in an organic vehicle, are painted on opposite sides of a silicon wafer. Wafers are then stacked back-to-back and diffused in an open tube.

After removing the oxide coats in HF, a matte surface is achieved by lapping or chemical treatment. An electroless nickel or gold plate can then be achieved. The wafers are diced to size, pretinned with soft solder, mounted on base pedestals, and etched in conventional etch solutions.

A variety of case structures has been conceived which minimize to some extent the masking, etching, and lead attachment problems associated with the handling of basic silicon cells not having refractory metal contacts.

It should be noted that small units may be soldered directly to copper tabs since soft solders are used. Coatings of silicon varnish are applied and cured prior to final case welding or soldering. Diffused rectifiers are being presently manufactured in current ranges to several amperes.

7) *Large Fused Types:* In all available types of large rectifiers, the silicon is supported by a base contact of Mo. In the large sizes and under the extreme conditions which such types are placed, the basic problem of strain in the silicon must be carefully considered. In most cases, the counterelectrode is likewise supported by a Mo plate.

The basic cells then consist of silicon sandwiched between two Mo members. The silicon employed ranges in size up to $\frac{5}{8}$-inch diameter and has a thickness of 8 to 10 mils. The parts can be fused together with advantage using the hard solder systems. For the counterelectrode in the p^+nn^+ unit aluminum is used, for the base solders Au-Sb or Ag-Pb-Sb. The n^+pp^+ type uses an aluminum solder to a molybdenum contact but no counterelectrode support member. Foils are in the range 1 to 3 mils thick. Fusions take place in graphite jigs, in vacuum, at temperatures close to 900°C with the silver alloy, 800°C with the gold alloy. Again, units are etched in the usual solutions. The basic cells are then soft soldered to copper pedestals. A series of welding or soldering operations completes the fabrication.

[54] K. D. Smith, "Economical diffused junction silicon varistors," presented at PGED Conference, Washington, D. C., October 26, 1956.

[55] H. E. Hughes, J. H. Wiley, and P. Zuk, "Silicon diode design characteristics and aging data," 1957 IRE WESCON Convention Record, pp. 80–89.

IV. Rectifier Properties

A. Appearance

Although the number of basic cells is quite limited there are a large variety of rectifiers despite efforts at standardization. The general appearances of the Ge cases are similar to those of the small silicon types. The characteristics of the germanium units are coded Ge_s in subsequent illustrations.

Large germanium rectifiers are pictured in Fig. 1. The range of air and water cooled applications give greater latitude in case and cooling structure design. A code of Ge_l is used in the illustrations.

The appearances of the smaller and medium sized silicon rectifiers are familiar (coded Si_s and Si_m in subsequent figures). Larger units (Si_l) are assembled in Fig. 2.

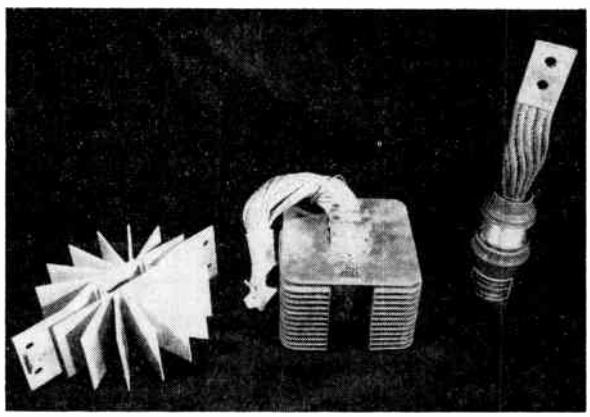


Fig. 1—Large germanium rectifier units.

Fig. 2—Large power silicon rectifier units.

B. Static Characteristics

A valid quantitative comparison of theoretical and experimental rectifier characteristics requires detailed information which is not usually obtained by the manufacturer for commercial purposes. In those cases in which detailed information has been obtained there is some question that the devices investigated were identical or closely similar to those discussed. Consequently, it is necessary to review specific results that have been obtained with special contacts in evaluating the theories. Only qualitative statements can then be made regarding the application of these results to the general run of different commercial devices. Of course the static characteristics *per se* with or without explanation are of first interest to users.

1) *Reverse Characteristics:* Pell[37] has examined the reverse current at low voltage of a variety of small fused and small and large grown germanium contacts and, with care to separate bulk and surface effects, has demonstrated that two components of bulk current are present: one follows the usual diode diffusion theory, and the other arises from charge generation in the space charge region, at traps about 0.3 ev from one of the bands. The absolute magnitude of the components were compared with results obtained from independent measurements of intrinsic resistivity and reasonable agreement was obtained. The diffusion component dominated at temperatures above $-20°C$. The nature of the traps could not be determined but the level was not inconsistent with those of Fe and Cu. Many investigators working at room temperature and above have found good agreement between the theory for the diffusion component and experiment in small rectifiers.

At higher reverse voltages, the agreement of characteristics of small germanium diodes has again been good using junction lifetime parameters measured on the diodes under test. The correct dependence of saturation current on resistivity has been noted. The avalanche breakdown mechanism for contacts with resistivities employed in the rectifiers being discussed has been reasonably well substantiated. However, the values of breakdown fields have been determined experimentally.

In any particular small rectifier, reverse current may exceed the theoretical value (again using the rectifier determined junction lifetime). In many cases, this can be traced to the existence of surface conduction channels.

In larger germanium rectifiers the agreement of theory and experiment is usually quite poor. A saturation current may be observed for only a small voltage range or not at all. Breakdown voltages fall far short of those predicted from breakdown fields in small units. The difficulties appear to be known although the ultimate in reverse characteristics has not yet been obtained. The problems lie in obtaining large area crystal slices free of lineage, aggregations, dislocations, and, of course, grain boundaries or cracks, and in achieving uniform penetration of the counterelectrode alloys. When unwet areas exist under the electrode alloy, large components of surface leakage at these spots exist. Post-etching procedures noted in Section III can only improve the exposed junction periphery.

The reverse currents in small etched silicon junctions are much smaller than in germanium, being in the range 10^{-5} to 10^{-7} amps/cm^2. If the reverse current were diffusion controlled, the current would be close to 10^{-10} amps/cm^2.

Sah, Noyce, and Shockley have recently treated the problem with care to show surface effects to be negligible in the freshly etched junctions used. Good agreement of theoretical calculations and experiments were obtained on the basis of charge generation from traps very near the center of the band.

The small silicon rectifiers of all types exhibit wide ranges of reverse current densities regardless of extreme care taken to avoid surface leakage currents. The magnitude of the observed range of currents and this variation is readily understood in terms of the dominant nature of current from charge generation centers deep within the forbidden band. Thus, the characteristics are extremely sensitive to the silicon material used. In unetched rectifiers and those with unclean surfaces, the reverse currents are higher because of the surface components. The nature of all the surface contributions to excess leakage has not been determined. In fused rectifiers, the edge strained crystal structure, surface oxide films, and ionic impurities all contribute. With further examination of bulk trapping levels and separation of the body effects, the quantitative analysis of the practical rectifier may be looked for in the near future. The characteristics of small diffused junctions exhibit the same range of currents.

The reverse breakdown in silicon is again of the avalanche type. Breakdown fields have been determined from the study of small junctions. These breakdowns represent upper limits to be expected. In practical small rectifiers, the breakdowns obtained are safely a factor of about twenty times the resistivity in the range of voltage below 1000 volts, a factor of 10 or less at higher voltage. Again, the perfection of the silicon is all-important in achieving the highest voltages. Both fused and diffused rectifiers have been made in reasonable yields with voltages over 1500 volts. However, the ultimate that is possible, according to current theories, has not been achieved.

The reverse current densities of large silicon rectifiers of both the fused and diffused types have been higher than in the smaller types in most cases. The reverse current densities range from 10^{-5} to 10^{-1} amps/cm^2 (in the operating voltage range 50 to 400 volts) depending upon the details of processing and the quality of silicon employed. The currents generally increase faster with voltage than in the small rectifier cases. With ultrapure silicon, the largest size rectifiers can be produced with good yield of units with voltage over 1000 volts and low leakages. As in the germanium large area rectifier, the limiting factors are the degree of crystal perfection and the uniformity of alloy fusion or doping impurity diffusion.

Fig. 3 presents the reverse characteristics per unit area of a collection of silicon and germanium rectifiers together with the characteristics of the junctions that were examined in comparing theory and experiments noted above.

2) *Forward Characteristics:* The forward characteristics of germanium follow reasonably well the exp (ev/kT) dependence at low voltage, the exp $(qv/2kT)$ low at higher voltages. Diffusion currents control in all ranges of temperatures in practical applications. At higher currents, the field drops in the fully conductivity-modulated base semiconductor determine the charac-

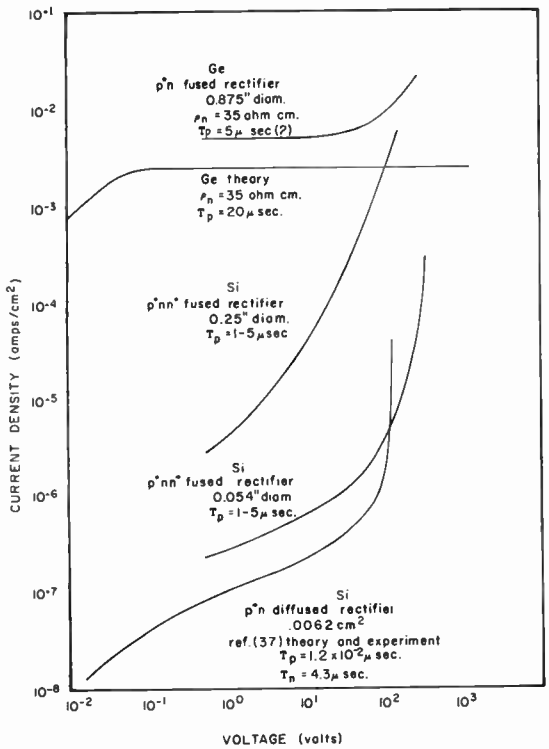


Fig. 3—Reverse characteristics per unit area of representative germanium and silicon rectifiers.

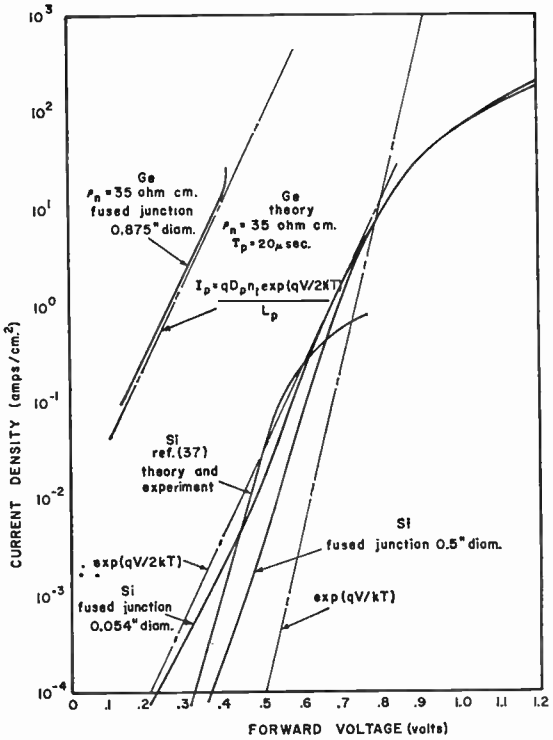


Fig. 4—Forward characteristics per unit area of representative germanium and silicon rectifiers.

teristics. Typical forward current densities in the commercially available units are given in Fig. 4.

The forward characteristics of most silicon rectifiers, large and small, exhibit a region with exp (qv/kT) dependence at very low voltages which alters to an exp ($qv/2kT$) for a range of higher voltages. At very high forward voltages, the rectifiers reach constant resistance regions. Sah, *et al.*, have also treated the forward characteristic and attributed portions of the characteristic between the very low-level region and a high-level diffusion current region (that of bulk conductivity modulation) to recombination effects. It is somewhat difficult to separate different regions of the characteristics since there is always a qv/kT region at low levels and the $qv/2kT$ region must be passed through in reaching the constant resistance region. The analysis of small rectifiers by Fletcher noted in the section on theory considers the importance of semiconductor thickness and rectifier structure in determining the exact form of the forward characteristic. The thickness ranges noted in Section III were chosen to give lowest possible forward resistances without obtaining an appreciable number of shorts. The very lowest forward drops seem to occur in p-type fused and thin diffused units. Fig. 4 compares forward current densities in practical germanium and silicon rectifiers with values obtained in junctions that have been employed in testing the theories of Section II.

3) *Temperature Dependence of Reverse Leakage:* For the most part, the characteristics of well etched germanium devices follow the exp (E_G/kT) law required for diffusion currents.

The temperature dependence of reverse leakage of some well etched silicon diodes follow reasonably well an exp ($E_G/2kT$) relation. Other rectifiers exhibit a variety of dependences. In some cases, the leakage even decreases with temperature. Such behaviors are attributed to surface leakages since they occur for the most part in poorly etched or unetched units. The subject of the temperature dependence of rectifier properties has not been completely explored and considerable additional work is required on the subject.

C. Typical Ratings

The subject of ratings of rectifiers is a very large one and a detailed treatment of the subject would be out of order in the present paper. Therefore, the general principles will be briefly reviewed, then a short summary of the range of current, voltage, and power handling capacities of single rectifier units will be given. Most of the previous discussion of theories, processes, and devices has concentrated on units which handle currents in the 0.2- to 200-ampere range and above. Such devices are employed in applications where first attention has not been on switching times (although this lack of attention has caused difficulties in some applications). Therefore, little will be mentioned on switching periods and switching transients. Again, reference is made to the literature[15] for the treatment of small diodes. In larger devices, much more work is needed on the subject.

The ratings of a rectifier are determined basically from the condition that the temperature of the semi-

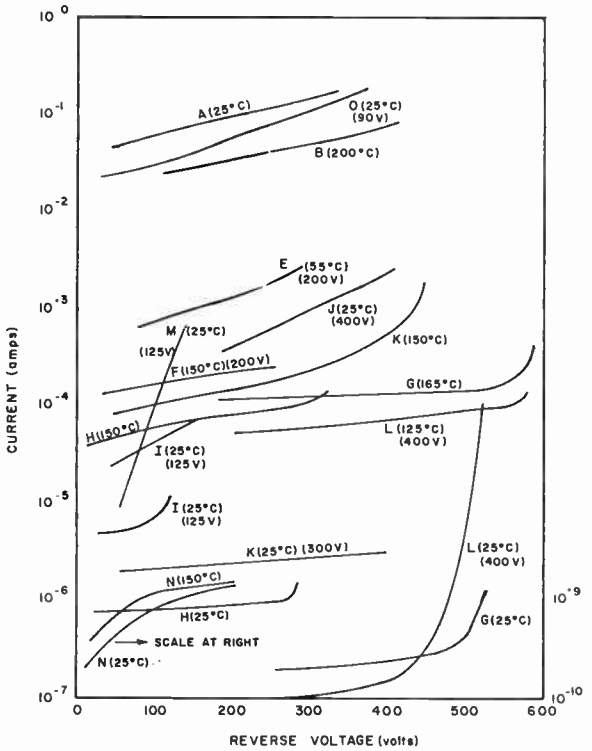


Fig. 5—Reverse characteristics of complete range of rectifiers.

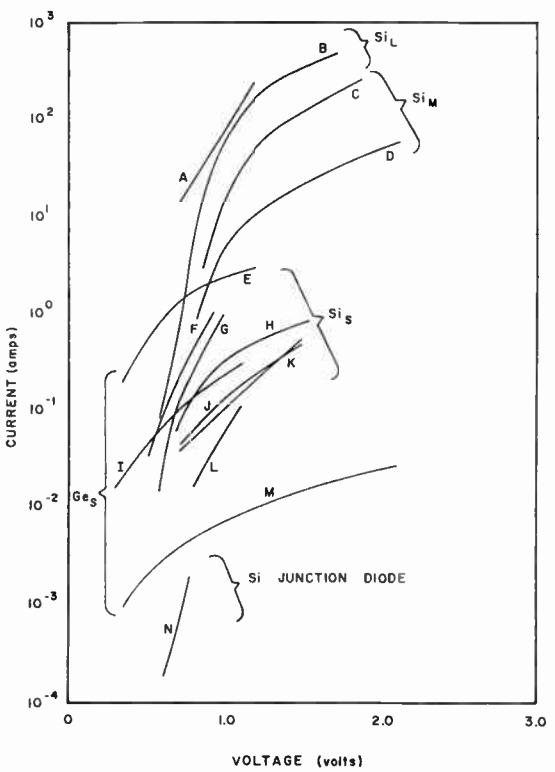


Fig. 6—Forward characteristics of complete range of rectifiers.

conductor junction does not exceed a steady-state value in which the heat dissipated in both forward and reverse directions equals that removed to the heat sink.

The reverse power dissipation rises exponentially with temperature at a given reverse voltage. On the other hand, the forward dissipation at a given current drops only slightly with temperature principally as a result of the temperature dependence of the internal contact potential.

The power removed to the heat sink consists of radiation convection and conduction losses. In the temperature range under consideration, effective radiation film coefficients can be used and the combined radiation and liquid or air convection losses per unit area can be expressed as a linear function of the local temperature drop to the ambient and the combined film coefficient.

In present practice, the maximum steady-state junction temperature permitted in germanium is about 90°C; in silicon it is from 180 to 200°C. At times higher junction temperatures have been employed, to possibly 110°C in germanium and 250°C or so in silicon, but these values are unusual.

In determining reverse ratings, an additional consideration arises: the rated voltage must be safely above the value of any transient encountered in the particular kinds of applications for which the units are designed. Fig. 3 shows that the reverse characteristic may have so-called hard or soft knees, that is, the breakdown voltage may be sharply or poorly defined.

The Ge and Si basic rectifying cells are so small that the film coefficients of the cases shown in Fig. 2 almost completely control the thermal resistances.

Small germanium rectifiers and all silicon rectifiers are air cooled. The larger germanium units are either air or water cooled. In most instances, the largest contribution to the power loss comes from the forward conduction cycle.

Consequently, it is seen that the current rating of most rectifiers involves equal considerations of the forward drop of the cell and of heat transfer from finned structures. Factors entering into the forward drop have been treated in Sections II and III. There is nothing unique about the heat transfer problem apart from the fact that rather small sources are under discussion. In such cases, care should be taken to employ the cylindrical functions in accounting for temperature drops within the cooling fins.[56]

1) *Computer Junction Diodes:* Figs. 5 and 6 code N present reverse and forward characteristics, respectively, of the smallest type of fused junction diode. This is one representative of a group of small silicon diodes that spans a voltage range from 15 to about 300 volts. The current carried in the forward direction with 1-volt drop varies from about 1 to 40 ma in different types. Reverse leakage at room temperature spans the range 0.001 to 0.5 μa; at 150°C, the range 0.5 to about 50 μa. The current handling capacity goes from 30 to about 90 ma at 150°C. These units are employed in computer and higher frequency applications where the switching transient is of utmost concern. The re-

[56] Henkels, *op. cit.*, Appendix III.

covery period of the diodes from a condition of near rated forward current to a condition of a 400 K ohm resistance in the reverse direction is of the order 1–10 μsec. Collector capacitance is 10 to 40 μμf. Higher conductance units have been designed which raise the forward current to 200 ma at 1 volt.

2) *Small Germanium and Silicon Rectifiers:* The range of the gold bonded and small button alloyed germanium types overlaps that of the variety of small silicon rectifiers. Thus, codes I and E, the germanium units, and codes F, G, H, J, K, L which represent a variety of techniques for producing the small silicon rectifiers span a current range 10^{-2} to 1 at 0.75-volt forward drop (Fig. 6). The peak inverse voltage of the germanium units is limited to about 300 volts. The leakage from Fig. 5 is 0.05 ma for the smaller, about 1 ma for the large germanium unit at the rated voltage and 50°C. The silicon units have peak inverse voltage ratings up to 1000 volts. The reverse leakages are relatively constant with voltage ranges from 0.07 to 0.5 ma at rated voltages and 150°C. In some instances, upper reverse current limits are somewhat higher, up to 1–3 ma (depending on the junction size and designed steady forward current and surge current ratings).

The dc current ratings of the germanium unit E may be as high as 150 ma at 55°C. The equivalent silicon units have ratings from below 0.1 to about 0.5 ampere at 150°C. A one-finned structure (approximately $2\frac{1}{4}$ square inches) increases the germanium rating to 0.5 ampere, and similar provisions for the stud mounted units in silicon raise the capacity to 1 or 2 amperes.

Current surge capacity varies greatly from 1 to 30 amperes.

There has been some difficulty in the life test of many of the silicon devices in this range.[57] It appears that many may be marginally rated.

3) *Intermediate Range Silicon Rectifiers:* Devices for the current range from 3 to 30 amperes exhibit forward currents at 1 volt that fall in the decade above those of the rectifiers just discussed (codes C and D in Fig. 6). The reverse currents at rated voltages, which range to 400 volts, may be as high as 50 ma at 150°C. All are mounted on fins of some kind in operation (sizes from 4 inches^2 to 25 inches^2).

Surge ratings up to a few hundred amperes in certain cases can be handled for seconds. These larger devices which, as seen from the discussion of Section III, have different, more stable structures than most of the small rectifiers have, for the most part, performed in an excellent manner. Life test results have been good.

4) *Large Germanium and Silicon Rectifiers:* The very large germanium rectifiers have forward drops as low as 0.55 volt at 600 amperes (25°C). The silicon units have forward currents to 100 or 200 amperes at 1-volt drop. The breakdown voltages of well made germanium units may be as high as 500 or 600 volts but working ratings below 100 volts are common. In silicon large area devices, breakdowns consistently over 1000 volts have been produced. However, most ratings are restricted to 300 or 400 volts. Current ratings generally do not exceed 250 amperes per cell in either germanium or silicon with ordinary heat dissipation schemes. Surge ratings of both germanium and silicon are high (1000–2000 amperes for 1 second). Silicon rectifiers are always provided with fins for forced air cooling in these ratings. Germanium is used in both water and forced air cooled installations.

[57] W. F. Bechtold and C. L. Hanks, "Failure rate studies on silicon rectifiers," AIEE Tech Paper No. 57-795; June, 1957.

V. Selected Applications

Referring again to Figs. 5 and 6, it is seen that a span of just about six decades of current is handled in the range of devices that have been considered. All of these devices are commercially available and some have been sold for a number of years. Yet, all the phases of research, development and production have taken place within the past ten years. The number of applications for this range of devices is, of course, extremely large, covering all the fields of electronics and electrical power equipments. A few applications will be noted below by way of illustration.

A. Computer Diodes

The smallest junction diodes are employed in a variety of radio and computer applications. As a video detector, the semiconductor device prevents the introduction of hum, screening is made simple, and the high efficiency is achieved. Diodes have also been employed in dc restorers and automatic gain control circuits. In the ratio detector in TV receivers with an FM sound channel, two diodes can be used in the discrimination circuit. In computers, flip-flop circuits and clamping and gating circuits, use has been made of the units.

B. TV Diodes

One of the volume uses for the small rectifiers with ratings of 0.5 ampere is in the B^+ supply in television receivers. The silicon and germanium rectifiers for the purpose are very compact, have excellent efficiencies, and retain these efficiencies during operation (in other words, the units do not age). Turn-on capacitor surge problems caused trouble in initial applications but proper design has eliminated the problems. The devices are widely used and substantial quantities are in the field.

C. Small Power Supplies

The devices with the higher ratings in the group marked Si_s and Ge_s and those in the group Si_m are assembled in a variety of single- and three-phase power supplies that requires a sizeable volume to list. Fig. 7 shows a few assemblies of a number of different manufacturers. The smaller bridges are free connection, the larger, forced air cooled.

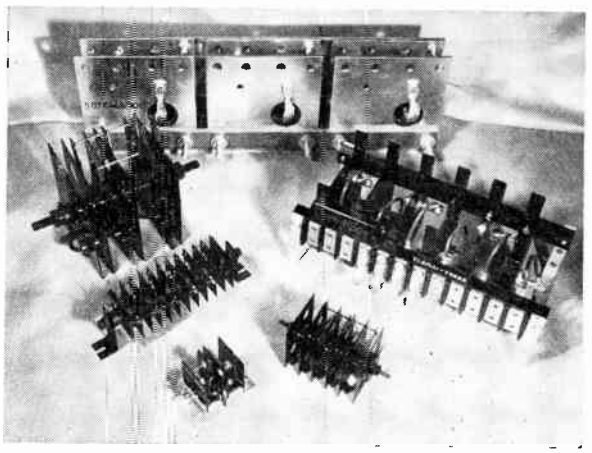


Fig. 7—Assorted rectifier bridges.

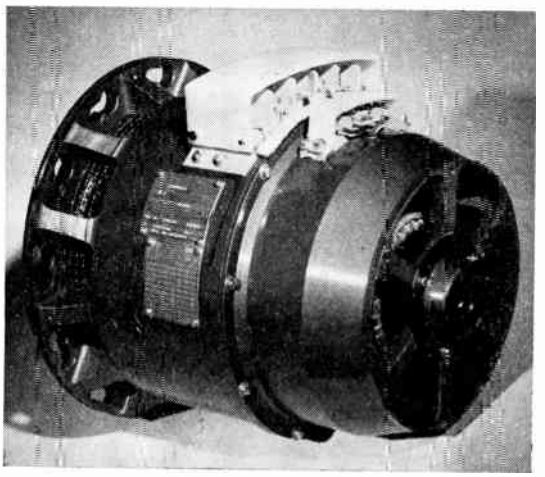


Fig. 8—40-kva brushless air cooled alternator.

D. Rotating Rectifier

One of the first applications of the rectifiers with ratings noted Si_m was to the development of a brushless alternator for high-altitude aircraft.[58] Both oil cooled and air cooled designs were produced. The use of silicon rectifiers in this application illustrates their operation under the most extreme environmental conditions. In addition to all the usual salt spray, humidity, and shock requirements of the military specifications, the rectifiers were required to withstand centrifugal forces up to 15,000 g and oil pressures of 400 pounds per square inch, both at ambient temperatures from −55 to 180°C. Cells were required to withstand 175 volts in the reverse direction and 40 amperes in the forward. The brushless generator incorporating the air cooled design to deliver 40 kva is shown in Fig. 8.

E. Battery Chargers

Fast rate battery charger manufacturers have also employed silicon rectifiers because of their ruggedness and space savings features. Six, twelve, and twenty-four volt chargers are available with ratings of a few amperes to over 100 amperes. Other features include long life, the ability of the cells to withstant short circuits, and the absence of any inflammable parts.

F. Utility Rectifier

The hermetically sealed features of silicon and germanium rectifiers have made them adaptable to many oil cooled applications, the largest of which is a 240-kw silicon rectifier designed to meet the conversion requirements of utility companies.

The silicon rectifying cells are oil cooled within the submersible welded construction. Balancing reactors are used internally to provide proper division of current among parallel cells. Fuses are mounted externally in an air-tight compartment. Space savings and high efficiency combined with no moving parts are added features which have proven the advantages of the new rectifiers in this utility application.

G. Germanium Electrochemical Installation

The majority of applications now installed are for electrochemical processes requiring comparatively low voltage, high current, and high efficiency.[59] The production of hydrogen, chlorine, and other gases involves operation at 65 volts dc; electroplating and anodizing use 12 to 48 volts dc at high current, and arc reduction furnaces require 80 volts dc open circuit. Efficiencies are high, 94 per cent at 37.5 volts, 92 per cent at 24 or 48 volts, and 90 per cent at 12 volts. The size and design of the individual cubicles depends upon the particular requirements, and may include the germanium stack assemblies, the transformer, control reactor, ac switchgear, fuse monitoring system, and heat exchanger type cooling system. For larger power requirements, these cubicles have been connected in parallel.

[58] H. W. Henkels, "Silicon rectifiers in brushless machinery," *Commun. and Electronics*, vol. 33, pp. 620–626; November, 1957.

[59] L. W. Burton, "Germanium rectifiers for industrial applications," AIEE Tech. Paper No. 56-78.

The Potential of Semiconductor Diodes in High-Frequency Communications*

A. UHLIR, JR.†, ASSOCIATE MEMBER, IRE

Summary—Graded *p-n* junctions that can be fabricated by solid-state diffusion are low-loss nonlinear capacitors at microwave frequencies. These diodes can be used to make low-noise amplifiers, amplifying frequency converters, harmonic and subharmonic generators, switches, limiters, and voltage-tuned passive circuits. Single junctions can control many watts of microwave power.

Point-contact diodes are nonlinear resistors and as yet are unchallenged as microwave rectifiers. At lower frequencies, nonlinear-resistance action can be obtained in *p-n* junctions by introducing recombination centers.

A *p-i-n* diode is resistive at high frequencies. The value of the resistance depends upon the dc current. This variable resistance can be used as a broad-band microwave switch or attenuator. At low current densities, the *p-i-n* structure functions as a transmission line and so can serve as a support, protection, and connection for small-area *p-n* junctions made in the same single crystal of silicon.

I. Introduction

THE need for microwave crystal diodes stimulated early work on germanium and silicon. In particular, germanium of fair purity resulted and was an essential ingredient in the discovery of the transistor. Concepts created in the work on microwave diodes showed how to use the periodic table in further development of transistor materials.

It is now time for transistor technology to repay its debt to the microwave art. Single-crystal germanium and silicon are now available. New methods of making *p-n* junctions have been developed, such as solid-state diffusion of impurities. The *p-n* junction concept has been laid down and subjected to extensive (though by no means exhaustive) analysis. How can these developments be put to microwave use?

One answer to this question is that the *p-n* junction itself will be a valuable microwave circuit element. Its most spectacular use will probably be in low-noise microwave amplifiers. This and other uses of junction diodes are discussed. Also, an attempt is made to explain why junction diodes have not replaced point-contact diodes as microwave rectifiers and why the latter have not been greatly improved by transistor technology.

A major theme of this paper is that diodes differ qualitatively in their electrical characteristics. For a given circuit function, a particular type of electrical characteristic will generally be optimum. An introduction to three diode types will be given in Section II. Two types, the nonlinear capacitor and nonlinear resistor, can be described by simple equivalent circuits with frequency-independent elements; structures and physical mechanisms will be considered in Sections IV and V. Another type, the *p-i-n* structure discussed in Section VI is in a certain sense a variable resistor. Various circuit uses of diodes will be considered to determine which type of diode is most suitable. Rectification is discussed in Section VII. Frequency converters (Section VIII) may use nonlinear resistor or nonlinear capacitors; if the latter, amplification is possible. The control of microwave power by diode switches is discussed in Section IX. Sections X and XI deal with harmonic and subharmonic generation. The use of the nonlinear capacitor as a passive, electronically-variable capacitor is considered in Section XII.

* Original manuscript received by the IRE, April 12, 1958. Based in part on investigations supported by the U. S. Army Signal Corps on contract DA 36-039 sc-73224.

† Bell Telephone Labs., Inc., Murray Hill, N. J.

II. Classification of Diodes

What is here called a nonlinear resistor is the ordinary conception of a rectifier, a diode that can convert ac power into dc power. The argument of the entire paper can be anticipated by suggesting that a rectifier may not be the right type of diode to use in frequency converters, switches, harmonic generators, and other circuits where rectification is not a specifically desired function.

For an ideal nonlinear resistor, the instantaneous current $i(t)$ is a function only of the instantaneous voltage $v(t)$. Practically, a device might be called a nonlinear resistor if, for the intended frequency range, its action can be analyzed approximately by assuming $i(t) = f(v(t))$. The anticipated relation between current and voltage is the familiar sort shown in Fig. 1. Generally, the polarity of the applied voltage makes a vast difference in the current. The polarity that gives a large current is called forward bias, the opposite, reverse bias.

An analogous description can be made of the nonlinear capacitor. The operation of such a device can be analyzed approximately by assuming that the instantaneous charge on the device is a single-valued nonlinear function of the instantaneous voltage applied to each terminal. The charge-voltage characteristic for a *p-n* junction is indicated in Fig. 2. The slope dQ/dV of this curve is the small-signal capacitance, which, like the charge Q, may be regarded as a function of the instantaneous voltage.

A time-varying capacitance can amplify. This principle has been utilized for a long time in vibrating-reed (or "dynamic capacitor") electrometers for amplifying low-frequency voltages from very high impedance sources. In Section VIII it is shown that analogous ef-

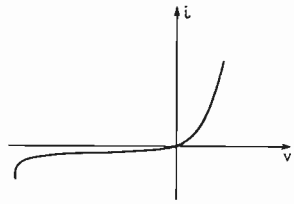


Fig. 1—Typical current-voltage characteristic of nonlinear resistor.

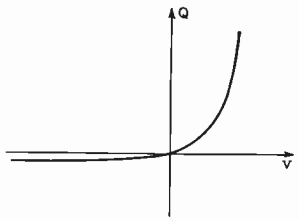


Fig. 2—Typical charge-voltage characteristic of nonlinear capacitor.

fects can be obtained at microwave frequencies by applying a time-varying voltage to a nonlinear capacitor, thus producing a time-varying capacitance. The direct analog of the vibrating-reed electrometer is the amplifying up-converter. Negative resistance amplification can also be obtained from time-varying capacitance and is the basis for a low-noise 6000-mc amplifier that is mentioned in Section VIII.

Amplification can similarly be obtained from nonlinear inductance. The name "varactor" has been proposed for any device whose operating principle is nonlinear reactance.[1] If this name gains currency, one would speak of nonlinear capacitor amplifiers, as well as amplifiers using saturable reactors, as varactor amplifiers. At present, the name "parametric amplifier" is often used. "Reactance amplifier" seems to be a much more descriptive term.

Another class of diodes seems to deserve the name "variable resistor." (It is regrettable that this term has been previously used as a synonym for nonlinear resistor.[2]) The structure is a layer of high-resistivity semiconductor with heavily-doped *p* and *n* regions on either side. A symbol for this structure is *p-i-n*. The shunt capacitance per unit area is remarkably small and the impedance at high frequencies is essentially resistive. The value of the resistance can be varied over a large range, being very high for reverse dc biases and very low when current flows through the diode in the forward direction. However, one cannot vary the resistance at a microwave rate, as is possible with certain nonlinear resistors (point-contact microwave diodes).

III. Impedance Measurements

The diode classifications outlined above define diodes by their two-terminal lumped-element characteristics. At high frequencies, ordinary components and transmission lines have dimensions that are appreciable compared to the wavelength, so that lumped-element concepts cannot be applied to them. However, the active regions of semiconductor diodes are generally small enough so that a lumped-element definition is tenable.

A suggestion by Waltz[3] makes possible microwave measurements of the actual junction or contact impedance of semiconductor diodes. This technique has been especially helpful in the development of *p-n* junction nonlinear capacitors, and this problem is solved in the following. The diode is mounted in some kind of a waveguide holder or crystal mount for coaxial line. The transmission line is connected to some impedance-measuring device adapted to transmission lines, such as a slotted line. The objective is to find the impedance of the junction itself. To do this, one must find the parameters that define the electrical transformation from the junction to the transmission line. The basis for the measurement is to make impedance *standards* that, like the junction, are small compared to the wavelength. They must be placed in exactly the same configuration as the *p-n* junction or point contact. The impedance standards may be, for example, small area pressure contacts to carbon of the type used in carbon resistors or pencil leads. (In evaluating gold-bonded germanium diodes, satisfactory "standard resistors" were made by bonding gold-gallium wire to *p* type germanium.) Two singular impedance standards are readily constructed, an open circuit and a short circuit.

The analysis of the data is not described here, except to note that it is particularly simple if the transformation between the diode and the transmission line can be regarded as lossless. Then the transformation at a single frequency may be represented by a length of line, a series reactance (or shunt susceptance), and an impedance transformation. The open circuit and short circuit determine length of line, and the product of the series reactance and the transformation ratio, so that relative junction impedances can be obtained without using a standard resistor.

If the diode is fabricated in some kind of a cartridge, one must use identical parts to fabricate the standard resistors. Care must be taken to insure that these resistors are inserted in the crystal mount in exactly the same way as the diode to be measured. The frequency limit of this technique depends upon the pains taken to insure equivalence of the electrical transformation when measurements are made on the standards and on the unknown. Dimensional tolerances should be held to within a few thousandths of a wavelength. Little difficulty is encountered in doing this at frequencies below 1000 mc.

IV. Nonlinear Capacitor Structures

The mechanism of *p-n* junction capacitance is described qualitatively, with mention of effects that can

[1] M. E. Hines, paper to be published.

[2] A. Uhlir, Jr., "Two-terminal *p-n* junction devices for frequency conversion and computation," Proc. IRE, vol. 44, pp. 1183–1191; September, 1956.

[3] M. C. Waltz, "A Microwave Resistor for Calibration Purposes," Bell Telephone Labs., Third Interim Rep. on Task 8 (Crystal Rectifiers), Signal Corps' Contract DA-36-039-sc-5589; April 15, 1955.

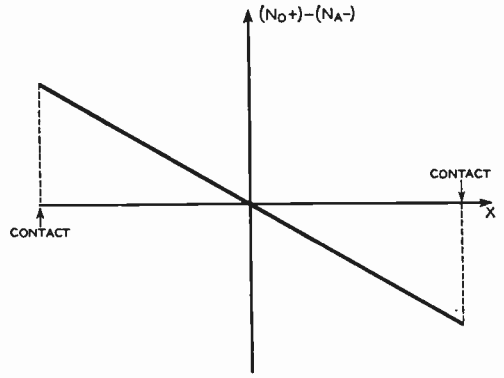


Fig. 3—Impurity distribution in linearly graded junction.

cause losses and lead to residual noise. Measurements on diffused silicon and welded-contact (gold-bonded) germanium diodes are given.

Generally speaking, semiconductor materials suitable for transistor fabrication exceed by a considerable margin the quality requirements for microwave nonlinear capacitors. *P-n* junctions are usually made by incorporating certain impurity atoms in the semiconductor lattice. These donor and acceptor atoms tend to ionize and become fixed positive and negative charges.

If the concentration of ionized donors is N_{D+} and the concentration of ionized acceptors is N_{A-}, the net charge density is $q(N_{D+}-N_{A-})$ and varies as shown in Fig. 3 for a linearly-graded junction, which will be used as an example. At zero bias the *p-n* junction is in equilibrium. For typical graded junctions, the concentration p of holes and n of electrons varies with distance as shown in Fig. 4(a). The place where the fixed charge density is zero is called the stoichiometric junction. It may be seen that in a small region around the stoichiometric junction, there are very few holes or electrons. This region is known variously as the depletion layer, exhaustion region, or space-charge region. The latter term refers to the fact that the net fixed charge is not neutralized by mobile carriers. Outside of the depletion layer the mobile carriers are present in almost exactly the right numbers to neutralize the the fixed charges. Evidently, holes are required to neutralize the negative fixed charges on the p side of the junction and electrons are required to neutralize the positive fixed charges on the n side of the junction.

Suppose the junction is biased slightly in the forward direction. This means applying to the contact on the p side a voltage that is positive with respect to the contact on the n side. This voltage will urge the hole and electron distributions to move toward each other, as shown in Fig. 4(b). For this motion to take place without leaving large unbalanced electric charges in the previously neutral p and n regions, it is necessary for additional holes and electrons to enter the semiconductor at the contacts. It is assumed that the contacts are of such a nature as to permit this. If a still larger forward voltage is applied, holes and electrons intermingle appreciably in a not-so-thoroughly-depleted layer and

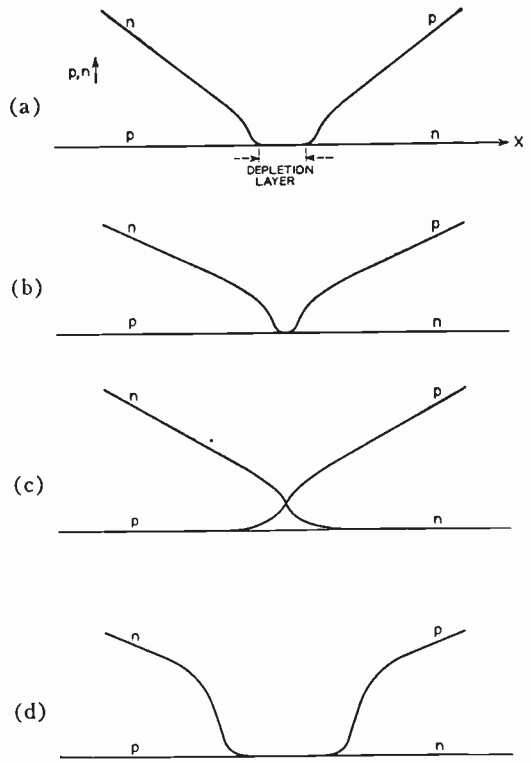


Fig. 4—Hole and electron concentrations in neighborhood of a graded *p-n* junction: (a) zero bias, (b) small forward, (c) large forward bias, (d) reverse bias.

on either side of it [Fig. 4(c)]. Despite the intermingling, one can recover the charge if it is allowed to return in a time that is short compared to the time required for appreciable recombination. (Data are given below on a diffused silicon nonlinear capacitor to show that recombination is indeed negligible if the frequency is much higher than 1 mc.) When a substantial amount of the stored charge is thus represented by intermingled holes and electrons, one speaks of "carrier-storage capacitance."

When the junction is biased in reverse direction, the depletion layer widens as shown in Fig. 4(d). When the reverse voltage reaches the so-called "breakdown voltage," the junction begins to conduct copiously, usually because of generation of carriers in the depletion layer by a process known as avalanche multiplication. In the range between this reverse breakdown voltage and the slight forward voltage corresponding to Fig. 4(b), the charging and discharging of the junction takes place by the motion of hole and electron distributions toward and away from each other, without appreciable intermingling of holes and electrons. This type of capacitance is referred to as a depletion-layer capacitance. The thermally generated reverse saturation current of the junction in shunt with this capacitance produces full shot noise. The reverse current of clean silicon *p-n* junctions is so small that this shot noise is not important at the relatively low impedance levels of high-frequency circuits. Reverse currents of germanium *p-n* junctions cannot be so confidently neglected if the temperature is appreciably above room temperature.

The two loss mechanisms thus far mentioned—forward conductance, caused by recombination, and reverse leakage, caused by generation—produce shot noise that is practically independent of frequency. There is, in addition, the possibility of a frequency-dependent conductance arising from inability of the holes and electrons to redistribute themselves instantaneously in response to very rapid changes in applied voltage. This phenomenon has been called dispersion of the capacitance.[4] The dispersion of depletion-layer capacitance must occur at very high frequencies; it has never been observed experimentally. The fast response of the depletion-layer capacitance is attributed to the fact that the distances the holes and electrons must move are only small fractions of the depletion-layer width.[4] The dispersion of the storage capacitance has been one of the main topics in the theoretical and experimental study of p-n junctions.

The depletion-layer capacitance is probably the basis of the microwave diode amplifiers that will be described. One reason for this belief is a theoretical analysis of the effect of series resistance on the gain and noise of variable capacitance frequency converters.[5] For a capacitance varying sinusoidally with time, the theory shows that a six-to-one dynamic range of capacitance is all one might wish; a three-to-one range is almost as good. These modest dynamic ranges can be obtained with depletion-layer capacitance alone.

The depletion-layer capacitance of a linearly-graded junction is given as a function of voltage v by the approximate formula[4]

$$C \approx \frac{C_0}{\sqrt[3]{1 - (v/\phi)}} \tag{1}$$

where ϕ is a constant that depends upon the impurity gradient but is about one-half volt for most silicon junctions. This formula is quite accurate for reverse biases (negative values of v) and holds reasonably well for moderate forward biases if ϕ is chosen empirically. A corresponding formula for abrupt junctions is

$$C \approx \frac{C_0}{\sqrt{1 - (v/\phi)}} . \tag{2}$$

It is sometimes suggested that the square-root relation (2) is "more nonlinear" than the cube-root relation, (1), and that the abrupt junction therefore is preferable to the graded junction. This argument is meaningless when not referred to a specific application and is probably false under most circumstances. The ease with which low series resistance can be obtained in graded junctions appears a decisive advantage in their favor.

Since the dynamic range of the capacitance is more than adequate, the parameters determining the gain and noise potentialities of the diode are, as shown in Fig. 5,

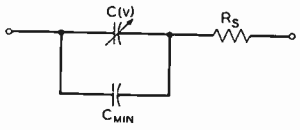


Fig. 5—High-frequency equivalent circuit of nonlinear capacitor.

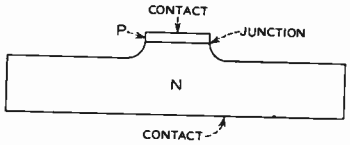


Fig. 6—P-n junction "mesa" diode.

the series resistance R_s and the minimum capacitance C_{min}, the latter being the capacitance for reverse voltages just short of breakdown. Then, if an arbitrary impedance level is permitted, a single figure of merit will serve to describe the diode. This figure of merit may be written as a *cutoff frequency* f_c, defined by

$$f_c = \frac{1}{2\pi R_s C_{min}} . \tag{3}$$

Diffused silicon nonlinear capacitors with cutoff frequencies typically 60 to 120 kc, are being made from graded junctions with impurity gradients at the junction of 10^{23} to 10^{24} cm^{-4}.[6] The preferred impurity distribution for minimum series resistance is one in which the impurity gradient is everywhere as large as, or larger than, the impurity gradient at the stoichiometric junction.

The series resistance is inversely proportional to area and the capacitance is proportional to area, so the cutoff frequency is independent of area. In other words, the graded junction is a "planar" or "one-dimensional" structure. In this respect it differs from most high-frequency diodes, such as point-contact and welded-contact diodes, which must have small contact size for good high-frequency performance (because their series resistances vary inversely as the contact diameter). It is also a point of difference from the transistor triode, which requires two-dimensional flow of majority and minority carriers in the base layer.[7]

Apart from the series resistance, the diffused silicon diodes indeed appear to be voltage-dependent capacitors up to microwave (perhaps much higher) frequencies. The purely capacitive impedance should have no shot noise, and the shot noise of the reverse current is usually negligible. As a tentative hypothesis, it is suggested that the series resistance exhibits thermal noise. Then the cutoff frequencies that have been obtained lead one to expect microwave and UHF diode amplifiers to have noise figures lower than those now obtained with the best electron tubes. Preliminary evidence of low noise is given in Section VIII.

[4] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

[5] D. Leenov, "Gain and noise figure of a variable capacitance up-converter," *Bell Sys. Tech. J.*, vol. 37; July, 1958.

[6] A. E. Bakanowski, N. G. Cranna, and A. Uhlir, Jr., "Diffused silicon and germanium nonlinear capacitors," presented at the IRE-AIEE Semiconductor Device Research Conference, Boulder, Colo., July, 1957.

[7] A. Uhlir, Jr., "Shot noise in p-n junction frequency converters," *Bell Sys. Tech. J.*, vol. 37; July, 1958.

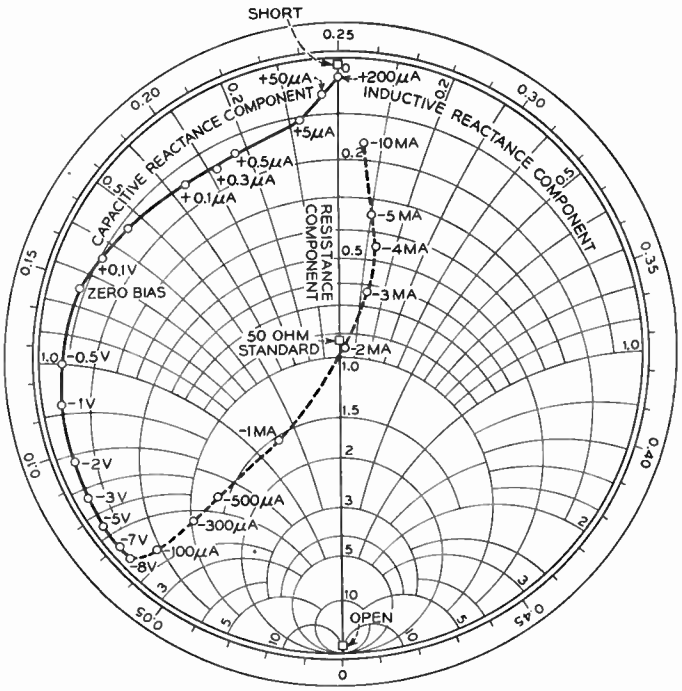


Fig. 7—Small-signal impedance of diffused silicon nonlinear capacitor at 1000 mc. Broken line is reverse breakdown region.

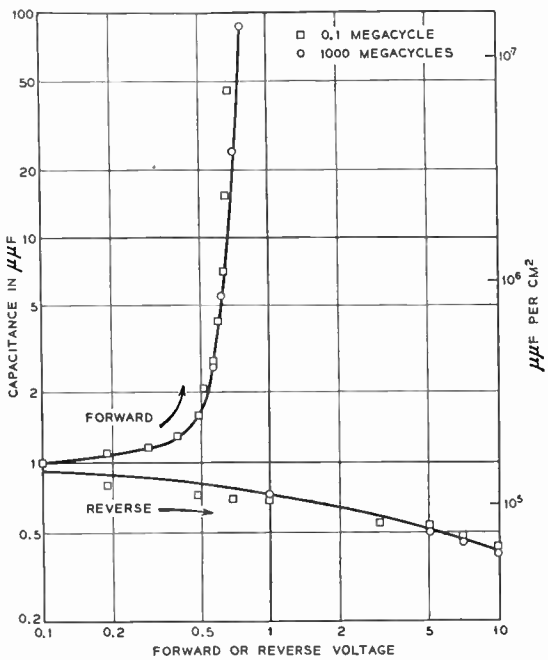


Fig. 8—Capacitance of a diffused silicon nonlinear capacitor as a function of voltage.

The diffused silicon "mesa-type" p-n junction structure is shown in Fig. 6. The 1000-mc small-signal measurements on a diode of this type are given in Smith chart form in Fig. 7. The solid curve represents the nonlinear capacitance bias range; the broken line shows the impedance when breakdown current is flowing. Some points corresponding to impedance standards are shown on the chart. The capacitance-voltage relation obtained from this chart is compared in Fig. 8 with the same relation determined by measurements at 100 kc. The resistance-reactance trajectory in rectangular coordinates

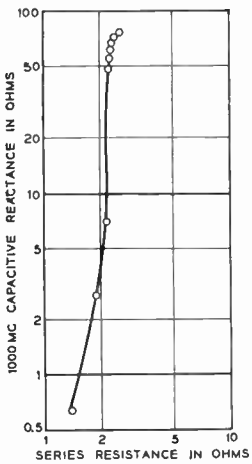


Fig. 9—Small-signal impedance of diffused silicon nonlinear capacitor, for various dc biases.

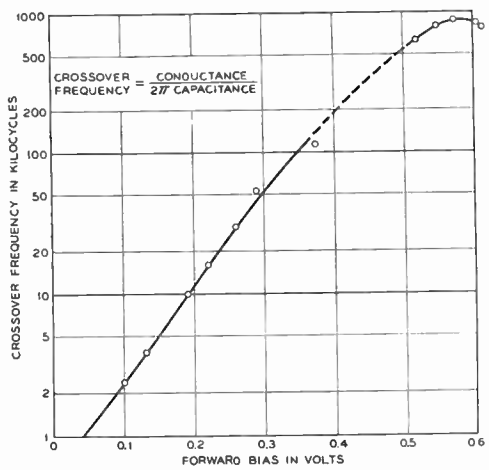


Fig. 10—Frequency at which conductance and susceptance are equal. Based on 100-kc measurements of 10^{-3} and 10^{-5} cm² areas of diffused silicon junction of gradient 5×10^{23} cm^{-4}.

(simply a transformation of the Smith chart) is shown in Fig. 9. The justification for the equivalent circuit of Fig. 5 is evident from Fig. 9 and the frequency-independence of capacitance shown in Fig. 8. Substantial reductions in series resistance have been realized since these graphs were prepared. Fig. 10 shows, as a function of bias, the crossover frequency at which the junction susceptance equals the junction conductance (measured at 100 kc). Evidently, the frequency-independent part of the conductance is negligible if the operating frequency is much higher than 1 mc.

Silicon p-n junctions are marketed by several concerns for use as electronically-variable capacitors in the VHF range. These devices do not have and do not require the high cutoff frequencies of the diodes intended for low-noise microwave amplifiers. Previously, experimental nonlinear capacitors made by alloying indium to n-germanium[8] had been tested in similar applications (*e.g.*, frequency control).

Historically, the first semiconductor diodes reported to give amplification were welded-contact germanium

[8] L. J. Giacolletto and J. O'Connell, "A variable-capacitance germanium junction diode for VHF," *RCA Rev.*, vol. 17, pp. 68–85; March, 1956.

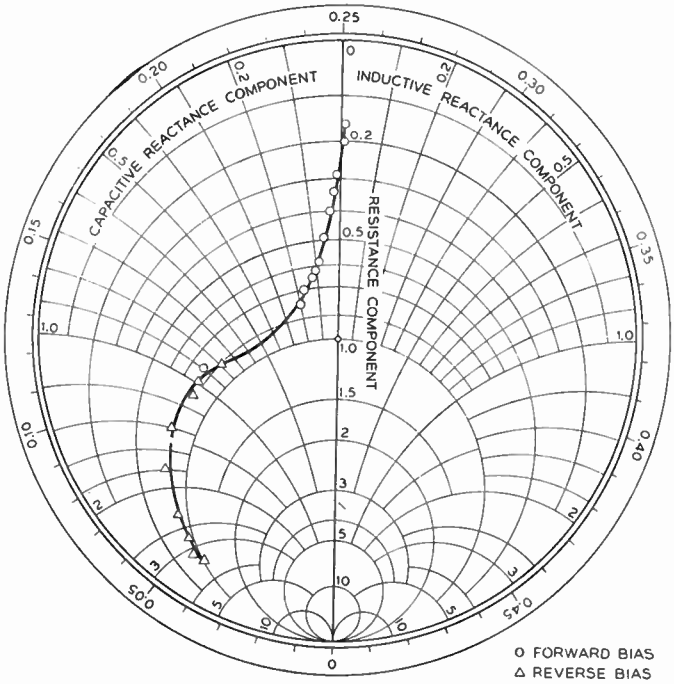


Fig. 11—Small-signal impedance of gold-bonded germanium diode at 9 kmc. Chart center is 60 ohms.

diodes.[9] A modern version of this type of diode, a gold-bonded germanium diode, has been designed for use in a microwave relay transmitting modulator.[10] The cutoff frequency is 40 kmc, which is more than adequate for an amplifying modulator between 60 to 80 mc and 6000 mc. A Smith chart plot of the 9-kmc small-signal admittance of one of these diodes is given in Fig. 11.[11] For large forward currents, the resistance decreases and acquires a small inductive component; these effects are attributed to conductivity modulation of the series resistance.

While the bonded diodes are economical and practical, future developments will doubtless be predicated on the demonstrated success of the diffused junction nonlinear capacitor. The impedance level of a large-area p-n junction is inconveniently low for high-frequency circuits. A way of building up the impedance is to put a number of such junctions in series. The power-handling capability is increased, first, by the large junction areas and, second, by the multiplicity of junctions.

A series stack can be approximated in a single crystal of semiconductor without intervening metallic contacts (which might add resistance). Consider a single crystal with alternate layers of n- and p-type semiconductor, as shown in Fig. 12. One readily obtains nonlinear capacitance with the impurity charge distribution shown in Fig. 13. The p-n junctions in this structure are graded considerably more steeply than the n-p junctions. The

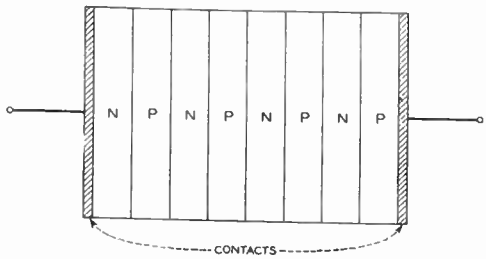


Fig. 12—Multiple-junction diode.

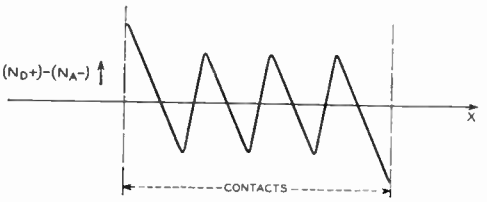


Fig. 13—Impurity distribution for nonlinear capacitance in a multiple-junction diode.

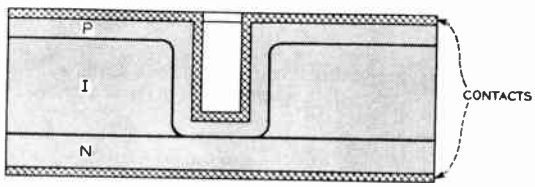


Fig. 14—Section through dimple diode made by diffusion in silicon.

steeper junctions have higher capacitance per unit area and may be regarded approximately as ac short-circuits. The equivalent circuit of the structure, then, is a series connection of the nonlinear capacitances of the more gradual junctions.

Another ramification is the production of a graded p-n junction embedded in a p-i-n diode, as shown in Fig. 14.[12] This dimple structure is resistant to atmospherically-induced changes in capacitance or breakdown voltage and can safely dissipate more power than equivalent mesa diodes. It is a way of contacting and handling very small p-n junctions.

In most diodes, reverse breakdown due to avalanche multiplication occurs at a number of localized discharges, each of which is called a microplasm. A large fraction of the dimple diodes break down at just one microplasm. When this microplasm turns on (starts to pass reverse current), a microwave transient is generated—an effect which appears to be the first observed conversion of dc power into microwave radiation by a p-n junction.[13]

A p-i-n diode of considerable linear extent should act as a transmission line. A sequence of p-n junctions in this kind of transmission line, as in Fig. 15, is a traveling-wave diode amplifier in a single piece of silicon.

[9] H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, N. Y., ch. 13; 1948.
[10] "Semiconductor diodes yield converter gain," *Bell Labs. Rec.*, vol. 35, p. 412; October, 1957.
[11] D. Leenov, private communication.
[12] N. G. Cranna and A. Uhlir, Jr., paper in preparation.
[13] J. L. Moll, A. Uhlir, Jr., and B. Senitzky, "Microwave transients from avalanching silicon diodes," to be published in PROC. IRE.

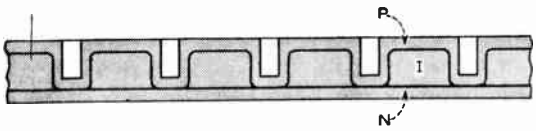


Fig. 15—Section through *p-i-n* transmission line with integral *p-n* junctions.

V. Nonlinear Resistor Structures

Nonlinear resistors fall into two groups. *P-n* junctions can be used at frequencies up to several hundred megacycles. Point-contact diodes are used at higher frequencies, including millimeter waves.[14]

Recombination processes are necessary for nonlinear resistance action in *p-n* junctions. It is suggested that point-contact diodes are *p-n* junctions in a broad sense and also require recombination. The increasing variety of semiconductor materials being used for microwave point-contact diodes is noted. Finally, the possibility of of making nonlinear resistors that do not depend upon recombination is considered.

The forward current in a *p-n* junction is maintained by recombination of holes and electrons. One is concerned with the situation like that shown in Fig. 4(c) in which the hole and electron distributions overlap appreciably. For each electronic charge that flows in the external circuit in the forward direction, one hole and one electron must recombine. This recombination can take place in the neutral *n* or *p*-regions or in the depletion layer.

The graded *p-n* junction has one effect that tends to make it a difficult structure in which to obtain nonlinear-resistance action. The built-in field in the neutral *p* and *n* regions is such as to make any minority-carrier admittance capacitive rather than resistive. However, if nonlinear-resistance action can be obtained in spite of this effect, the graded junction will be very desirable because of the low series resistance relative to the depletion-layer capacitance.

To enhance recombination (a practice referred to as "ruining lifetime"), impurities may be added to the crystal to serve as catalysts for the recombination process; gold in silicon is an example.[15,16] Also, to lower lifetime, the geometrical arrangement of the lattice may be rendered imperfect by mechanical abuse at room temperature or elevated temperatures, or by irradiation with electrons, neutrons, etc. A sandblasted surface near the *p-n* junction is used in one experimental germanium diode.[17]

Increased understanding of recombination processes may eventually lead to microwave *p-n* junction nonlinear resistors. In the meantime, point-contact diodes will serve as microwave nonlinear resistors, as shown

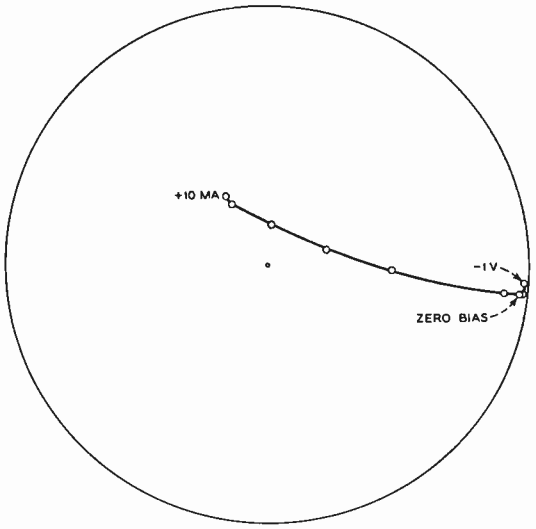


Fig. 16—Small-signal impedance of 1N23B silicon point-contact diode. Presented as 1000-mc reflection coefficient when mounted in commercial crystal mount at end of 50-ohm line.

by the measurements in Fig. 16. The physical structure of the point contact (to say nothing of the physical mechanism) is conjectural, for reasons that will be apparent from a brief description of processes used in fabrication.

Silicon point-contact diodes are made from *p*-type silicon doped to low resistivity (10^{-2} ohm-cm) with boron or aluminum. Lately, aluminum is preferred on the basis of empirical evidence that it gives better burnout resistance.[18] "Burnout" refers to any impairment of diode performance by electrical overload.

Point-contact diodes made in single-crystal silicon with controlled resistivity are more uniform than,[19] but are otherwise similar to, diodes made from polycrystalline silicon. High minority carrier lifetime in the starting material is not desired for obtaining nonlinear resistor action. Neither is it disadvantageous, because the effective lifetime at the contact is probably determined by mechanisms, speculated on below, for which the original lifetime is irrelevant.

The silicon is sliced and given a heat-treatment which may increase the resistivity of a thin surface layer. A high-resistivity surface layer would decrease the capacitance per unit area without proportionately increasing the series resistance and would assist in removing carriers from the contact according to one mechanism to be described. The diode is assembled by bringing a sharp tungsten point in contact with the surface. Good rectification is not obtained until the contact is mechanically disturbed, for example, by tapping the assembled unit with a small hammer.

Low-noise germanium point-contact diodes have been

[14] W. M. Sharpless, "Water type millimeter wave rectifiers," *Bell Sys. Tech. J.*, vol. 35, pp. 1385–1402; November, 1956.
[15] G. Bemski, "Recombination in semiconductors," this issue, p. 990.
[16] A. E. Bakanowski and J. H. Forster, paper in preparation.
[17] R. H. Rediker and D. E. Sawyer, "Very narrow base diode," Proc. IRE, vol. 45, pp. 944–953; July, 1957.
[18] E. J. Feldman, "Improved S-Band Crystal Diodes," Microwave Crystal Rectifier Symposium Record, Fort Monmouth, N. J., p. 196; February, 1956.
[19] J. H. Bollman, "Use of Single-Crystal Silicon in Microwave Varistors," Bell Telephone Labs., First Interim Rep. on Improved Crystal Rectifiers, Signal Corps' Contract DA36-039-sc-73224; May 15, 1957.

developed.[20] Single-crystal n-type germanium of 10^{-2} ohm-cm resistivity is used. These diodes are not tapped; instead, the required artistry is to pass an electrical "forming" current through the assembled diode. This germanium microwave diode gave a clear-cut improvement in noise figure over the silicon diodes that were current at the time of its introduction. Soon thereafter improved silicon diodes appeared. At present, the noise-figure race is so close that one or the other type of diode can be made to appear better by more aggressive selection of the best units from large batches or by altering circuit design.

Recently, n-type gallium arsenide has been used to make point-contact microwave diodes.[21] The results are good; further studies will be needed before it can be affirmed that this material is superior to silicon or germanium.

Point-contact diodes for use as microwave rectifiers are made by much the same techniques as the diodes for superheterodyne use. However, the requirement of high impedance, to be explained in Section VII, generally leads the designer to use smaller contact areas in diodes for rectifier use.

The point contact may be regarded as a kind of p-n junction if one realizes that the surface of a semiconductor is likely to have an electric charge.[22] The magnitude and even the sign of this charge depends upon the surface treatment, but under given conditions may be fairly constant. The surface charge may be regarded as fixed in comparison to the mobile carriers. If the surface charge is opposite in sign to the fixed charge due to impurities in the bulk of the semiconductor, a kind of p-n junction results. It seems beyond doubt that such a surface p-n junction exists at the emitter of a point-contact transistor made on n-type germanium, and at the emitter and collector contacts of a surface-barrier transistor. For transistor action to occur in either of these devices, it is necessary that the surface charge be quite strong. Then, in a layer of semiconductor just beneath the surface, there will be a high concentration of carriers neutralizing the surface charge. Forward current flows by injection of these carriers into the bulk material, where they are minority carriers. Nonlinear resistance action requires some mechanism that prevents the return of these minority carriers. It is questionable that the initial bulk minority-carrier lifetime is short enough to accomplish this automatically, even in polycrystalline semiconductors.

One possible mechanism for nonlinear resistance in a point-contact diode with carrier injection is simply the difficulty the injected carriers may have in finding their way back to the small contact. This mechanism cannot explain why rectification frequencies are as high as they are, if one assumes (as in ordinary p-n junction theory) that the minority carriers move by simple diffusion. However, the flow of forward current is accompanied by an electric field that hastens the departure of minority carriers from the neighborhood of the contact. For this field to be adequate, the resistivity very near the contact must be higher than the bulk resistivities that are used. Heat-treatment or forming may very well produce such a resistivity alteration.

Carrier injection into the bulk semiconductor may not be important in microwave point-contact diodes, which are made on bulk materials of much lower resistivity than preferred for transistors. It is possible that most of the forward current is carried by carriers moving from the bulk material into the surface layer, there to recombine.

The author has constructed a theory of shot noise in p-n junction frequency converters.[7] It should also be applicable to point-contact nonlinear resistors whether the carriers are emitted from surface to bulk or vice versa. The theory shows the importance of local-oscillator waveform in determining the noise figure of nonlinear resistor superheterodyne circuits.

Future progress in nonlinear resistors may employ structures that use collection, rather than recombination, to remove carriers. For example, n-i-n and n-p-n diodes are nonlinear resistors up to frequencies which compare with the transit time for an electron to go from one n region to the other. If symmetrical, such structures could not function as passive rectifiers, but should be usable frequency converters.

Another proposal is the "drift diode," with the impurity distribution shown in Fig. 17.[23] The impurity gradient near the junction gives rise to an electric field, even under equilibrium conditions, that is in such a direction as to discourage the return of injected carriers.

VI. P-I-N Diodes

When the first p-n junctions became available, everyone who studied them was impressed by their superb low-frequency rectification characteristics, compared to the previously available point-contact diode. But some device engineers were not content and proposed the p-i-n structure,[24,25] shown in Fig. 18(a). The symbol I stands for intrinsic or high-resistivity semiconductor. The intrinsic layer gives a very much larger breakdown voltage than can be obtained in simple p-n junctions. The somewhat more surprising feature, which makes the structure an excellent power rectifier, is that in forward bias the intrinsic region is filled with injected car-

[20] G. C. Messenger and C. T. McCoy, "Theory and operation of crystal diodes as mixers," Proc. IRE, vol. 45, pp. 1269–1283; September, 1957.

[21] D. A. Jenny, "A gallium arsenide microwave diode," Proc. IRE, vol. 46, pp. 717–722; April, 1958.

[22] J. Bardeen and W. H. Brittain, "Physical principles involved in transistor action," *Phys. Rev.*, vol. 75, pp. 1208–1223; April 15, 1949.

[23] C. H. Knowles, "Characteristics of the Drift P-N Junction," presented at the IRE-AIEE Semiconductor Device Research Conference, Purdue, Ind., 1956.

[24] R. N. Hall, "Power rectifiers and transistors," Proc. IRE, vol. 40, pp. 1512–1518; November, 1952.

[25] M. B. Prince, "Diffused p-n junction silicon rectifiers," *Bell Sys. Tech. J.*, vol. 35, pp. 661–684; May, 1956.

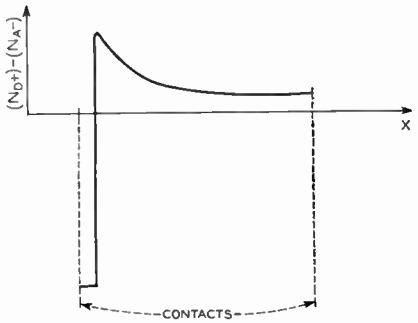


Fig. 17—Impurity distribution in drift diode.

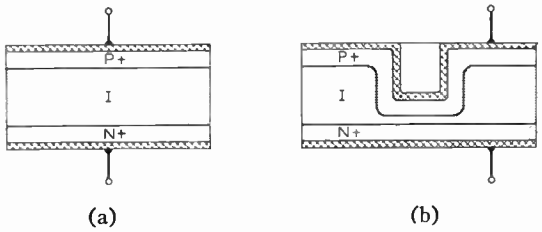


Fig. 18—(a) *P-i-n* diode. (b) Dimple structure of *p-i-n* diode with small effective area. In example, dimple is 5 mils diameter, thickness at bottom of dimple is about 3 mils.

riers and no longer has a high resistivity. Therefore, the forward drop at large currents is moderate. While this structure does its intended job of rectifying power frequencies such as 60 cycles per second or 400 cycles per second, it becomes a poor rectifier at frequencies as low as a megacycle (depending upon the thickness of the intrinsic region).

As long as thinking about microwave diodes revolved about the rectifier, there was little inclination for anyone to place in microwave circuits a device that could not even rectify one megacycle. Moreover, the low-frequency capacitance of these diodes was of the order of 20 mmf and up; what could be more absurd than to put such a device in a high-frequency circuit?

But, let us consider the *p-i-n* diode shown in Fig 18(b). To be sure, the dimensions of this diode are slightly smaller than those customarily used in the smallest power rectifiers, but they are still of the same order of magnitude and are enormous compared to the dimensions of the active area of point-contact microwave rectifiers; the zero-bias capacitance at 100 kc is 13.8 mmf. The raw data of a 1000 mc measurement of this diode is shown in the Smith chart of Fig. 19.[26] The parameter that is varied is the dc bias. Also shown are the measurements made on a short circuit and an open circuit constructed in the same diode cartridge and mounted in the same crystal holder. At zero bias and reverse biases the impedance is extremely high compared to the chart-center impedance of 45 ohms, while at moderately large forward currents the impedance of the diode is very small compared to 45 ohms. The effective shunt capacitance in reverse bias is only 0.3 mmf.

[26] Fabricated by N. G. Cranna; measured by D. E. Iglesias.

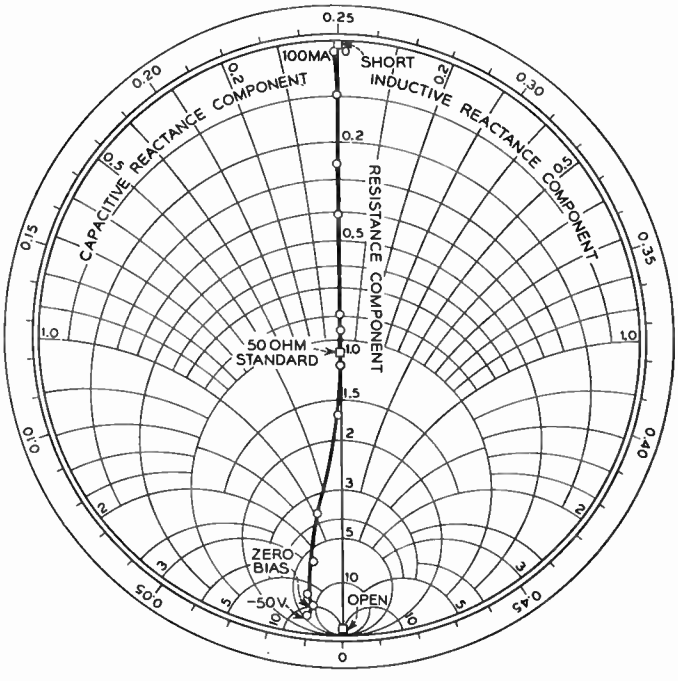


Fig. 19—Small-signal 1000-mc impedance of *p-i-n* dimple diode of Fig. 18(b).

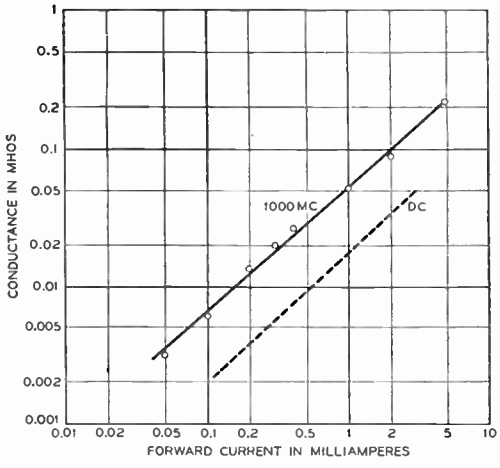


Fig. 20—Conductance of *p-i-n* dimple diode.

The value of the conductance is plotted as a function of the dc current in Fig. 20. The dashed line in this figure gives the approximate value of the dc conductance at corresponding dc currents. At high frequencies, one observes only the conductivity-modulated resistance of the intrinsic region. Accordingly, the high-frequency conductance is larger than the dc conductance.

Evidently the *p-i-n* diode can be used as an electronically-variable attenuator of microwave frequencies. The variable resistivity of an intrinsic region can also be utilized in distributed structures to provide variable attenuation;[27] the limit in this direction would be a *p-i-n* transmission line with variable attenuation.

[27] E. M. Gyorgy and G. L. Pearson, private communication.

VII. Rectifiers

In proposing the use of a nonlinear resistor in a circuit, one should keep in mind the ultimate system objective. If this objective necessarily implies that ac power must be converted to dc power without an external source of power, then an asymmetrical nonlinear resistor (rectifier) must indeed be used. One such situation might be a light, portable receiver to operate without batteries. In the laboratory, the combination of a rectifier and a meter makes a convenient passive detector of electromagnetic radiation of all frequencies up to the limiting rectification frequency of the diode. The broad-band detection capabilities of the rectifier make it attractive for counter-measures, for which purpose it may be combined with a lightweight, low-power, transistor video amplifier.

Unfortunately, the sensitivity of rectifier receivers is poor, for reasons that have been recognized for a long time,[28] and just briefly are outlined here. When the incoming signal is small, the rectifier acts as a square law device. This means that the output dc voltage is proportional to the square of the RF voltage. Accordingly, the efficiency of rectification is proportional to RF power and decreases when the available power decreases. Just when efficiency is most needed, it is lacking. For a given amount of power, the best efficiency can be obtained by making both the source impedance and the diode impedance as high as possible, to obtain as large a voltage as possible. There are limits on how high these impedances may be for reasonable circuits and diode contact areas.

In detecting faint signals, the noise competing with the rectified output is the low-frequency noise of the diode and the noise of the video amplifier. In the usual receiver the diode is used at zero dc bias; it would be contrary to the second law of thermodynamics for the diode to exhibit any but thermal noise. (A biased diode can have any noise between one-half thermal and infinity.[29])

Theoretically, improved sensitivity can be obtained by lowering the temperature, for two reasons. The important reason is the improvement in efficiency of rectification, for a given input power. This effect may be predicted from the theoretical rectifier characteristic

$$i = i_s\{e^{qv/kT} - 1) \tag{4}$$

which shows that the nonlinearity improves as temperature is lowered. The other reason is the reduction of the thermal noise of the diode, but this reduction is of limited advantage unless the noise in subsequent amplifiers can be correspondingly reduced. Specially designed diodes would be required for very low temperature use.

[28] Torrey and Whitmer, *op. cit.*, ch. 11.
[29] Torrey and Whitmer, *op. cit.*, ch. 6.

VIII. Frequency Converters and Diode Amplifiers

A frequency converter is a circuit that can accept signals at certain frequencies and deliver proportionate signals at one or more other frequencies. Nonlinear devices may be used to make frequency converters. Many frequency converters use one or more diodes as nonlinear elements. With nonlinear capacitors, one can make amplifying frequency converters which give output signals of greater power than the input signals. Diode amplifiers can be derived as special amplifying frequency converters that have the input frequency or frequencies included among the output frequencies.

Harmonic generators, which convert power of one frequency to a multiple frequency, are discussed in Section X and are excluded from the present connotation of the term "frequency converter."

As frequency converters, semiconductor diodes have the advantages of low-power requirements, freedom from microphonics, unlimited life when protected from overloads, and low cost. Moreover, the nonlinear capacitor is a low-noise or medium-power high-frequency amplifying device, with all of the above advantages and vastly improved resistance to electrical burnout, compared to point-contact diodes.

A circuit element whose value varies with time is a frequency converter. A picturesque representation of such an element might be a rotary variable capacitor driven by a large motor. The capacitance—the ratio of charge to voltage—is then a periodic function of time. Another type of such element would be a rheostat whose slider was made to move periodically by some mechanical drive. Then the resistance— the ratio of voltage to current—is a periodic function of time. Such conceptual objects would be linear frequency converters. Their efficiencies and impedances would not depend upon the magnitude of the impressed electrical signals (except that the electrostatic forces between the capacitor plates should not be large enough to react upon the motor drive).

A nonlinear capacitor or a nonlinear resistor can be made to imitate the mechanically driven frequency converters. These imitations may be used at microwave frequencies. The scheme is to apply a relatively large periodic voltage to the nonlinear element. Then it is found that small signals applied at frequencies other than the fundamental and harmonics of the large voltage are converted in frequency just as if the device were a time-varying resistance or capacitance. The small-signal response of such a frequency converter is linear, as can be shown very simply. Suppose that $v(t)$ is written

$$v(t) = v_l(t) + v_s(t) \tag{5}$$

where v_l is the large-signal part, periodic or not, and v_s

can be regarded as arbitrarily small. Then $f(v)$ can be expanded in a Taylor's series;

$$f(v(t)) = f(v_l(t)) + f'(v_l(t))v_s(t). \tag{6}$$

For the nonlinear resistor, $i=f(v)$ and it is natural to define a small-signal conductance $G=di/dv=f'(v)$. Then, from (6), one has

$$i(t) = i_l(t) + i_s(t) \tag{7}$$

where

$$i_l(t) = f(v_l(t)) \tag{8}$$

and

$$i_s(t) = G(t)v_s(t). \tag{9}$$

The last equation is the one of interest, because it shows that small variations in current are linearly related to small variations in voltage. In similar fashion one may define, for the nonlinear capacitor, $C=dQ/dv$ and obtain, for the small-signal charge Q_s,

$$Q_s(t) = C(t)v_s(t). \tag{10}$$

The above relations in the time domain show the underlying simplicity of the mathematical approach used to linearize the problem. The detailed operation of frequency converters can best be analyzed in the frequency domain. The procedures used in such an analysis will be outlined. For the nonlinear resistor, suppose that $G(t)$ is periodic with a fundamental frequency b. This condition could prevail in a mechanically-varied resistor, but, of course, here we are most interested in periodic $G(t)$ resulting from the application of a periodic large-signal beating-oscillator voltage $v_l(t)$ to a nonlinear resistor type of diode. One can write

$$G(t) = \sum_{n=-\infty}^{+\infty} G_n e^{2\pi jnbt}. \tag{11}$$

If the applied voltage is given by

$$v(t) = \sum_{s\pm} \sum_{m=-\infty}^{\infty} v(mb+s)e^{2\pi j(mb+s)t}, \tag{12}$$

the resulting current is

$$\begin{aligned} i(t) &= G(t)v(t) \\ &= \sum_{s\pm} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G_n v(mb+s)e^{2\pi j((m+n)b+s)t} \end{aligned} \tag{13}$$

and may be seen to contain the same frequency components as the impressed voltage, that is, one can deal with a closed set of frequencies. The impressed voltages and currents are related by a conversion matrix. Thus

$$i(mb+s) = \sum_n Y_{mn}v(nb+s \tag{14}$$

where the conversion matrix is

$$Y_{mn} = G_{m-n}. \tag{15}$$

Exactly the same steps may be followed for the time-varying capacitance, with the result

$$Q(mb+s) = \sum_n C_{m-n}v(nb+s). \tag{16}$$

It is customary to treat current rather than charge as a variable in circuit analysis. For any frequency ν,

$$i(\nu) = 2\pi j\nu Q(\nu). \tag{17}$$

Thus

$$i(mb+s) = 2\pi j(mb+s)\sum_n C_{m-n}v(nb+s) \tag{18}$$

so that the conversion matrix is given by

$$Y_{mn} = 2\pi j(mb+s)Cm\text{-}n. \tag{19}$$

A frequency converter is made up of the periodically varied element, described by a conversion matrix, plus impedances terminating all of the frequencies $mb+s$. Having chosen these terminations, one may, in principle, determine the performance of the frequency converter by linear network analysis.

If any other set of frequencies $mb+s'$ is considered, another conversion matrix can be constructed. The linear problem for this new set of frequencies may be solved entirely separately from the first set of frequencies. Thus, a signal containing a spectrum of frequencies can be resolved into frequency components; separate analyses are conducted for each closed set of frequencies generated by the frequency components of the original signal.

When diodes that are neither nonlinear resistors nor nonlinear capacitors are used in frequency converters, they can still be described by conversion matrices for sets of frequencies $mb+s$. The elements of the conversion matrix are complex, in general. For example, a variety of conversion matrices arise from considering p-n junction action as nonlinear injection of minority carriers, followed by diffusion, drift, and recombination.[2]

The general conversion matrix involves an infinite number of frequencies, which usually makes an exact circuit analysis difficult. Most analyses neglect all but two or three frequencies, with no justification except the reasonableness of the results. An exception is an analysis of gain and noise in up-converter amplifiers,[5] in which a nonlinear capacitor is assumed having the equivalent circuit of Fig. 5. A way of treating the infinite number of frequencies is given which is logically consistent with the presence of series resistance.

The problem now is to obtain some general insights from the conversion matrices and any other considerations that can be applied. By the conversion matrix analysis of particular situations, one finds that amplify-

ing frequency converters are possible with time-varying capacitance, but one soon suspects that a time-varying resistance cannot amplify. The truth of the latter surmise is easily demonstrated by the following argument. A mechanically-varied rheostat cannot have an electrical output that exceeds the electrical input. Therefore, a linear-for-small-signals frequency converter made with a nonlinear resistor cannot amplify, because it has exactly the same conversion matrix as a hypothetical mechanically-varied rheostat, as long as nonlinear resistors with negative resistance ($f'(v)<0$) are excluded.

The correspondence between the rheostat and the nonlinear resistor must not be carried too far. The rheostat can be an amplifier (of the mechanical signal). Also, noise in a rheostat is obviously thermal noise, while the nonlinear resistor mechanisms suggested above imply shot noise. Thus, a given $G(t)$ waveform has definite signal transmission properties but may produce different amounts of noise, depending upon the physical mechanism.

To obtain insight into the properties of linear frequency converters utilizing nonlinear capacitors, the analysis thus far is used only to suggest what signal frequencies should be considered. In the search for general principles, one turns naturally to conservation of energy within the (almost) lossless nonlinear capacitor. But this principle by itself is quite empty, for any reasonable amount of power will be cheerfully supplied from the beating oscillator (often called the "pump") if so doing produces the desired signal transfer.

What is needed is a "second law," like the second law of thermodynamics or the principle of conservation of momentum in dynamics. The required second law is given by Manley and Rowe in a general analysis of nonlinear reactances.[30] The results obtained for several simple and important types of frequency converters will be discussed.

Diagrams are invaluable for discussions of this kind, but no particular representations have been universally adopted. One obstacle to universality is that it is mathematically most convenient to use negative frequencies, so that Fourier series can be written, as in (7), with complex exponentials. Therefore, the transition to positive frequencies will be made explicit and finally a simplified diagram will be suggested.

Fig. 21 is a pictorial representation of the Fourier components of a periodic function such as the beating-oscillator voltage. Positive and negative harmonic frequencies are used. The complex amplitude of each component is the vector in the x-y plane perpendicular to the frequency axis at the corresponding frequency mb. Since the voltage is a real function of time, $v(-mb)$ is the complex conjugate of $v(mb)$.

Graphs like Fig. 21 are more complicated than necessary for a general discussion. In Fig. 22(a), the spectrum of Fig. 21 is schematized by replacing the two-dimensional vector by a vertical line whose length might be the magnitude of the vector; what matters is that the presence of a line denotes the existence of a vector. One could represent the vectors $v(mb+s)$ according to the same scheme. However, it is helpful to consider a hypothetical signal spectrum $mb+s$, $mb+s'$, $mb+s''$, $\cdots$ founded upon a perhaps infinite set of frequencies s, s', s'', $\cdots$, as indicated in Fig. 22(b). In the equations [*e.g.*,(8)], one also encounters the frequencies $mb-s$, shown in Fig. 22(c). Because the signal voltage is real, Fig. 22(c) does not represent any information not already represented in Fig. 22(b). Therefore, in calculations with conversion matrices, it is sufficient and advisable to consider either Fig. 22(b) or 22(c), but not both.

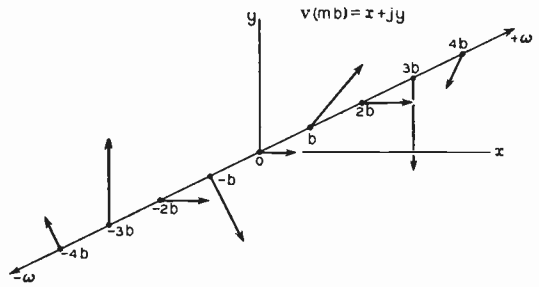


Fig. 21—A representation of the Fourier components of the local-oscillator voltage.

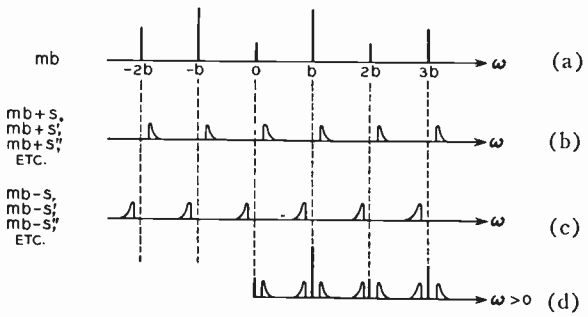


Fig. 22—Frequency components in a linear frequency converter: (a) local oscillator, (b) small-signal spectrum, (c) conjugate small-signal spectrum, (d) combined spectrum for positive frequencies.

In circuit design, as opposed to circuit analysis, there is much to be said for thinking only of positive frequencies. Then one combines Fig. 22(a) through 22(c), discarding negative frequencies, to obtain Fig. 22(d). Note that some of the signal spectra are inverted when measured in positive frequencies; these correspond to negative frequencies in Fig. 22(b) and are called "lower sidebands."

The possibility of using the sidebands around 2b, 3b, etc., is occasionally useful; the practice is known as harmonic mixing. However, the most important frequency converter applications use the three signal bands shown in Fig. 23(a). A diagram equivalent to Fig. 23(a) is shown in Fig. 23(b); the distinction between inverting and noninverting signal bands is indicated by small vertical arrows.

The first circuit example is the upper-sideband converter shown in Fig. 24(a). It is assumed that filters

[30] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—Part I. General energy relations," PROC. IRE, vol. 44, pp. 904–913; July, 1956.

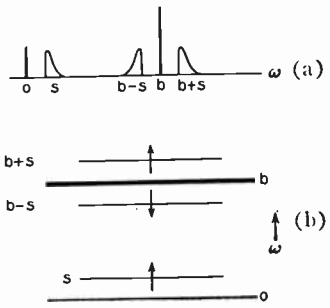


Fig. 23—(a) Frequencies involved in nonharmonic frequency conversion. (b) Diagram corresponding to (a); small arrows distinguish inverting and noninverting signal bands.

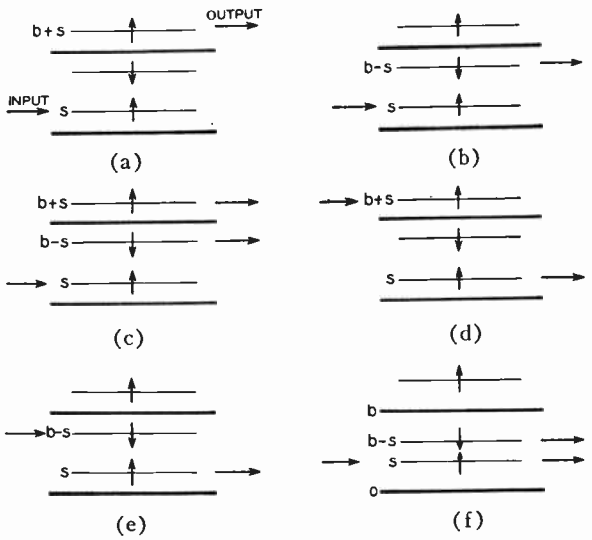


Fig. 24—Frequency relations in various types of frequency converters. (a) Upper sideband up-converter. (b) Lower sideband up-converter. (c) Double sideband up-converter. (d) Upper sideband down-converter. (e) Lower sideband down-converter. (f) Amplifier.

are used, if necessary, to prevent power leaving or entering the nonlinear capacitor at the lower sideband. Manley and Rowe[30] show that the power gain from input to output is exactly equal to the ratio $(b+s)/s$, and that the circuit is stable. The reduction of gain by series resistance has been analyzed;[5] the results can be expressed in terms of the ratios of the signal frequencies to the cutoff frequency. Some gain and bandwidth calculations on nonlinear capacitance up-converters have been discussed in the literature.[31] Another analysis puts forth the general principle that the 3-db bandwidth is about 40 per cent of the input frequency s, for upper-sideband operation.[32]

A lower-sideband up-converter is diagrammed in Fig. 24(b). Here, the power gain is $-(b-s)/s$. What does a negative power gain mean? As Manley and Rowe explain, it means that unlimited amplification is possible, but instability is also possible: for certain terminations oscillations will occur and deliver power to the terminations at both $b-s$ and s. The greater the amplification, the less the bandwidth. There is no direct theoretical noise penalty associated with this "regenerative gain," since an ideal nonlinear capacitor is not a source of noise, no matter how operated.

Hines[1] has pointed out a remarkable difference between double-sideband up-converters [Fig. 24(c)] using nonlinear capacitors and those using nonlinear resistors. The Manley-Rowe relations show that unlimited gain is possible for the nonlinear capacitors. Moreover, for a given input power, the more power delivered to the upper sideband, the more delivered to the lower sideband. Contrast this to the miserly economy of a nonlinear-resistor up-converter, wherein, to obtain maximum output at one sideband, power must be reflected at the other sideband.

An upper-sideband down-converter, shown in Fig. 24(d), produces a stable power loss of $s/(b+s)$ when a nonlinear capacitor is used, so nonlinear resistors are preferred in this type of circuit. The lower-sideband down-converter shown in Fig. 24(e) can give unlimited gain, but is not a generally satisfactory circuit when large frequency ratios are involved. The basic reason is the following situation. For given terminations, the gain in such a down-converter is less by $(s/(b-s))^2$ than the gain in the reverse direction; *i.e.*, with the circuit operating as a lower-sideband up-converter [Fig. 24(b)].

One kind of nonlinear-capacitor amplifier is diagrammed in Fig. 24(f). Unlimited gain is possible; the theoretical amplitude-gain-bandwidth product is approximately equal to the frequency being amplified.[32] As shown, the output is at two frequencies, the original and one generated by frequency conversion. The new frequency can be discarded, but it is usually wise to send it on to the next stage. The most obvious reason for preserving the new signal is that it and the amplified signal at the original frequency together determine the original information regardless of possible fluctuations in the local oscillator. In addition, a single stage of amplification by a lumped nonlinear capacitor is approximately bilateral. To obtain unidirectional gain, one might hope to arrange nonlinear capacitors in sequence in a transmission line (or in a single piece of silicon, as shown in Fig. 15), with a directional phasing of the local-oscillator supply to the successive diodes. This scheme fails if only one signal frequency band is transmitted from diode to diode, for then each diode merely presents to the signal a negative impedance that is independent of the local-oscillator phase.

Practically, of course, it is much easier to transmit both signal bands than to stop one of them. A UHF traveling-wave amplifier using four nonlinear capacitors has been built that gives low noise and unidirectional gain.[33] A few milliwatts of pump power are sufficient for this circuit.

In Fig. 24(f), no frequency values are indicated but the spacing of the lines suggests that s and $b-s$ are

[31] C. F. Edwards, "Frequency conversion by means of a nonlinear admittance," *Bell Sys. Tech. J.*, vol. 35, pp. 1403–1416; November, 1956.

[32] A. E. Bakanowski, "The Nonlinear Capacitor as a Mixer," Bell Telephone Labs., Second Interim Rep. on Task 8, (crystal rectifiers), Signal Corps Contract No. DA-36-039-36-5589; January 15, 1955.

[33] R. S. Engelbrecht, private communication.

approximately equal to each other. This particular frequency relation has the advantage that single-tuned circuits can be used to tune both s and $b-s$. In the process of amplifying s it is not only possible but also necessary that an amplified signal emerge at an inverting frequency such as $b-s$. Therefore $b-s$ must be terminated with an "idler" impedance capable of absorbing power; *i.e.*, having a resistive component. Depending upon its physical nature, the idler termination may be a source of noise. In a receiver, one might use the antenna as an idler. If the effective source temperature is low, idler noise will be low; if not, a low-noise receiver is of little value. Another possibility is to make the idler frequency much higher than the signal frequency; then the idler noise is deamplified by the frequency ratios $(b-s)/s$.[34] Idler noise can be obviated if systems considerations permit simultaneous reception of signals at s and $b-s$. On this basis, a 3-db noise figure, with 35 db of gain, has been measured for a 6-kmc amplifier using a diffused silicon nonlinear capacitor and a few hundred milliwatts of 12-kmc pump power.[35] For signal-side-band use, one would have a noise figure of 4.5 db to 6 db, depending on whether the idler termination is noiseless or exhibits room-temperature thermal noise.

This experimental result can surely be improved by further development, but it is already considerably better than a microwave superheterodyne receiver using point-contact nonlinear resistors, which might give an over-all noise figure of 6.5 db in converting 6000 mc to 30 mc. It should be noted, of course, that point-contact noise figures have been bettered at certain frequencies with traveling-wave electron-beam tubes; *e.g.*, less than 4 db at 3000 mc.

Nonlinear resistors, either point-contact diodes or some future junction devices, will continue to have some uses. They are generally to be preferred when a high frequency must be converted to a much lower frequency. In such down-converters, nonlinear capacitors are difficult to stabilize. An important use for down-converters is in frequency standards; one can obtain a signal whose frequency is the difference between some standard high-frequency oscillator (such as an ammonia maser) and a high frequency which is to be compared with the standard. As long as a device can be fairly described as a nonlinear resistor, its behavior is independent of frequency. Therefore, nonlinear resistors could be exceedingly broad-band devices. This potentiality has not been realized, because the usual point-contact nonlinear resistor contains a cat's-whisker with sufficient inductance to limit the bandwidth.

Up-converters accept a signal at a given frequency and have an output at a very much higher frequency. They are typically found in transmitting modulators. In up-converters, it is possible to use p-n junction nonlinear capacitors to get gain and nevertheless have unconditionally stable circuits. In addition to amplification, the p-n junction has the advantage of good power-handling capability. Bonded germanium diodes and diffused silicon diodes have been found satisfactory in transmitting modulators.[10]

Amplifying up-converters may also be used in low-noise receivers. Up-converters employing nonlinear capacitors have demonstrated lower noise figures than are commonly obtained with electron tubes. For example, an up-converter from 460 mc to 9375 mc has been built with 9 db of gain and a 2-db noise figure.[36]

Junction diodes can be used in frequency shifters to generate carrier frequencies for the several channels of a microwave relay system.[10] Having generated one stable carrier frequency by harmonic generation from a low-frequency oscillator, one can obtain a neighboring carrier frequency by mixing the first with an accurate low frequency. The reliability and power-handling ability of p-n junction nonlinear capacitors favors their use in this application.

IX. Control of Microwave Power

Microwave circuit elements whose impedances can be altered enable one to vary the transmission of microwave power or its distribution to branches of the circuit. It may not matter if the impedance change is resistive or reactive. If a reactive change is satisfactory, the nonlinear capacitor has a power-handling advantage over point-contact nonlinear resistors, for two reasons. One is the larger area of junction nonlinear capacitors. The other is the fact that a good nonlinear capacitor absorbs but little of the power that it controls. There is no known limit on the speed of switching by either nonlinear resistors or nonlinear capacitors. This statement may seem extraordinary in comparison, for example, with the present limitations of ferrite switches, but its truth is evident if one considers that a switch with either type of diode is just a kind of frequency converter.

If a relatively slow (*e.g.*, 1 mc) variation of the microwave power is all that is required, the p-i-n "variable resistor" diode permits the construction of broad-band switches and attenuators. The p-i-n diode has the advantage of even larger size than the p-n junction nonlinear capacitor suitable for the same frequency (excepting multiple-junction or series-parallel nonlinear capacitors). Despite large sizes possible with the p-i-n structure, one must be careful when it is used with large powers, because, included among its wide range of resistive values, there is usually one resistance that matches well enough to absorb a large fraction of the available power. When the available microwave power exceeds the allowable dissipation of the p-i-n diode, one must make sure that the diode does not exhibit the matching impedance except for a short time during a switching operation.

[34] S. Bloom and K. K. N. Chang, "Theory of parameteric amplification using nonlinear reactances," *RCA Rev.*, vol. 18, pp. 578–593; December, 1957.

[35] G. F. Herrman, M. Uenohara, and A. Uhlir, Jr., "Noise figure measurements on two types of variable reactance amplifiers using semiconductor diodes," this issue, p. 1301.

[36] G. T. Knapp of the New York Bell Telephone Co. collaborated with the author in obtaining these measurements.

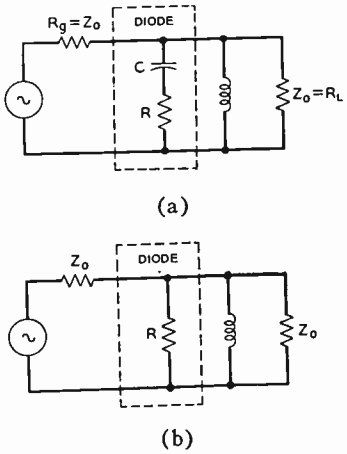


Fig. 25—Equivalent circuits of a diode transmission-line switch. (a) Reverse bias (low-loss condition), (b) forward bias (high-loss condition).

Let us consider what can be done in switching microwave power by a simple circuit in which a nonlinear capacitor is shunted across a transmission line of characteristic impedance Z_0. It will be assumed that an inductor is shunted across the line to tune for maximum transmission when the diode exhibits its minimum capacitance, as in Fig. 25(a). When the diode is biased appreciably in the forward direction, the capacitance becomes very large and the effective impedance of the diode is just its series resistance, as shown in Fig. 25(b).

The insertion loss of a shunt element of admittance Y is

$$\text{Insertion loss} = (1 + \tfrac{1}{2}Z_0Y)^2. \tag{20}$$

At the band center, the insertion loss under reverse bias is

$$\text{Insertion loss, reverse bias} = \left\{1+\frac{Z_0}{2R_s}\cdot\frac{1}{1+(f_c/f)^2}\right\}^2, \tag{21}$$

where f_c is the cutoff frequency defined by (3). For forward bias, the effect of the tuning inductor is usually negligible and the insertion loss is

$$\text{Insertion loss, forward bias} \approx \left\{1+\frac{Z_0}{2R_s}\right\}^2. \tag{22}$$

The fraction of the incident power dissipated in the series resistance is

$$\frac{P_{\text{dissipated in diode}}}{P_{\text{available}}} = \frac{Z_0/R}{1+(f_c/f)^2} \div \text{insertion loss} \tag{23}$$

for reverse and

$$\frac{P_{\text{dissipated in diode}}}{P_{\text{available}}} = \frac{Z_0}{R} \div \text{insertion loss} \tag{24}$$

for forward bias.

For an example, a calculation is given for 6-kmc switching with a diode of 100-kmc cutoff frequency. A reverse bias insertion loss of 0.5 db permits a forward bias insertion loss of 25 db. In reverse bias, 11 per cent of the incident power would be dissipated in the diode; in forward bias, 10 per cent. The loaded Q under reverse bias is less than unity, so the switch is quite broad band.

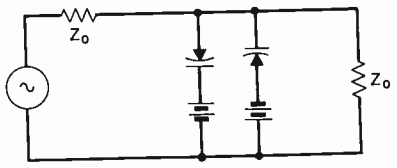


Fig. 26—A protective limiter circuit using two nonlinear capacitors.

Still better performance is possible at lower frequencies. For the same diode at 400 mc, a reverse bias loss of 0.2 db would permit, theoretically, a forward bias loss of 58 db and a loaded Q less than 2. The corresponding dissipated powers would be 2.6 per cent and 0.25 per cent of the incident power.

The reverse bias calculations must be modified unfavorably when the signal voltage is large, because the capacitance is nonlinear and the breakdown voltage must not be exceeded. However, the power that can be controlled will still be much larger than the allowable power dissipation of the diode(s).

On the assumption that thermal conduction through the semiconductor is the determining process the allowable dissipation P per diode can be estimated from

$$\begin{aligned} P &\approx 1.5\kappa d\Delta T \quad &&\text{for mesa diodes} \\ P &\approx 4\kappa d\Delta T \quad &&\text{for dimple diodes} \end{aligned} \tag{25}$$

where d is the junction diameter, κ is the thermal conductivity, and ΔT may conservatively be taken as 100°C for silicon diodes in contact with a heat sink near room temperature. For example, these relations give 1 watt for a 0.003-inch diameter mesa diode and 5 watts for a 0.005-inch diameter dimple diode, in good accord with experience.

The limiter circuit shown in Fig. 26 is closely related to the switch. Again, the diodes' capacitance can be tuned for good low-level transmission. High-level signals are clipped. The bias batteries may sometimes be eliminated. One application of such a limiter is in maintaining a fairly constant output amplitude from a variable-frequency oscillator; the accuracy might be improved by a feedback circuit that adjusts the bias voltage in accordance with a power monitor. The limiter can also be used as a protective circuit. Since most of the incident high-level power is reflected, a limiter of this kind can protect against powers that exceed by many times the allowable dissipation of the diodes.

X. Harmonic Generation

A nonlinear device generates harmonics when excited by a single-frequency generator. The nonlinear resistance of point-contact rectifiers has long been used for this purpose. In this section, theoretical and experimental evidence is given in support of the idea that nonlinear capacitors should be much better harmonic generators. Harmonic generation seems to be an expedient way of obtaining millimeter waves. Another application

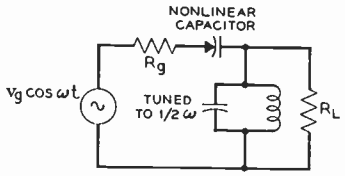


Fig. 27—Subharmonic generator.

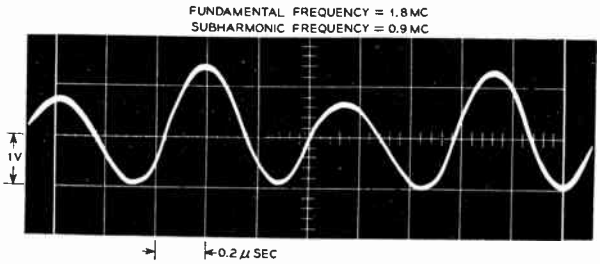


Fig. 28—Waveform observed across tank circuit of Fig. 27.

is in frequency standards where a low-frequency crystal-controlled oscillator output is multiplied to obtain standard microwave frequencies. In either of these applications the most serious limitation is the weak output obtainable with present crystal rectifiers.

It is expected that much higher efficiency in harmonic generation can be obtained with nonlinear capacitors. Ideally, such a diode cannot convert any of the incident power into dc power, nor can it dissipate any of this power. Hence, if it is possible to put an ideal lossless filter between the generator and the diode, passing only the fundamental, and to have another lossless filter that permits only the desired harmonic to leave by way of the output, then it should be possible to get nearly perfect efficiency in harmonic generation.[30] Since *p-n* junction nonlinear capacitors can be made with much better power-handling capability than point-contact diodes, as well as better efficiency, they should indeed make superior harmonic generators.

Some experiments have been very encouraging even though they made use of miscellaneous filters and tuning elements that happened to be available. Diffused silicon diodes of the mesa type were used. For example, 58 mw of available power at 430 mc was doubled to give 17 mw of 860 mc, a "conversion loss" of 5 db. In tripling, 30 mw of 330 mc gave 5 mw of 990 mc, or a 7-db conversion loss. Three hundred milliwatts of 1200 mc gave 0.4 mw of 8400 mc (29 db down to 7th harmonic).

XI. Subharmonic Generation (Frequency Division)

In frequency measurement and the establishment of frequency and time standards, it would be valuable to be able to divide a given frequency by an integer. From the study of differential equations, such as the Matthieu equation, it is known that a time-varying capacitance may be used to divide a frequency by 2; that is, to generate the one-half harmonic.[37]

This conclusion is also suggested by the small-signal analysis of frequency converters. In the "amplifier" scheme diagrammed in Fig. 24(f), it is possible to generate spontaneous oscillations at any pair of frequencies s and $b-s$. If the signal frequency s is just equal to one-half of the beating oscillator frequency b, then the lower-sideband frequency $b-s$ is equal to s. Hence, the two signal frequencies can both be tuned by a single resonant circuit. This fact makes it particularly easy to generate oscillations at $b/2$. Indeed, it is found experimentally that there is a distinct tendency to lock in at the one-half harmonic when a singly-tuned resonant circuit is approximately tuned to this frequency.

An experimental test at low frequency was made in the circuit shown in Fig. 27, where an alloy germanium transistor was used as a low-frequency nonlinear capacitor by tying together the emitter and collector. The resulting oscillogram is shown in Fig. 28.[38] Because the resonant circuit was not infinitely sharp, an appreciable amount of the fundamental frequency appears at the output. It is clear from this figure that an elementary description of one-half harmonic generation is to say that "alternate cycles are different." Similar waveforms have been observed on a traveling-wave oscilloscope with 450-mc input. Generation of the one-half harmonic of 12 kmc is easily observed in the apparatus used as a 6-kmc negative resistance amplifier.

For a fixed fundamental oscillator, there are two possible phases of subharmonic oscillation. Ordinarily, the phase is determined by chance. However, it has been suggested that the two phase possibilities could be used in computers to represent binary digits; for this application, it is necessary to devise methods of establishing the desired phase.[39]

XII. Voltage-Tuned Circuits

Direct applications for the voltage-dependent capacitance of *p-n* junctions are to be found in varying the tuning of oscillators, amplifiers, and filters.[8,40] The capacitance presented to *small* high-frequency signals depends only on the dc bias voltage applied to the junction.

The high-frequency Q of a nonlinear capacitor is determined by the series resistance and is a maximum when the capacitance is a minimum; *i.e.*, at the maxi-

[37] L. Brillouin, "Wave Propagation in Periodic Structures," McGraw-Hill Book Co., Inc., New York, N. Y., sec. 45; 1946.

[38] A. Uhlir, Jr., "Possible Uses of Nonlinear Capacitor Diodes," Bell Telephone Labs., Eighth Interim Rep. on Task 8, Signal Corp. Contract No. DA-36-039-S65589; July 15, 1956.

[39] J. von Neuman, U. S. Patent No. 2,815,488; December, 1957.

[40] W. Y. Pan and O. Romanis, "Automatic Frequency Control of Television Receivers Using Junction Diodes," in "Transistors I," RCA Labs., Princeton, N. J., pp. 598–608; 1956.

mum reverse voltage. This maximum Q is given by f_c/f, where f_c is the cutoff frequency defined in (1) and f is the frequency of use. At 1 kmc, a diffused silicon junction with an f_c of 100 kmc would have a Q of 100 at -7 volts bias. It would have two and one-half times as much capacitance at zero bias and accordingly would have there a Q of 40. Use at slight forward bias is not excluded, but when appreciable forward current flows, shunt conductance is added to the series resistance losses.

Since the dc current is very small throughout the generally useful voltage range, the dc power required to maintain the desired bias is exceedingly small: 10^{-9} watts, for example, with silicon junctions. Accordingly, low-power feedback circuits can be used to tune an oscillator in response to an error signal.

The high-frequency voltage must be small compared to the relative change of capacitance with voltage, if the effective capacitance is to be independent of signal level (and if the diode is not to become active and perhaps break into oscillation). It is sometimes possible to put the diode at a low-impedance point of a tuned circuit. Otherwise, several diodes can be connected in series (they may be in parallel at dc to keep the control voltage small) or the multiple junction structure of Fig. 13 can be used.

For reverse biases, the capacitance of graded silicon junctions is remarkably independent of temperature, as shown in Fig. 29.[12] Also, for reverse biases, the relative variation of capacitance with voltage is not sensitive to variations in the impurity gradient. These facts make control of capacitance to within 1 per cent seem feasible. Present production techniques do not permit satisfactory yields of single junctions selected to such close tolerances. However, it is easy to combine two or more junctions, after measurement, to obtain a desired capacitance. "Channels" caused by surface charges can produce reversible changes in junction capacitance. These changes are ordinarily negligible at high frequencies, because the series resistance of the channel is usually high. For applications requiring exceptional long-term stability of capacitance, the dimple diode of Fig. 14 affords protection against surface effects.

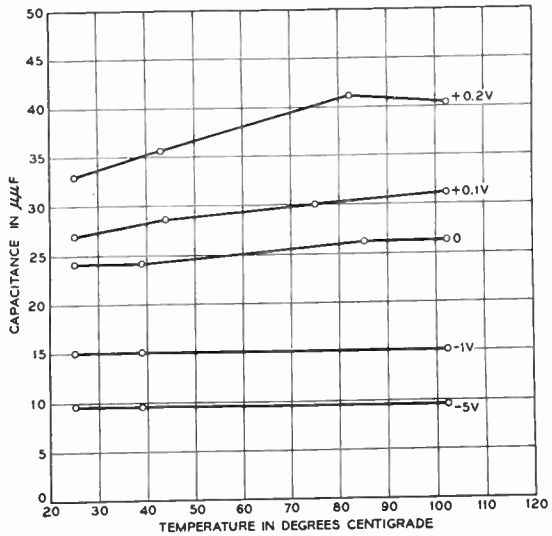


Fig. 29—Capacitance of graded junction in silicon, as a function of temperature and bias. Gradient is 10^{23} cm^{-4}.

XIII. Conclusions

Low-noise UHF and microwave amplification can be obtained at room temperature with the nonlinear capacitance of p-n junction diodes of special but simple design. Nonlinear capacitors are also capable of performing many other circuit functions, with the exception of rectification. P-n junction diodes are small, reliable, and reproducible. They can control substantial amounts of high-frequency power while consuming very little control power.

XIV. Acknowledgment

The author wishes to thank his colleagues for permission to quote their results in advance of publication. Also, R. M. Ryder has made many specific and helpful suggestions.

But for the patience of the Bell Telephone Laboratories and the U.S. Army Signal Corps, nonlinear capacitors would not have been refined to the point where low-noise microwave amplification could be proved.

New Concepts in Microwave Mixer Diodes*

G. C. MESSENGER†, ASSOCIATE MEMBER, IRE

Summary—**Recently developed techniques in three broad areas have been drawn upon to help solve the problem of improving microwave mixer diode performance. Specifically examined were the advantages offered by: 1) pertinent processes from the latest transistor technology, 2) new semiconductor materials, notably gallium arsenide, indium arsenide, and indium antimonide, and 3) cooling.**

These techniques and their theoretical bases are described, and supporting experimental evidence is presented. Use of these techniques can extend the frequency range in which detector sensitivity is good by an order of magnitude, from about 10^5 mc to 10^6 mc. In some cases, the already good sensitivity below 10^4 mc can be improved by an order of magnitude (from a receiver temperature of 2000°K to 200°K). In addition, burnout resistance in radar-detector applications can be improved by an order of magnitude.

The technology described is also applicable to a number of other microwave-detector problems, *e.g.*, those encountered in various types of video receivers, these applications are dealt with to some extent.

INTRODUCTION

THE basic theories underlying microwave mixer diode operation have been summarized by Torrey and Whitmer.[1] The original point-contact silicon rectifier that they described has been continuously improved upon and has been applied to additional frequency bands.[2] Its sensitivity is now close to the theoretical limit possible with present techniques; further major improvements will derive from exploitation of new approaches rather than from refinement of conventional technology.

The work of Torrey and Whitmer shows that sensitivity can be improved by reducing the series resistance (R_s) and/or by reducing the series resistance-barrier capacitance product (R_sC_b). Messenger and McCoy[3,4] have developed a figure of merit for the semiconductor base material of the mixer crystal based on the improvement in sensitivity that can be achieved by reducing the value of R_sC_b. The figure of merit is:

$$a\epsilon^{1/2}/N^{1/2}b$$

where a is the whisker-contact radius, ϵ is the semiconductor dielectric constant, N is the majority carrier density, and b is the majority carrier mobility.

This figure of merit applies only to nondegenerate semiconductors in the extrinsic conduction range. It enables one to choose among semiconductor materials (the smaller the figure of merit, the more promising the material), and to determine the optimum impurity-doping density. The use of this figure was instrumental in the development of a germanium X-band crystal diode[5] with a broad-band noise figure of 4.5 db, which is very close to the theoretical limit for this type of structure.

There are three new approaches in diode-design procedure that can provide further improvement: applying transistor technology (particularly small-area contacts and geometry control), using new materials having lower figures of merit, and cooling the mixer crystal (with or without concurrent cooling of the antenna).[6]

Transistor technology can be applied in the following areas: 1) increasing geometrical precision with surface-barrier techniques,[7] 2) jet etching with infrared thickness control, 3) diffusion techniques, and 4) micro-alloying techniques.[8] These techniques can provide the small geometry required for efficient conversion at high frequency. They are essential in making the change from the present structure, a point contact on a semi-infinite semiconductor base, to a "micro-etch" structure, a small-area point contact on a semiconductor membrane of small thickness compared with the contact radius. This micro-etch structure reduces both R_s and R_sC_b by an order of magnitude, making it possible to extend the upper frequency limit for good conversion to 10^6 mc (330 microns) and thus bridging the gap between microwave and infrared detectors. Small-area contacts permit an order-of-magnitude reduction in the $1/f$ noise power spectrum of the diode, a useful reduction for video receivers or doppler receivers. The small-area contact with a low value of R_s is ideally suited for applications involving certain forms of parametric conversion.[9–11] The micro-etch structure can provide much higher power dissipation than the present structure because

* Original manuscript received by the IRE, March 10, 1958; revised manuscript received, April 15, 1958.

† Res. Div., Philco Corp., Philadelphia, Pa.

[1] H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, M.I.T. Rad. Lab. Ser., vol. 15, 1948.

[2] There are several laboratories active in this field, including Sylvania, Microwave Associates, Bomac, and Kemtron, where such series as the 1N23, 1N23B, 1N23C, 1N23D, and 1N23E, each a minor improvement over its predecessor, have been developed. Similar work has gone on concurrently in British laboratories such as the Telecommunication Research Establishment.

[3] G. C. Messenger and C. T. McCoy, "A low noise-figure microwave crystal diode," 1955 IRE CONVENTION RECORD, pt. 8, pp. 68–73.

[4] G. C. Messenger and C. T. McCoy, "Theory and operation of crystal diodes as mixers," PROC. IRE, vol. 45, pp. 1269–1283; September, 1957.

[5] This unit was a selected 1N263.

[6] G. C. Messenger, "Cooling of microwave mixers and antennas," IRE TRANS ON MICROWAVE THEORY AND TECHNIQUES, vol. 5, pp. 62–68; January, 1957.

[7] W. E. Bradley, *et al.*, "The surface-barrier transistor, Part I–V," PROC. IRE, vol. 41, pp. 1702–1720; December, 1953.

[8] A. Rittmann, G. C. Messenger, R. A. Williams, and E. Zimmerman, "Microalloy transistors," IRE TRANS. ON ELECTRON DEVICES, vol. 5, pp. 49–54; April, 1958.

[9] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—Part I. General energy relations," PROC. IRE, vol. 44, pp. 904–913; July, 1956.

[10] A. Uhlir, "Two terminal p-n junction devices for frequency conversion and computation," PROC. IRE, vol. 44, pp. 1183–1191; September, 1956.

[11] H. Welker and H. Weiss, "Group III–Group V compounds," in "Solid State Physics," edited by F. Seitz and D. Turnbull, Academic Press, Inc., New York, N. Y., vol. 3; 1957.

considerably larger contact areas can be used without increasing R_sC_b.

New semiconductor materials with lower values of $a\epsilon^{1/2}/N^{1/2}b$[11] (in the nondegenerate case) can increase the sensitivity of both micro-etch and conventional structures. Some new materials, particularly GaAs, can operate at higher ambient temperatures; other materials, notably InSb, can operate at lower (liquid nitrogen) temperatures. The great variety of properties to be found in the growing number of potentially useful diode materials holds promise of higher sensitivity and greater temperature range.

In preamplifiers operating in the megacycle frequency range, the noise in microwave diodes is primarily shot noise and thermal noise, which vary directly with Kelvin temperatures.[12] Therefore, cooling the mixer crystal reduces the over-all receiver noise. If the antenna looks at a background temperature of 290°K, the reduction is limited to several db. But if the antenna, by virtue of high directivity, looks at a low-background temperature, the combined effect of a cooled crystal and low-background temperature can increase sensitivity by an order of magnitude.

The Micro-Etch Structure

In attempting to improve the conventional diode structure by reducing the contact radius, it is found that eventually the contact becomes too weak, mechanically, for device use; the smallest practical contact radius is about 4×10^{-4} cm. To overcome this problem it was suggested that a wedge-shaped contact be used,[13] but this suggestion proved impracticable. The final solution, described in this paper, involves reducing the width of the semiconductor blank until it becomes a membrane whose thickness is small compared with the diameter of the point contact; this micro-etch structure affords another order-of-magnitude improvement in sensitivity by means of geometry control.

If the micro-etch structure, Fig. 1, is compared with the conventional point-contact structure, Fig. 2, it is seen that the micro-etch structure offers the possibility of reducing the series resistance without a corresponding increase in capacitance. The design theory of the two structures is compared with respect to reducing the figure of merit, ωR_sC_b, as follows:

Conventional Structure

$$R_s = \frac{1}{4aNqb} \tag{1a}$$

$$C_b = \pi a^2\sqrt{\frac{\epsilon qN}{2(\Phi_0 - V)}} \tag{2a}$$

$$\omega R_sC_b = \frac{\omega\pi a\epsilon^{1/2}}{4\sqrt{2(\Phi_0 - V)qNb}} \tag{3a}$$

[12] Messenger, *op. cit.*
[13] Torrey and Whitmer, *op. cit.*

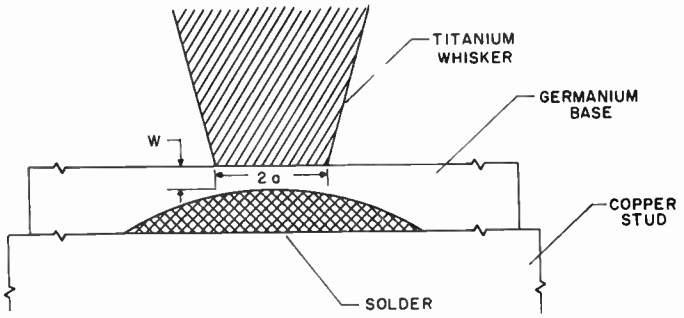


Fig. 1—Micro-etch mixer crystal.

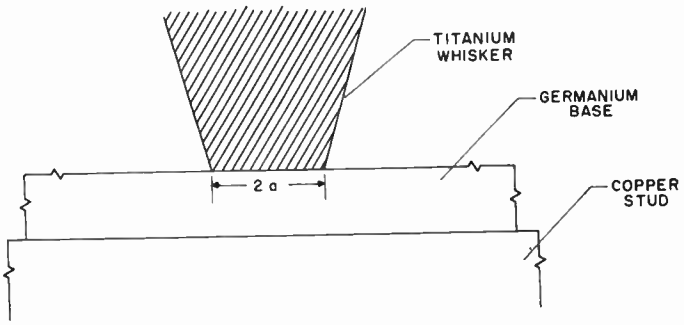


Fig. 2—Conventional 1N263 mixer crystal.

Micro-etch Structure

$$R_s = \frac{W}{\pi a^2Nqb} \tag{1b}$$

$$C_b = \pi a^2\sqrt{\frac{\epsilon qN}{2(\Phi_0 - V)}} \tag{2b}$$

$$\omega R_sC_b = \frac{\omega W\epsilon^{1/2}}{\sqrt{2(\Phi_0 - V)qNb}} \tag{3b}$$

Here ω is angular frequency, q is electronic charge, Φ is contact potential, W is base width, and V is voltage across the barrier.

For a given operating voltage $(\Phi_0 - V)$, and a given frequency ω, the dependence on geometry and semiconductor properties can be put into a figure of merit, M:

$$M = \frac{a\epsilon^{1/2}}{N^{1/2}b} \tag{4a}$$

$$M = \frac{W\epsilon^{1/2}}{N^{1/2}b} \tag{4b}$$

Notice that in the micro-etch structure the thickness of the base, W, replaces the radius of the contact, a, as the determining geometrical factor. By using electrochemical techniques, W can be reduced at least a factor of ten further than a can be reduced in the conventional structure. This allows either an order of magnitude more burnout resistance at a given frequency with a given noise figure, or an order of magnitude higher frequency for a given noise figure and a given burnout

resistance, or an improvement in noise figure at a given frequency with a given burnout resistance, at the designer's option.

Using present technology the minimum practicable value of W is about 4×10^{-5} cm as compared with the minimum practicable value of 4×10^{-4} cm for a. The lower limit of 4×10^{-5} cm is determined by infrared thickness control process.

The fabrication of the micro-etch structure can be accomplished in several ways; a typical method for germanium diodes is as follows.[14,15] The semiconductor blank is polished on one side, exactly as in the conventional method, and is then placed in a jet-etching station where a pit approximately 2×10^{-2} cm in diameter is etched from the unpolished side. The membrane thickness is controlled by infrared transmission to a value of about 10^{-4} cm. A suitable metal is then electrochemically plated in the pit. Next comes the micro-alloying step. The penetration of the micro-alloy is set for 5×10^{-5} cm, leaving a residual membrane thickness of 5×10^{-5} cm. The mechanically strong membrane produced by the micro-alloy step is then soldered to a standard diode stud. From this point on the assembly operation employs standard diode-fabrication techniques.

The infrared thickness control determines the membrane thickness efficiently in the range from 5×10^{-4} to 4×10^{-5} cm. The micro-alloy process then acts as a vernier in obtaining the final desired thickness; the alloying depth in this process can be accurately controlled in the range from 5×10^{-6} cm to 5×10^{-5} cm. Possible substitute methods of obtaining the final thickness include a controlled diffusion process and a timed-polish process on the face of the blank.

For conventional 1N263 type germanium diodes the measured values of R_s vary from 4 to 10 ohms; the average is about 6 ohms. Early experimental micro-etch versions of the 1N263 type diode have exhibited values of R_s varying from less than 1 ohm to 4 ohms; the average is about 2 ohms. One of the practical problems involved is to make certain that the resistance between the blank and its stud, and the resistance in the joint between the whisker and its stud, are both much less than 1 ohm. (The resistance of the 0.25-cm length of 3-mil titanium whisker is considerably less than 1 ohm.)

The ultimate limitation on the reduction in R_s is twofold. The resistance of the whisker wire is of the order of several hundredths of 1 ohm at dc, and at X-band about $\frac{1}{2}$ ohm due to skin effect; it would do little good to reduce the diode R_s below the latter value. The width of the membrane cannot be reduced below approximately 2×10^{-6} because of an effect analogous to transistor punch through, *viz.*, the field, created by a potential of one volt on the whisker wire, reaches completely through the membrane. The barrier thickness, D, is given by

$$D = [\{2\epsilon(\Phi_0 - V)\}/qN]^{1/2} \quad (5)$$

It may be possible by carefully utilizing present techniques to reduce the series resistance to about 1/10 ohm at dc and $\frac{1}{2}$ ohm at X-band in the 1N263 type structure.

The inherent advantages of the micro-etch structure can be discussed in terms of several possible extensions of the state of the art. At present the conventional germanium point-contact structure has an average noise figure of 10 db at 70 kmc. The order-of-magnitude reduction in the R_sC_b product should permit the same sensitivity at 700 kmc, that is, approximately 400 microns, or very long wavelength infrared. This development therefore makes feasible sensitive detectors that will bridge the gap between present microwave detectors and long wavelength infrared detectors. At low rf frequencies (less than 10 kmc), there will be only a small improvement in sensitivity since the R_sC_b product causes only a small deterioration in sensitivity in present structures. Fig. 3 summarizes the potential

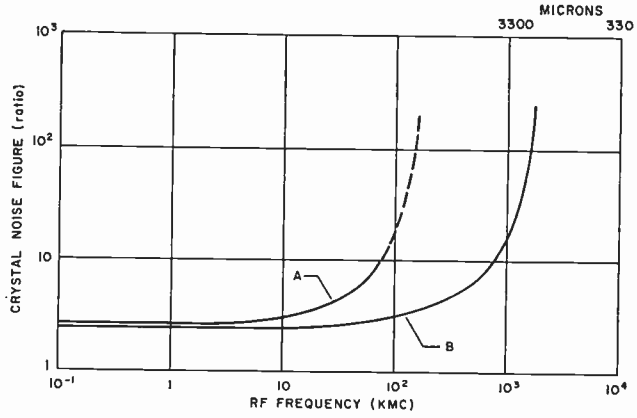


Fig. 3—Comparison of crystal noise figure vs frequency for best conventional diodes, curve A, and expected values for micro-etch diode, curve B. The frequency dependent parameter ωC_bR_s enables the same performance to be realized at a higher frequency in the micro-etch structure, since for equal values of ωC_bR_s the lower value of R_s in the micro-etch structure compensates the higher value of ω. The better performance at low frequency is due directly to the reduction of R_s.

improvement in crystal noise figure (as a function of frequency) that can be derived by using the micro-etch structure rather than the standard structure. The improved curve (curve B) is obtained from curve A by calculating the improvement due to the lower value of R_s at low frequencies, and by assuming equivalent performance for the same values ωC_bR_s at high frequencies, then the reduced value of R_s makes possible a proportionate increase in operating frequency.

The micro-etch structure makes possible the use of small-area contacts in microwave diodes. The present state of the art in transistor technology makes feasible

[14] Bradley, *et al.*, *op. cit.*
[15] Rittmann, Messenger, Williams, and Zimmerman, *op. cit.*

contacts as small as 10^{-3} cm in diameter.[16] This is a factor of two or three too large to use in the conventional 1N263 structure; the added capacity would increase the R_sC_b product and seriously degrade performance. With the micro-etch structure, however, the larger value of capacity can be offset by the smaller value of series resistance, resulting in the same over-all R_sC_b product and approximately the same performance.

The principal advantages of the micro-etch structure, then, are: 1) assuming that a good heat sink is provided, the resistance to burnout will be an order of magnitude larger due to the larger contact cross section, 2) a large reduction in the R_sC_b product is obtained, 3) the new type contacts that can be used (transistor contacts of the surface-barrier, micro-alloy and alloy-junction types) have a much reduced noise-power spectrum in the lower-frequency ranges where $1/f$ noise predominates. For example, actual noise measurements show that the noise power from a 1N263 diode at 10 kc is about 20 db higher than the noise power from an alloy-junction contact of the type used in a 2N207B audio transistor. A mixer diode incorporating a low-flicker-noise contact should provide a large improvement in sensitivity in doppler-type receivers or in video receivers, where the noise-power spectrum of the diode at low frequency is important.

The advantages of capacitive mixing have been pointed out by Uhlir.[17] In one proposed system using a capacitive upconverter having gain, and followed by a standard crystal mixer, the R_sC_b product of the upconverter diode again provides a good figure of merit. Thus, the micro-etch structure will facilitate the fabrication of good capacitive mixer diodes.

In summary, the performance of just about every microwave diode type can be improved by a reduction in the R_sC_b product. In addition a number of new possibilities for the use of microwave diodes will open up as a result of a substantial reduction in the R_sC_b product.

New Materials

In the past several years a host of new semiconductor materials have been prepared in various laboratories, primarily for application to transistor devices. Several of these are promising as basic materials for microwave diodes.[18] Three possibilities for fabricating better microwave devices follow from using these materials: 1) those with lower values for the figure of merit promise better performance, 2) those with higher gap energies may possibly work efficiently at higher ambient temperatures, 3) some of them may rectify efficiently at very low temperatures where they will be useful in cooled mixer applications.[19]

Detailed experimental information on the variation of mobility with impurity density is not available for most semiconductors; therefore, the denominator of the figure of merit is not presently ascertainable. However, in general, the figure of merit will roughly correlate with the mobility (without consideration of its variation). On this basis Table I is presented in which are listed a number of semiconductor materials roughly in order of decreasing figure of merit as microwave diode materials. It is recognized that the mobility in many of these new materials will increase with time as the state of the art in their preparation improves, so that this table must be regarded as a qualitative indicator.

The high energy-gap materials may potentially be capable of maintaining their sensitivity at high ambient temperatures. It should be realized that the temperature dependence of a point-contact diode is determined by the contact potential between the whisker and the semiconductor so that a large energy gap is a necessary though not sufficient requirement for high-temperature operation.

TABLE I

Semiconductor Materials Listed in Order of Estimated Utility for Making Microwave Diodes

Type	Semiconductor	Mobility	Energy Gap
		cm²/volt sec	ev
N	InSb*	57,000	0.17
N	InAs	27,000	0.37
N	HgTe	11,800	0.4
N	HgSe	10,000	0.65
N	GaAs	4000	1.35
N	GaSb	4000	0.69
N	Ge	3600	0.72
N	InP	3500	1.25
N	Si	1400	1.09
P	Si	440	1.09

* Useful only at very low temperature since it is intrinsic at room temperature.

Specifically, it has been found experimentally[20] that germanium and silicon[21] microwave diodes exhibit the same temperature dependence even though silicon has a higher energy gap. Recent experiments in our laboratories on GaAs diodes, however, indicate that as rectifiers they are much less sensitive to temperature than either germanium or silicon diodes. One of these experimental GaAs diodes maintained good rectification characteristics up to 350°C.

The concept of operating mixer diodes at low ambient temperature to improve their sensitivity enables the use of low energy-gap semiconductors (which may be intrinsic at room temperature) with high figures of merit. Of these, InSb is very promising.

[16] Private communication from R. A. Williams, Philco Corp.
[17] Uhlir, *op. cit.*
[18] Welker and Weiss, *op. cit.*
[19] Messenger, *op. cit.*
[20] Messenger and McCoy, "Theory and operation of crystal diodes as mixers," *loc. cit.*
[21] Uhlir, *op. cit.*

TABLE II

EFFECTS OF COOLING TECHNIQUES ON EXCESS NOISE

Assumed Conditions	Excess Crystal Mixer Noise Temperature (F_x-2) 290°			Excess IF Noise Temperature $L_x(F_{if}-1)$ 290°	Antenna Noise Temperature $2(\tau_a)$ 290°	Receiver Noise Temperature τ_s 290°		
	Mixer Diode at 290°	Mixer Diode at 72.5°	Mixer Diode at 6°			Mixer Diode at 290°	Mixer Diode at 72.5°	Mixer Diode at 6°
1) Average 1N263 $F_{if}=1.5$ $\tau_a=1.0$ $L_x=3.4$ $l_x=1.2$	610°	150°		490°	580°	1680°	1320°	
2) Selected 1N263 Special Preamplifier $F_{if}=1.2$ $\tau_a=1.0$ $L_x=2.8$ $l_x=1.0$	230°	60°		160°	580°	970°	800°	
3) Calculated Micro-etch Crystal $L_x=2.5$ $l_x=1.0$ $F_{if}=1.2$ $\tau_a=1.0$	150°	40°		150°	580°	880°	770°	
4) Selected 1N263 $F_{if}=1.2$ $\tau_a=0.05$ $L_x=2.8$ $l_x=1.0$	230°	60°		160°	30°	420°	250°	
5) Calculated Micro-etch Crystal $F_{if}=1.2$ $\tau_a=0.05$ $L_x=2.5$ $l_x=1.0$	150°	40°		150°	30°	330°	220°	
6) Calculated Micro-etch Crystal $L_x=2.5$ $l_x=1.0$ $F_{if}=1.0$ $\tau_a=0.05$	150°	40°	3°	0°	30°	163°	73°	33°

COOLING OF MICROWAVE MIXER DIODES

Following McCoy[22] let us express the receiver-noise sensitivity, τ_s, of a broad-band receiver as:

$$\tau_s = (F_x - 2) + L_x(F_{if} - 1) + 2\tau_a \qquad (6)$$

where F_x-2 is the excess-mixer noise figure, L_x is the mixer conversion loss, $F_{if}-1$ is the excess IF noise figure, and τ_a is the normalized antenna temperature. The receiver temperature can then be found by multiplying τ_s by the standard reference temperature of 290°K.

It has been shown[23] that the excess-mixer noise figure is proportional to Kelvin temperature if it contains shot and thermal noise contributions. Van der Ziel[24] has shown that transistor noise at high frequencies is also composed of shot and thermal noise. In the diode case an average reduction in noise figure of about 2 db for 1N263 diodes was observed in going from 370°K to 230°K. However, recent experimental results[25] for transistors indicate a large increase in $1/f$ noise as the temperature is decreased, which may complicate the problem of obtaining low transistor noise figures by cooling techniques.

If it is assumed that the excess noise in diode converters is proportional to the ambient temperature, then for utmost sensitivity the converter should be operated at the lowest feasible ambient temperature. This is determined by the energy gap between impurity levels and the conducting or valence band. For germanium the normal N-type doping levels are about 0.01 ev below the band, and serious changes in carrier density begin to occur at about 100°K. With some redesign involving adjustment of the impurity density, however, operation at the temperature of liquid air should be feasible. The excess noise from the crystal should be reduced by a factor of about four at the liquid-air temperature. Table II, based on X-band operation with 30 mc IF, summarizes the improvements which should be attained by cooling techniques.

[22] C. T. McCoy, "Present and future capabilities of microwave crystal rectifiers," PROC. IRE, vol. 46, pp. 61–65; January, 1958.
[23] Messenger, *op. cit.*
[24] A. van der Zeil, "Theory of shot noise in junction diodes, junction transistors," PROC. IRE, vol. 43, pp. 1639–1646; November, 1955.
[25] C. A. Lee, and G. Kaninsky, "Temperature dependence of noise in transistor structure," paper delivered at the annual meeting of the American Physical Society, New York, N. Y.; January, 1958.

Conditions 1) and 2) represent the present state of the art and are experimental values. The remaining conditions are calculated on the basis of the general ideas proposed in this paper. Note that the excellent sensitivities shown for conditions 4)–6) require a cold antenna background ($\tau_a = 0.05$).

Condition 2), representing the best current technology, shows that a large percentage of the receiver noise is antenna background noise. Condition 4) should therefore be presently realizable.

On a more speculative basis, if the rectifying properties of semiconductor contacts could be maintained down toward the temperature of liquid helium, the excess mixer noise could be virtually eliminated. In addition, it is probable that the excess IF noise can also be virtually eliminated, either through the development of better low-noise tubes or perhaps by cooled transistors. In this case, as is shown in Condition 6) nearly all of the noise in the receiver is background noise, and the receiver has become nearly ideal. These possibilities must be qualified, however, by the many practical problems involved in using cooling techniques. Typical of these problems in the design of mixers and crystal cartridges which will operate well at very low temperatures. In addition, it must be noted that the necessity for maintaining cooling facilities would be a real hardship in many systems.

Receiver Applications

An excellent comprehensive summary of the present and future capabilities of Microwave Crystal Receivers has recently been published.[26] The conclusions stated in that paper are reproduced here.

> "The low-noise capabilities of present day microwave and millimeter crystal-superheterodyne receivers are determined primarily by the crystal mixer. The crystal-mixer noise, in essence, may be quantitatively related to a few fundamental physical parameters, *viz.*, rectifying contact geometry, dielectric constant, carrier concentration, carrier mobility, and temperature of the semiconductor.
>
> "For the future, improved contact geometry, better physical constants from new semiconductors, and lower temperatures promise limitless reduction in the excess noise from the crystal mixer. Therefore, the antenna noise will soon become the dominant limitation to over-all receiver sensitivity for all frequencies between 100 mc and 100,000 mc."

In addition, the following potential receiver applications are noted. For applications requiring superior burnout in crystal mixers, an area adaptation of the micro-etch structure should provide a significant improvement over present devices. For applications requiring low $1/f$ noise levels, an area adaptation of the micro-etch structure should provide improvement of at least one order of magnitude. With a high-frequency detector design, detection will be possible in the frequency range from 10^5 to 10^6 mc.

Conclusions

Three broad areas have been outlined for further improvements in microwave diodes. These areas include a basic modification in contact geometry, application of new semiconductor materials, and use of cooling. These improvements are mutually compatible and may be used in any combination for optimizing results in particular problems. Potential capabilities which may be realized through these techniques include 1) extension of the upper-frequency limit for good crystal detectors from 100,000 to 1,000,000 mc, 2) improvement of the sensitivity of present microwave receivers by another 10 db, 3) extension of the useful ambient-temperature range for mixing, 4) virtual elimination of all excess mixer noise from a receiver, and 5) increase of burnout resistance by an order of magnitude over present crystal types.

[26] McCoy, *op. cit.*

Narrow Base Germanium Photodiodes*

D. E. SAWYER†, MEMBER, IRE AND R. H. REDIKER†, ASSOCIATE MEMBER, IRE

Summary—The operation of germanium photodiodes at room temperature both as reverse-biased and photovoltaic detectors is analyzed. This analysis takes into account generation of hole-electron pairs in the base as an exponentially decreasing function of distance from the surface. General expressions are derived for the steady-state and the time-varying detector signal components. The intrinsic frequency behavior (that associated with carrier diffusion from the point of generation to the *p-n* junction) is the same for reverse-biased and photovoltaic operation. The frequency-cutoff behavior is compared with that of a homogeneous base transistor and, for the case of small loss of photogenerated carriers by surface recombination, it is shown that the cutoff frequencies are essentially the same for identical base width. Photodiodes may be useful at modulating frequencies well above this cutoff frequency if the radiation is penetrating since for this type of radiation the intrinsic frequency response does not decrease rapidly above cutoff. The equivalent circuits for both reverse-biased and photovoltaic operation are obtained as is the noise-equivalent circuit for reverse-biased operation. It is shown that at frequencies where $1/f$ noise may be neglected, for most small-signal applications reverse-biased operation is greatly superior to photovoltaic operation. With a reduction in base width the intrinsic frequency cutoff will be increased, the bulk and surface recombination loss of photogenerated carriers decreased, and the diffusion capacitance associated with the *p-n* junction in photovoltaic operation reduced. Thus, improvements in both reverse-biased and photovoltaic operation are realized. For frequencies at which $1/f$ noise sources may be neglected, photodiode operation can be determined from a knowledge of the intrinsic response, the equivalent circuit, the low-frequency noise-equivalent power, and a quantity which specifies the increase of the noise-equivalent power with frequency. Equations are developed to specify these quantities and a design is given for a narrow-base photodiode which can be operated at modulating frequencies from the low kilocycle range to above two megacycles with a noise equivalent power of 3×10^{-12} watts cycle$^{-1/2}$ ($\lambda=1.5$ μ).

Introduction

SEMICONDUCTOR photodiodes and phototransistors have been extensively reported in the literature.[1] In the present analysis we have extended previous treatments of photodiodes to include fundamental frequency limitations and the frequency behavior of the noise-equivalent power. In order to increase the frequency response of photodiodes the base region can be made narrower as is done in conventional transistors. Such a narrow-base germanium photodiode is shown in Fig. 1. In determining the frequency response of narrow-base photodiodes whose base widths are of the same order as the absorption length of the incident radiation in the semiconductor, one has to take into account the fact that the minority carriers are produced within the base region. In this respect the theory presented diverges from conventional transistor theory where all the carriers are injected into the base region at a given boundary. The effects of carrier generation within the base region on photodiode operation both as a reverse-biased detector and as a photovoltaic detector are presented.

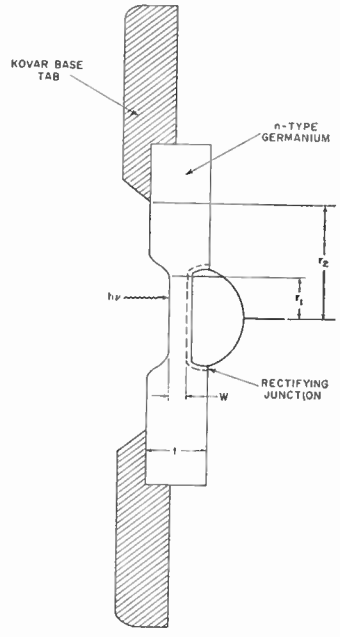


Fig. 1—Cross section of a narrow-base photodiode (not to scale).

Narrow-base germanium photodiodes have been fabricated with intrinsic response times of less than 75 mμsec. These photodevices have a very desirable and unique combination of small size, high sensitivity, very fast speed of response, and spectral range into the near infrared. They have been used with success in many applications among which are: detectors in a semipermanent storage-optical memory system;[2] research tool for the study of the decay of near infrared spectral

* Original manuscript received by the IRE, February 28, 1958. The research reported in this document was supported jointly by the Army, Navy and Air Force under contract with Mass. Inst. Tech.

† Lincoln Lab., M.I.T., Lexington, Mass.

[1] See for example:
a) S. Benzer, "The photodiode and photo-peak characteristics in germanium," *Phys. Rev.*, vol. 70, p. 105; July, 1946.
b) F. S. Goucher and M. B. Prince, "Interpretation of α-values in p-n junction transistors," *Phys. Rev.*, vol. 89, pp. 651–653; February, 1953.
c) B. J. Rothlein and A. B. Fowler, "Germanium photovoltaic cells," IRE TRANS. ON ELECTRON DEVICES, vol. 1, pp. 67–71; April, 1954.
d) G. A. Boutry and F. Desvignes, "Photodiodes and phototransistors, considered as infra-red radiation detectors," *Nuovo cimento*, suppl. to vol. 2, pp. 541–563; March, 1953; a survey with 59 references.
e) L. P. Hunter, "Handbook of Semiconductor Electronics," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 5-1 to 5-14; 1956.

[2] D. M. Baumann, "A high scanning-rate storage device for computer applications," *J. Assoc. Computing Machinery*, vol. 5, pp. 76–88; January, 1958.

components in a gas following ionization;[3] and detector in a transistorized "light pen" for high-speed computer oscilloscope readout.

The analysis presented below is for germanium photodiodes operating at room temperature. Many of the results of the analysis should be applicable, however, to other semiconducting materials and temperatures. Specific applications to α and β detectors[4] will not be discussed although many of the considerations developed are relevant.

General Analysis

The following analysis is for monochromatic radiation whose absorption in the semiconductor is characterized by a single exponential function.[5] By superposition, the results can be generalized for any input spectrum. The following assumptions have been made:

1) The characteristics of the rectifying contact are such that only carriers which are minority carriers in the semiconductor base can cross the rectifying junction.

2) The base is assumed homogeneous and extrinsic. Conductivity modulation is neglected.

3) Photogeneration in the junction space-charge region, for which case operation of the photodiode is similar in many respects to that of an ionization chamber, is neglected here.

4) Photogeneration which occurs within the heavily doped alloy-contact recrystallized layer may be neglected.

5) Carrier trapping is negligible. The decay of excess carriers in the bulk can be characterized by a single exponential, exp $-(t/\tau)$.

In this analysis the base region is assumed n type. The results are applicable, with a suitable change in notation, for a junction photodiode with a p-type base. The continuity equation including generation in the bulk is[6]

$$\frac{\partial p}{\partial t} = \frac{p_n - p}{\tau} - \frac{1}{q}\vec{\nabla}\cdot\vec{J}_p + g_p' \tag{1}$$

where p is the hole density, p_n is the equilibrium hole density, τ is the bulk-hole lifetime, g_p' is the net rate of generation of holes within the base due to photon absorption, and J_p is the hole-current density.

If the number of photons per unit area per unit time which enter the semiconductor is denoted by I and their absorption coefficient denoted by α the net generation rate of holes at a distance x from the surface is

$$g_p' = \alpha I e^{-\alpha x} \tag{2}$$

where the quantum yield for germanium has been taken equal to unity.[7] The relationship between I and the incident radiation is found to be

$$I = \frac{(1 - R)\lambda}{hc} W \tag{3}$$

where R is the surface reflection coefficient, W is the incident power per unit area of wavelength λ, h is Planck's constant and c is the speed of light. For W in watts/cm^2 and λ in microns

$$I = \frac{0.806}{q}(1 - R)\lambda W \text{ quanta/cm}^2\text{ sec} \tag{3a}$$

where q is the electronic charge.

In order to obtain both the steady-state behavior and frequency response of the photodiode let

$$I = I_0(1 + Me^{i\omega t}) \tag{4}$$

where M the degree of modulation must always be smaller than unity. Using the relation

$$J_p = -qD_p\frac{dp}{dx}, \tag{5}$$

where D_p is the minority-carrier diffusion constant, and combining (1), (2), and (4), we obtain for the one-dimensional diffusion equation

$$\frac{\partial p}{\partial t} = D_p\frac{\partial^2 p}{\partial x^2} + \frac{p_n - p}{\tau} + \alpha I_0 e^{-\alpha x}(1 + Me^{i\omega t}). \tag{6}$$

The solution of (6) may be written

$$p(x, t) = p_{dc}(x) + p_{ac}(x, t). \tag{7}$$

With $\alpha^2 \neq 1/D_p\tau$, these solutions are[8]

$$p_{dc} = Ae^{x/L} + Be^{-x/L} + p_n - \frac{I_0 e^{-\alpha x}}{\alpha D_p(1 - 1/\alpha^2 D_p\tau)} \tag{8}$$

[3] M. P. Van Yukov and L. D. Khazov, "Photoelectric method of recording the change in time of the spectra of light flashes," *Doklady Akademii Nauk SSSR*, vol. 92, pp. 523–524; December, 1953 (AEC Translation NSF-tr-158, 1953).

[4] A. V. Airapetiants and S. M. Ryvkin, "Characteristics and operative mechanism of germanium n-p alpha counters," *Sov. Phys. Tech. Phys.*, vol. 2, pp. 79–88; January, 1957.

[5] For values of the absorption coefficient as a function of wavelength, see W. C. Dash and R. Newman, "Intrinsic optical absorption in single-crystal germanium and silicon at 77°K and 300°K," *Phys. Rev.*, vol. 99, pp. 1151–1155; August, 1955.

[6] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y., p. 298; 1950.

[7] For radiation of wavelength between 1 and 2 microns the quantum yield can be taken as unity. See F. S. Goucher, "The photon yield of electron-hole pairs in germanium," *Phys. Rev.*, vol. 78, p. 816; June, 1950. For wavelengths shorter than 0.575 micron the number of generated pairs is proportional to the total energy absorbed so that in this range λ in (3) should be replaced by the constant value 0.575. See S. Koc, "The quantum efficiency of the photo-electric effect in germanium for the 0.3–2. μ wavelength region," *Czech. J. Phys.*, vol. 7, pp. 91–95; July, 1957; also J. Drahokoupil, M. Malkovska, and J. Tauc, "Quantum efficiency of the photovoltaic effect in germanium for x-rays," *Czech. J. Phys.*, vol. 7, pp. 57–65; July, 1957.

[8] The solution for p_{ac} if $\alpha^2 = 1/D_p\tau$ is the same as that given by (8) except that the last term on the right-hand side becomes

$$\frac{I_0}{2D_p} xe^{-\alpha x}.$$

and

$$p_{ac} = E \exp \frac{x(1+i\omega\tau)^{1/2}}{L} + G \exp - \frac{x(1+i\omega\tau)^{1/2}}{L} - \frac{MI_0 e^{-\alpha x} e^{i\omega t}}{\alpha D_p \left(1 - \frac{1}{\alpha^2 D_p \tau} - \frac{i\omega}{\alpha^2 D_p}\right)}, \quad (9)$$

where $L=(D_p\tau)^{1/2}$. The boundary conditions at $x=0$ for (8) and (9) are respectively taken to be

$$s(p_{dc} - p_n) = D_p \frac{dp_{dc}}{dx} \quad (10)$$

and

$$sp_{ac} = D_p \frac{dp_{ac}}{dx}, \quad (11)$$

where s is the surface recombination velocity. The boundary conditions at $x=w$ depend on the manner in which the photodiode is used. In the analysis that follows the base width w is assumed to be independent of applied voltage.

Operation as a Reverse-Biased Photodiode

The minority-carrier density $p(w)$ at $x=w$ is related to the equilibrium minority-carrier density p_n by

$$p(w) = p_n e^{qV/kT}. \quad (12)$$

If V, the junction voltage, is negative and $|V| \gg kT/q$, $p(w)$ can be taken as zero. Thus by (7),

$$p_{dc}(w) \approx p_{ac}(w) \approx 0. \quad (13)$$

Determining the constants in (9) from the boundary conditions (11) and (13) and using (5) the junction alternating-current density is found from the diffusion equation to be

$$j_{ac} = \frac{qMI_0 e^{i\omega t}}{\left(1 - \frac{1}{\alpha^2 D_p \tau} - \frac{i\omega}{\alpha^2 D_p}\right)} \left\{ \frac{(1 + s/\alpha D_p) - e^{-\alpha w}\left[(1 + s/\alpha D_p)\cosh\frac{w}{\xi} + \left(\frac{s\xi}{D_p} + \frac{1}{\alpha\xi}\right)\sinh\frac{w}{\xi}\right]}{\cosh\frac{w}{\xi} + \frac{s\xi}{D_p}\sinh\frac{w}{\xi}} \right\} \quad (14)$$

where $\xi = L(1+i\omega\tau)^{-1/2}$. For the sake of brevity, we shall write (14) as

$$j_{ac} = qMI_0 e^{i\omega t} F(\omega, \alpha), \quad (14a)$$

although as can be seen, $F(\omega, \alpha)$ is a function of D_p, s, τ, and w as well as ω and α.

The solution for j_{dc} may be written

$$j_{dc} = j_0 + qI_0 F(\alpha) \quad (15)$$

where

$$j_0 = qp_n s \left\{ \frac{\cosh\frac{w}{L} + \frac{D_p}{sL}\sinh\frac{w}{L}}{\cosh\frac{w}{L} + \frac{sL}{D_p}\sinh\frac{w}{L}} \right\} \quad (16)$$

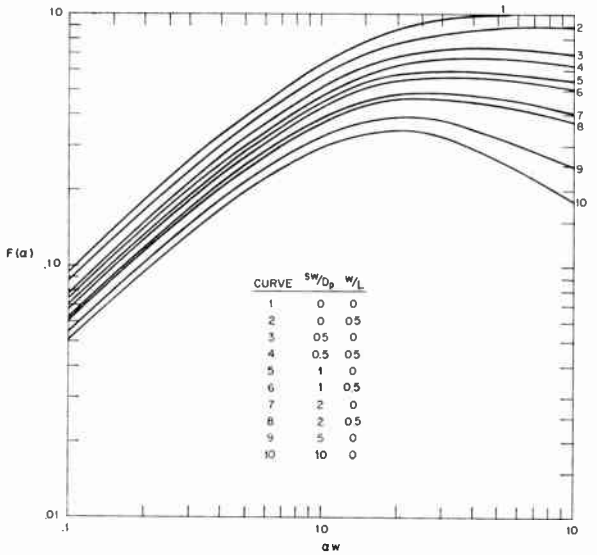


Fig. 2—The function $F(\alpha)$ for various sw/D_p and w/L values plotted from (14).

and $F(\alpha)$ is given by $F(\omega, \alpha)$ with ω set equal to zero. The quantity j_0 is the calculated diode saturation or "dark" current density. The function $F(\alpha)$ determines the increase in diode-junction current which is produced by steady-state illumination. This function may be described completely in terms of the dimensionless parameters αw, sw/D_p, and w/L. The parameter αw may be interpreted as the ratio of the base width to the mean depth of penetration $1/\alpha$ of the incident radiation. The parameter sw/D_p and w/L are associated with carrier loss by surface and volume recombination, respectively. In Fig. 2, $F(\alpha)$ is plotted as a function of αw for different values of sw/D_p and w/L. The uppermost curve in the figure represents the total number of minority carriers generated in the base as given by $[1-\exp(-\alpha w)]$. In the other curves the loss of these carriers by surface and volume recombination has been included. For values of αw larger than ten which are not shown in the figure, $F(\alpha)$ may be conveniently obtained directly from (14) by setting $\exp(-\alpha w)$ equal to zero. Good agreement between observed germanium photodiode steady-state behavior and that calculated from considerations employed here has been reported.[9]

Figs. 3 and 4 show the amplitude and phase dependence of $F(\omega, \alpha)/F(\alpha)$ as a function of frequency for different values of αw. The abscissa in these figures is the ratio of signal angular modulating frequency ω to the

[9] H. U. Harten and W. Schultz, "Influence of diffusion length and surface recombination on the barrier-layer photoelectric effect in germanium," *Z. Phys.*, vol. 141, pp. 319–334; July, 1955.

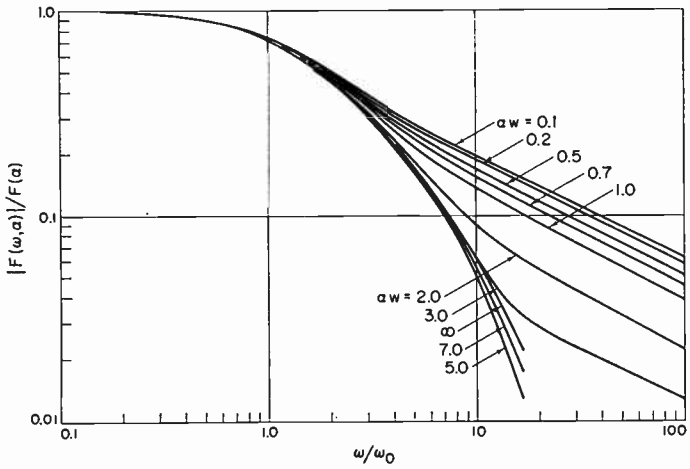


Fig. 3—The function $|F(\omega, \alpha)|/F(\alpha)$ for various αw values plotted from (14). The abscissa is the ratio of signal modulating angular frequency, ω, to the cutoff frequency, $\omega_0 = 2.4D_p/w^2$, of a transistor of base width identical to that of the photodiode. The curve for $\alpha w = 6$ (not shown) lies midway between those for $\alpha w = 5$ and $\alpha w = 7$; that for $\alpha w = 4$ lies very slightly below the curve for $\alpha w = 7$.

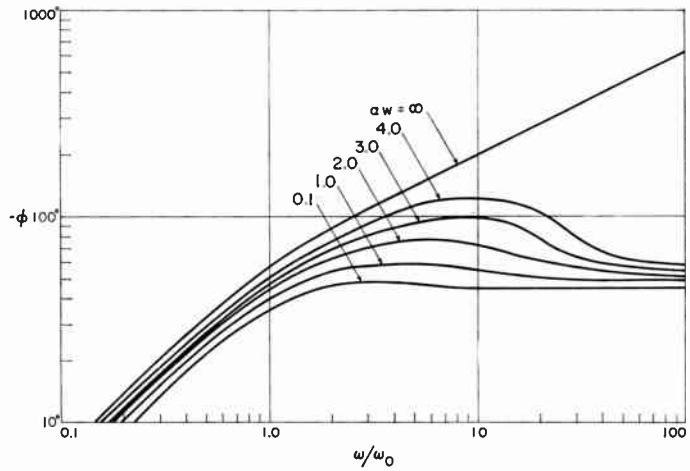


Fig. 4—Plot of the phase angle ϕ vs ω/ω_0 for various αw values. The quantity ϕ is defined as $\phi = \tan^{-1}[\text{Im } F(\omega, \alpha)/\text{Re } F(\omega, \alpha)]$.

cutoff frequency $\omega_0 = 2.4D_p/w^2$ for a transistor of base width identical to that of the photodiode. In the computation of these curves we have assumed that $\omega\tau \gg 1$ and $\alpha L \gg 1$. These assumptions are equivalent to requiring that the bulk-diffusion length L be much larger than both the base width and the mean penetration depth of the incident radiation. It will be convenient to designate as the photodiode-intrinsic cutoff frequency ω_{ico} that frequency at which the amplitude of $F(\omega, \alpha)/F(\alpha)$ is 0.707. For the case of $sw/D_p \ll 1$ plotted in Figs. 3 and 4, ω_{ico} depends only very slightly on the value of α. Even when the incident radiation is such that the mean generation depth $1/\alpha$ would be ten times the base width, ω_{ico} remains approximately equal to the transistor-equivalent cutoff frequency ω_0.

The frequency behavior for $\omega > \omega_{ico}$, however, is affected by the product αw especially when the product becomes of the order of unity or smaller. Thus while ω_{ico} is relatively unaffected, as αw is decreased the device becomes more usable at signal frequencies larger than ω_{ico} because $F(\omega, \alpha)$ decreases less rapidly with frequency in this range. At these frequencies, because of transit time dispersion the current contribution from carriers generated in the bulk just under the surface is reduced and the contribution from those carriers generated in closer proximity to the alloy junction becomes important. The current density j_{ac} can be determined by adding the contributions from the carriers generated at distances from the junction ranging from w to essentially zero, taking into account the phase and the transit time dispersion of each contribution. This is equivalent to what has been done in the derivation of (14).

Families of curves similar to Fig. 3 have been plotted for various sw/D_p values. These curves resemble those of Fig. 3 with one significant difference: for the finite values for αw shown on the figures, the intrinsic cutoff frequency ω_{ico} is increased, being approximately $1.6\omega_0$ for $sw/D = 1$, and approximately $2.5\omega_0$ for $sw/D = 5$ as contrasted with the approximate value of ω_0 for $sw/D_p = 0$. The reason for this difference is not difficult to envision. For large sw/D_p values, carriers generated close to the surface readily recombine at the surface, thus the base region just under the surface may be relatively ineffective in supplying photogenerated minority carriers for diffusion across the base width, w, to the junction contact. Consequently the maximum diffusion distance for carriers contributing to the signal current becomes somewhat less than the base width w with an increase in cutoff frequency over that for the same diode with $sw/D_p \ll 1$. However, increasing the value of sw/D_p increases the loss of generated carriers due to surface recombination and decreases $F(\alpha)$ as can be seen in Fig. 2. Therefore while the frequency cutoff may be increased, the absolute response to a given input signal at any frequency may be reduced.

Experiments were performed with a wide-base photodiode to verify the frequency behavior shown in Fig. 3 of $F(\omega, \alpha)/F(\alpha)$ for various values of the parameter αw. The diode alloy junction was 1.7 mm in diameter and a light tube was used to mask off from incident radiation all of the germanium surface opposite the junction except a central area 1 mm in diameter. A tungsten filament at a color temperature of approximately 2800°K was the radiant source for a Perkin-Elmer Model 88 monochromator fitted with a sodium chloride prism. The measured spectral half width at the exit slit was approximately 0.025 micron. The beam from the monochromator was sinusoidally modulated from 10 to 100 kc by a rotating disc and was brought to focus on the diode. An oscilloscope was used to measure the diode junction current with appropriate precautions taken so that the circuit response was "flat" past 100 kc. With radiation of 1.0 micron, which for this wide-base diode typifies essentially nonpenetrating radiation ($\alpha w \sim 500$), measurements yielded an ω_{ico} value of 38 kc. The output current, was determined as a function of the modulation frequency ω for various spectral wavelengths from 1.0 to 1.82 microns; this latter value represents for this

diode an αw value of approximately 0.3. The frequency behavior of $|F(\omega, \alpha)|/F(\alpha)$ shown in Fig. 3 for $\alpha w \leq 3$ and $\omega > \omega_0$ was clearly observed. However, while the results were indicative, one could not state with certainty that the predicted drop of $|F(\omega, \alpha)|/F(\alpha)$ with $3 \leq \alpha w < \infty$ and $\omega > \omega_0$ to values below the transistor-like frequency dependence curve ($\alpha w = \infty$) had been observed.

Photovoltaic Operation

In this treatment V is separated into dc and small-signal ac components

$$V = V_{dc} + v_{ac}. \tag{17}$$

Expanding (12) for $v_{ac} \ll kT/q$, (7) becomes

$$p(w) = p_{dc}(w) + p_{ac}(w)$$

$$= p_n e^{qV_{dc}/kT} + \frac{qv_{ac}}{kT} p_n e^{qV_{dc}/kT}. \tag{7a}$$

The boundary conditions at $x = w$ are determined by equating the current flowing through the detector load with the diffusion current crossing the junction:

$$i_{dc} = -qD_p \left.\frac{dp_{dc}}{dx}\right|_w = \frac{V_{dc}G_L}{A} \tag{18}$$

$$j_{ac} = -qD_p \left.\frac{dp_{ac}}{dx}\right|_w = \frac{v_{ac}Y_L}{A} \tag{19}$$

where the diode-load admittance is $Y_L = G_L + iB_L$ and A is the junction area taken to be equal to the window area. The small-signal solution for v_{ac} is

$$v_{ac} = \frac{qMI_0Ae^{i\omega t}F(\omega, \alpha)}{Y_1 + Y_L} \tag{20}$$

where

$$Y_1 = \frac{q^2Ap_nSe^{qV_{dc}/kT}}{kT} \left(\frac{\cosh\frac{w}{\xi} + \frac{D_p}{s\xi}\sinh\frac{w}{\xi}}{\cosh\frac{w}{\xi} + \frac{s\xi}{D_p}\sinh\frac{w}{\xi}} \right) \tag{20a}$$

is independent of the photogeneration process. The quantity V_{dc} is determined from the following transcendental equation:[10]

$$G_LV_{dc} + j_0A(e^{qV_{dc}/kT} - 1) = qI_0AF(\alpha). \tag{21}$$

The quantities $F(\omega, \alpha)$ and $F(\alpha)$ are identical to those defined in the previous section. For values of

$$\omega \ll \omega_0 = \frac{2.4D_p}{w^2}$$

the admittance Y_1 may be represented by the parallel combination of a resistor r_d and a capacitor C_d whose values for $sw/D_p \ll 1$ are

$$r_d = \frac{kT}{qj_0A}\exp\left(-\frac{qV_{dc}}{kT}\right) \tag{22}$$

and

$$C_d = \frac{1}{r_d}\cdot\frac{w}{s}. \tag{23}$$

This RC representation is valid for ω-values up to ω_d the angular-cutoff frequency of this combination:

$$\omega_d = \frac{1}{\tau_d} = \frac{s}{w}. \tag{24}$$

Junction Transition-Region Capacitance and Extrinsic Base Resistance

The transition-region capacitance for an abrupt, alloy junction is

$$\frac{C_t}{A} = \left[\frac{\epsilon}{2\mu_n\rho(\overline{V} + V_0)}\right]^{1/2} \tag{25}$$

where ϵ is the dielectric constant, ρ the resistivity, μ_n the majority carrier mobility, and $\overline{V}$ the junction voltage considered positive in reverse bias, *i.e.*, $\overline{V} \equiv -V$. The so-called internal contact potential V_0 is a function of material and junction properties.[11] To a good approximation, the applied junction voltage $\overline{V}$ in (25) can be considered as zero for photovoltaic operation. The corresponding transition-region capacitance, which will be denoted by $C_t(0)$, is considerably larger than $C_t(V)$, the transition-region capacitance in reverse-biased operation.

The extrinsic base resistance of the diode structure of Fig. 1 may readily be calculated as

$$r_b' = \frac{\rho}{8\pi w}\left[1 + \frac{4w}{t}\ln\frac{r_2}{r_1}\right]. \tag{26}$$

Equivalent Circuits for Reverse-Biased and Photovoltaic Operation

The ac signal equivalent circuits for reverse-biased and photovoltaic operation are obtained with the aid of (14), (20), (25), and (26) and are shown in Figs. 5 and 6, respectively. In these figures we have added a junction-shunting conductance G_j always found in practice. The quantity i_s represents an rms current generator of magnitude $j_{ac}A/\sqrt{2}$. At the lower frequencies where the reactances may be neglected, the load-signal amplitude for reverse-biased operation is limited only by G_j, while that for photovoltaic operation is limited by the real part of Y_1. For typical values of G_j of 10^{-6} to 10^{-7} mhos

[10] This equation appears in the expressions for photovoltaic power converters. See for example, W. G. Pfann and W. Van Roosbroeck, "Radioactive and photoelectric p-n junction power sources," *J. Appl. Phys.*, vol. 25, pp. 1422–1434; November, 1954; also M. B. Prince, "Silicon solar energy converters," *J. Appl. Phys.*, vol. 26, pp. 534–540; May, 1955.

[11] D. R. Muss, "Capacitance measurements on alloyed indium-germanium junction diodes," *J. Appl. Phys.*, vol. 26, pp. 1514–1517; December, 1955; also, R. F. Schwarz and J. F. Walsh, "The properties of metal to semiconductor contacts," Proc. IRE, vol. 41, pp. 1715–1720; December, 1953.

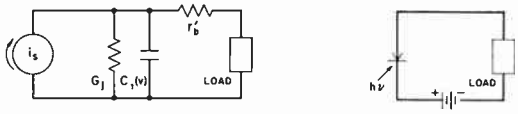


Fig. 5—Equivalent circuit for a photodiode operated as a reverse-biased detector. G_j is the junction-shunting conductance, C_t is the transition-layer capacitance, r_b' is the ohmic-base resistance and the current generator i_s has the rms value $j_{sc}A/\sqrt{2}$ where i_{sc} is given by (14).

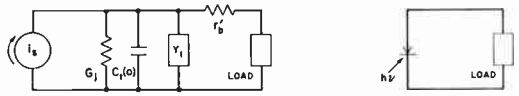


Fig. 6—Equivalent circuit for a photodiode operated as a photovoltaic detector. This circuit is identical to that of Fig. 5 except for the additional shunt admittance Y_1 given by (20a). The transition-region capacitance $C_t(0)$ evaluated at zero applied voltage may be considerably larger than $C_t(V)$ the transition region capacitance in Fig. 5.

and saturation currents of the order of one microampere, the load voltage for a given light signal may be up to 40 to 400 times larger for reverse-biased than photovoltaic operation. Also, the equivalent capacitance shunting the current generator i_s is always larger (and may be considerably larger) in photovoltaic operation. Thus for a given resistive load, the cutoff frequency (ω_{cco}) determined by the passive elements in the equivalent circuit will always be higher in reverse-biased operation. Since the low-frequency amplitude may be greater and the capacitive shunting reduced by operating the diode reverse biased, this mode of operation may be more satisfactory for detection applications than operation of the diode as a photovoltaic device.

Noise Equivalent Power

The minimum detectable signal for the photodiode is established by the noise generated by the detector and detector preamplifier. The three types of noise which must be considered are thermal, shot, and the so-called 1/f noise. With the diode reverse biased it is found in practice that at the lower frequencies, 1/f noise establishes the noise level and consequently the value of minimum detectable power. A good small-area germanium junction diode, however, should show negligible 1/f noise above the kilocycle range.[12,13] In the analysis which follows it is assumed that the signal frequencies of interest are above the 1/f region. Only the reverse-biased case will be discussed, since it can be shown that for this case photovoltaic operation increases the value of minimum detectable power.

If it is assumed that the extrinsic base resistance, r_b', of the photodiode may be neglected, then the noise and signal equivalent circuit of the photodiode and load admittance is as shown in Fig. 7. We have assumed that the load admittance is that of a conductance G_L in parallel with a capacitance C_L. G_{eq} is the sum of G_j the junction-shunt conductance and G_L; C_{eq} is the sum of $C_t(V)$ and C_L. The noise sources are an equivalent noise-current source of magnitude

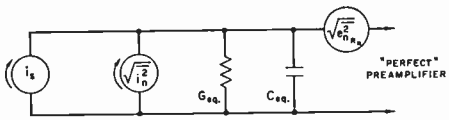


Fig. 7—Signal and noise equivalent circuit for a photodiode operated as a reverse-biased detector. The current generator is identical to that in Fig. 5. G_{eq} and C_{eq} are the resultant conductance and capacitance of the equivalent circuit including the load, and the noise current and voltage generators are described in the text.

$$\overline{i_n^2} = (2qi + 4kTG_{eq})df$$

and an equivalent thermal noise-voltage source of magnitude

$$\overline{e_{nR_n}^2} = 4kTR_n dj$$

where i is the junction current, i_s is the rms signal current, and the actual preamplifier is characterized by an equivalent noise-input resistor R_n in combination with a "perfect" preamplifier.

The signal-to-noise ratio is given by the ratio of the square of the signal voltage, e_s^2, to the mean-square noise voltage, $\overline{e_n^2}$, both referred to the input of the "perfect" amplifier.

$$e_s^2 = \frac{i_s^2}{G_{eq}^2 + \omega^2 C_{eq}^2} \tag{27}$$

and

$$\overline{e_n^2} = \left(\frac{2qi + 4kTG_{eq}}{G_{eq}^2 + \omega^2 C_{eq}^2} + 4kTR_n\right)df. \tag{28}$$

Thus the signal-to-noise ratio is

$$\frac{e_s^2}{\overline{e_n^2}} = \frac{i_s^2}{(2qi + 4kTG_{eq} + 4kTR_nG_{eq}^2 + \omega^2 C_{eq}^2)df}. \tag{29}$$

This result is similar to that for a case considered by van der Ziel.[14] The noise-equivalent power is that value of incident power for which the signal-to-noise ratio is unity. The corresponding signal current is given by

$$i_{s_{nep}}^2 = [2qi_0 + 4kTG_{eq}(1 + G_{eq}R_n) + \omega^2 4kTR_nC_{eq}^2]df. \tag{30}$$

In (30) and in the following discussion $i_{s_{nep}}$ is justifiably assumed to be much smaller than i_0, the diode-saturation current, so that i is replaced by i_0.

$i_{s_{nep}}$ can be related to the total incident power, $AW \equiv \underline{W}$, and diode parameters by (3), (4), and (14a)

$$i_s = \frac{qM(1 - R)\underline{W}\lambda}{(1 + M)hc\sqrt{2}} F(\omega, \alpha). \tag{31}$$

[12] A. van der Ziel, "Theory of shot noise in junction diodes and junction transistors," Proc. IRE, vol. 43, pp. 1639–1646; November, 1955.

[13] A small percentage of junction diodes fabricated in the laboratory have shown erratic noise characteristics exhibiting current spikes and noise bursts, but since these are the exceptions, they have not been discussed here.

[14] A. van der Ziel, "Noise," Prentice-Hall, New York, N. Y., p. 97; 1954.

Combining (30) and (31), $\underline{W}_{nep}$ may be written as

$$\underline{W}_{nep} = \underline{W}_0(1 + (\omega/\omega_n)^2)^{1/2} \tag{32}$$

where

$$\underline{W}_0 = \frac{(1 + M)hc\{[4qi_0 + 8kTG_{eq}(1 + G_{eq}R_n)]df\}^{1/2}}{qM(1 - R)\lambda\,|\,F(\omega, \alpha)\,|} \tag{33}$$

and

$$\omega_n = \left[\frac{2qi_0 + 4kTG_{eq}(1 + G_{eq}R_n)}{4kTR_nC_{eq}^2}\right]^{1/2}. \tag{34}$$

In order to specify both the signal response and the signal-to-noise ratio of a photodiode three cutoff frequencies must be considered: ω_{ico}, the cutoff frequency associated with the photodiode intrinsic response $F(\omega, \alpha)$, ω_{cco} the cutoff frequency associated with the passive elements in the equivalent circuit, and ω_n which specifies the circuit signal-to-noise degradation with frequency as given by (32).

With saturation currents the order of a fraction of a microampere or greater and normal values of junction-shunt conductance G_j a value of load conductance can be chosen so that $G_{eq} \ll 2kT/qi_0$ and thus the thermal noise due to G_{eq} will be negligible in respect to shot noise. Furthermore, this value of G_{eq} will in general be such that $G_{eq}R_n \ll 1$ for typical R_n values. With these readily attained conditions $\underline{W}_0$ and ω_n become

$$\underline{W}_0 \approx \frac{2(1 + M)hc(qi_0df)^{1/2}}{qM(1 - R)\lambda\,|\,F(\omega, \alpha)\,|} \tag{33a}$$

$$\omega_n \approx \left(\frac{20i_0}{R_nC_{eq}^2}\right)^{1/2}. \tag{34a}$$

If, as suggested above, G_L is reduced to maximize the signal-to-noise ratio, the output-signal cutoff frequency determined by the equivalent circuit (ω_{cco}) will be reduced in many cases to a frequency much lower than the desired upper limit of the pass band. On the other hand, if ω_{cco} limits the frequency response, an obvious method of broadbanding the detector is to increase G_L. It is seen from (33) and (34) that increasing G_L degrades the signal-to-noise ratio at all frequencies. A possibly more satisfactory method of frequency compensation is to perform the compensation following the preamplifier.[15] Compensation can be done successfully in this case because the signal-to-noise ratio may be constant to a much higher frequency than ω_{cco} the frequency at which the signal response may be limited. However, if the cutoff frequency ω_{ico} of the photodiode intrinsic response $F(\omega, \alpha)$ limits the signal-frequency response, broadbanding may be of limited use because ω_{ico} as well as ω_n is a limiting frequency for the signal-to-noise ratio [see (32)].

[15] R. G. Breckenridge, B. R. Russell, and E. E. Hahn, "Photoconductivity Conference," John Wiley and Sons, Inc., New York, N. Y., p. 74; 1956.

Advantages of a Narrow Base

For incident radiation of wavelength shorter than 1.52 microns the results of the previous analysis indicate several advantages which accrue when the photodiode base width is made narrow. Since the absorption coefficient α of germanium at 300°K for these wavelengths is larger[5] than 5×10^{-3} cm^{-1}, almost all the radiation entering the semiconductor can be absorbed in a base width of a few microns. With base widths of this order, photodiodes with intrinsic cutoff frequencies $\omega_{ico}/2\pi$ above 20 mc are possible. [See (14) and Figs. 3 and 4.] Furthermore, in this range of high absorption, these photodiodes may be more efficient than photodiodes with a wider base since as the base width is reduced the surface and bulk losses are reduced (see Fig. 2). If we relax our frequency response requirement from tens of megacycles to a few megacycles, then for wavelengths less than 1.52 microns (with transistor-quality single-crystal germanium), it is possible to simultaneously satisfy the following inequalities $1 \ll \alpha w$, $w \ll L$, $w \ll 2\tau s$, and $sw/D_p \ll 1$ so that the expression for $F(\alpha)$

$$F(\alpha) \approx 1 - sw/D_p \tag{35}$$

representing the fractional loss sw/D_p of the photogenerated carriers by surface recombination. The diode saturation current density as calculated from the one-dimensional model then becomes simply

$$j_0 \approx qp_ns.$$

In addition to the advantages listed above which apply to both reverse-biased and photovoltaic operation, as the base width is reduced the reactive component of Y_1 which enters in photovoltaic operation and is usually associated with the diffusion capacitance, C_d, is also reduced [see (23)]. Thus reducing the base width increases the cutoff frequency ω_{cco} associated with the passive elements in the equivalent circuit. However, as was indicated previously, the total capacitance shunting i_s can never be reduced below that for reverse-biased operation and ω_{cco} is always smaller for photovoltaic than for reverse-biased operation.[16]

For a photodiode with a narrow base it is possible for ω_{ico} to be larger than either ω_{cco} or ω_n so that, as was mentioned in the section on noise equivalent power, frequency compensation may be used to extend the utilization range of the photodiode to ω_{ico} with negligible degradation of the signal-to-noise ratio at frequencies less than ω_n. However, it is possible with a sufficiently narrow base that the assumption made in that section that r_b' is negligible might not be applicable. If reasonable care is taken in the photodiode

[16] From these considerations it would appear that photovoltaic operation is always less desirable than reverse-biased operation. However, if it is desired to retain the steady-state signal component, photovoltaic operation may be chosen since variation in the temperature-sensitive quantity p_n (and possibly G_j) is reflected only as a change in signal amplitude and not as a spurious detector load voltage.

TABLE I

THE EFFECTS OF CHANGES IN BASE RESISTIVITY AND DEVICE GEOMETRY ON PHOTODIODE PARAMETERS. A DASH SIGNIFIES LITTLE OR NO EFFECT, AN ARROW INDICATES THE DIRECTION OF THE EFFECT

	$F(\alpha)$ (14)	j_0 (16)	$C_t(V)$ (25)	$C_t(0)$[10]	C_d (23)	R_b' (26)	r_d (22)	$\underline{W}_0$ (33)	ω_{ico} (14)	ω_{cco} (27)
Reduction of ρ	—	↓	↑	↓	↓	↓	↑	↓	—	↓
Reduction of w	↑	—	—	—	↓	↑	—	—	↑	—
Reduction of junction area with r_2/r_1 constant	—	↓	↓	↓	↓	—	↑	↓	—	↑
Reduction of s	↑	↓	—	—	↑	—	↑	↓	—	—

design, this assumption concerning r_b' may be applicable for photodiodes with values of $\omega_{ico}/2\pi$ in the megacycle range, as shown in the next section.

The advantages listed above are predicated on the wavelength of the incident light being shorter than 1.52 microns. If, however, the wavelength of the incident radiation is longer than 1.52 microns and the steady-state sensitivity is of paramount importance, it may be advantageous to use a wider base. In this case if the values of α, s, and L are known, one can choose the optimum base width for maximum steady-state sensitivity with the aid of Fig. 2.

NARROW-BASE PHOTODIODE DESIGN

Table I shows the effects of changes in base resistivity and device geometry on the electrical parameters of the photodiode. The pertinent equation number in the text is shown beside the parameter headings. In obtaining the results tabulated in Table I, it was assumed that the total power incident on the window area is constant and that all the radiation that enters is absorbed in the base region ($\alpha w \gg 1$).

The relative importance of the various device parameters and thus the determination of the actual design of a narrow-base photodiode may strongly depend on the characteristics of the optical system and preamplifier used with the photodiode. Satisfactory devices for particular requirements have been made with germanium resistivities of from 0.5 to 4.0 ohm-cm and with junction diameters from 0.022 to 0.040 inch. A base width of from ten to twenty microns has been a satisfactory compromise between response time and extrinsic base resistance. The narrow-base alloy photodiodes such as shown in Fig. 1 have been fabricated using the bath-etching technique described in the literature.[17]

Measured saturation currents have ranged from 0.45 to 3.5 μa while the junction ohmic conductance has been consistently smaller than 1×10^{-7} mhos. Spectral measurements on these diodes have shown a constant quantum yield within an experimental error of ten per cent from 0.55 micron, the lower measurement limit, to 1.52 microns dropping sharply at this latter value which corresponds to the threshold energy for the more probable direct transition.[5] In reverse-biased operation decay time of the photoresponse following a light pulse[18] from a xenon flash tube is from 60 to 100 mμsec for a sufficiently small value of load resistor so that response is not limited by the cutoff frequency (ω_{cco}) of the diode preamplifier circuit. Measurements with a photomultiplier indicated that the decay time of the light pulse itself is in the neighborhood of 50 mμsec. In fast-response service a mask or focusing means must be used to restrict the incident radiation to the germanium area directly opposite the alloy junction. If this is not done, the pulse-decay tail will be lengthened by the late arrival at the junction of carriers photogenerated a distance from the alloy junction greater than w, the base width.

As an illustrative example of what may be achieved in practice with narrow-base photodiodes, the frequency behavior of the signal response and the noise equivalent power will be calculated for a representative photodiode. Referring to Fig. 1, this diode has a junction radius r_1 of 0.028 cm, a window frame radius r_2 of 0.125 cm, a base width w of 1.25×10^{-3} cm and a wafer thickness t of 0.015 cm. The base resistivity, ρ, is 2.0 ohm-cm. Saturation current, transition-region capacitance, and junction shunt conductance have been measured on diodes of this type and are respectively about $i_0=1.5$ μa, $C_t(6)=10$ $\mu\mu$f and $G_j<10^{-7}$ mhos. The intrinsic base resistance has been calculated to be 95 ohms. With a preamplifier of 400 ohms equivalent input-noise resistance and a detector load of 100 kohms in parallel with 10 $\mu\mu$f, the values of the intrinsic and circuit cutoff frequencies and the noise-degradation frequency are $\omega_{ico}=6.7\times10^7$ sec^{-1}, $\omega_{cco}=5\times10^5$ sec^{-1}, and $\omega_n=1.5\times10^7$ sec^{-1}, respectively. Frequency compensation *following* the preamplifier can be used to extend the signal-frequency response to ω_n with a negligible increase in the low-frequency noise equivalent power $\underline{W}_0$ below this frequency. This value of $\underline{W}_0$ is 3×10^{-12} watts/cycle$^{1/2}$ for $\lambda=1.5$ microns, M equal to unity, and with the reflection coefficient[19] R taken as

[17] R. H. Rediker and D. E. Sawyer, "Very narrow base diode," PROC. IRE, vol. 45, pp. 944–953; July, 1957.

[18] For a description of the light pulser see: D. T. Stevenson and R. J. Keyes, "Measurement of carrier lifetimes in germanium and silicon," *J. Appl. Phys.*, vol. 26, pp. 190–195; February, 1955.

[19] W. G. Spitzer and H. Y. Fan, "Determination of optical constants and carrier effective mass of semiconductors," *Phys. Rev.*, vol. 106, pp. 882–890; June, 1957.

0.36. If we choose to broadband the diode and the preamplifier input circuit and make $\omega_{cco}=1.5\times10^{7}$ sec^{-1} by reducing the load resistance to 3.3 kohms, then the previous value of $\underline{W}_0$ is increased by a factor of three and becomes $\underline{W}_0=9\times10^{-12}$ watts/cycle$^{1/2}$.

Photovoltaic operation of these narrow-base photodiodes in practice is found to degrade the high-frequency performance of the device even more than one might expect since experimentally determined values of the capacitance sum $C_d+C_t(0)$ have been found to be several times larger than that calculated from (23) and (25). The excess capacitance is believed to be due to the fact that the diffusion capacitance for a structure such as Fig. 1 cannot be accurately calculated unless one takes into account the carriers injected laterally from the junction into areas where the base width is more accurately specified by the wafer thickness t than by the smaller value w. Since the diffusion capacitance per unit area is an increasing function of the base width [see (23)], this lateral injection phenomenon contributes significantly to the diffusion capacitance.

ACKNOWLEDGMENT

The authors wish to thank C. R. Grant, L. Krohn, W. H. Laswell, and Mrs. M. L. Barney for their help in various aspects of the fabrication of the photodiodes. They are indebted to S. Zwerdling for optical measurements. They are grateful to Mrs. B. J. Houghton for computations and to C. W. Tillinghast for computations performed using the IBM-704. Profitable discussions were had with R. H. Kingston on noise. The authors are indebted to R. J. Keyes for discussions throughout the period of this research and for optical measurements.

Advances in the Understanding of the *P-N* Junction Triode*

R. L. PRITCHARD†, SENIOR MEMBER, IRE

Summary—**During the past ten years the junction triode has been studied extensively in attempts to improve the understanding of the device. The resulting increase in understanding has made possible improvements in both device design and in transistor-circuit design. Highlights of these studies including references to approximately 100 papers, are reviewed in this paper from the point of view of relating electrical characteristics to the physical construction of the device.**

First the ideal-diode model resulting from Shockley's 1949 analysis of the junction triode is reviewed. Then the differences between experimentally observed electrical characteristics and the corresponding characteristics of the ideal model, together with their explanations presented in the literature, are described. The dc characteristics are discussed first, followed by a detailed description of the increased understanding of the small-signal parameters. Other topics reviewed briefly include the effect of non-one-dimensional current flow, and switching characteristics of the triode.

INTRODUCTION

DURING the past ten years, the *p-n* junction triode has evolved from the "paper transistor" of Shockley [1] to a widely used electron device, of which approximately 27 million units were produced commercially in 1957. Part of the reason for this tremendous progress is the increased understanding of the device, which in turn has led to increased designability. Of course, important technological improvements, such as the recent introduction of solid-state diffusion, have made it practical to obtain this designability. Junction transistors today can be designed and built to perform specific functions, and their behavior in circuits can be predicted. This paper surveys the developments that have taken place in the understanding of the junction triode during the last ten years.

Transistor Types

Before discussing the operation of the junction triode, a brief description of the device might be in order. The first junction transistors [2], [3] were of the grown-junction type, in which the transistor comprises a rectangular bar cut from a germanium crystal that has been grown from a melt with suitable added impurities. Shortly thereafter, the alloy technique was developed [4], in which small dots of indium were fused, or alloyed, into a germanium wafer of suitable conductivity. These two basic types of transistor construction are illustrated in Fig. 1. Attempts to reduce the dimensions of alloy transistors later led to the introduction of the electrochemical and plating techniques and the development of the surface-barrier transistor [5].

The first grown-junction transistors were *n-p-n*, whereas the alloy units were *p-n-p*. However, *p-n-p* grown-junction transistors and *n-p-n* alloy units have since been developed with germanium, along with several types of *silicon* transistors.

* Original manuscript received by the IRE, March 25, 1958.
† Texas Instruments, Inc., Dallas, Texas.

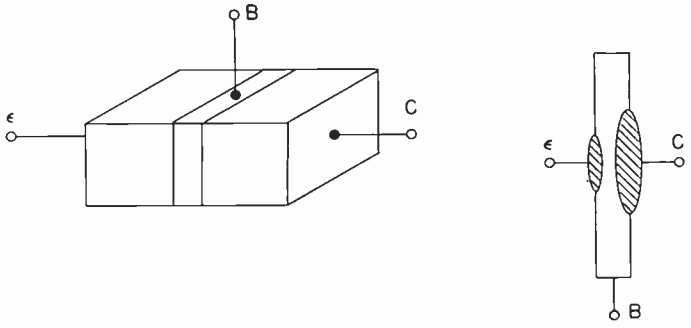


Fig. 1—Classical types of junction-transistor construction, illustrating (a) rectangular bar construction of grown junction and (b) opposed dot construction of fused, or alloy, junction, or surface-barrier transistors.

During the past several years the introduction of solid-state diffusion, which is reviewed by Smits [6], has led to the development of a number of other types of transistors, *e.g.*, the diffused-base transistor [7]. In many cases better performance can be obtained with the diffused-type units than is possible with conventional classical structures described above.

Basic Transistor Operation

As is well known,[1] a junction triode comprises three regions, emitter, base, and collector, separated by two transition regions or *p-n* junctions as shown in Fig. 2. A *p-n-p* transistor is discussed, but an *n-p-n* operates in quite an analogous fashion. For normal transistor operation with the emitter junction forward biased and the collector junction reversed biased, holes are injected from the emitter into the base, which they traverse by diffusion,[2] and ultimately are collected by the collector. Hence, the collector current I_C is proportional to emitter current I_E:

$$-I_C = \alpha I_E - I_{CO}, \tag{1}$$

where the proportionality constant α is the short-circuit forward current transfer ratio, and I_{CO} is the collector reverse current. These two quantities are two of the most important parameters of a junction transistor.

On the other hand, with a reversed biased collector, the emitter-base junction behaves like a simple *p-n* diode, having a voltage-current relation

$$I_E = I_{EO}[\epsilon^{qV_{EB}/kT} - 1], \tag{2}$$

where I_{EO} is the emitter reverse current, *viz.*, the emitter current that would flow if the emitter-base diode were reversed biased.

The junction triode also may be operated with the roles of the emitter and collector interchanged (inverse operation), in which case additional relations similar to (1) and (2) also exist. In some transistors (bilateral devices), normal and inverse operation lead to essentially the same performance, but in most units, better performance is obtained with normal operation.

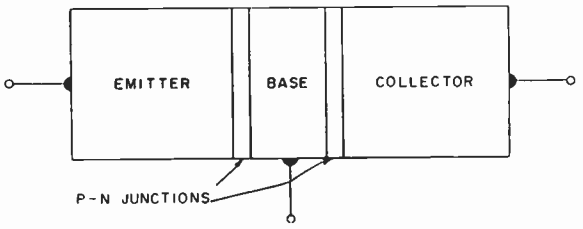


Fig. 2—Schematic diagram of junction transistor, illustrating emitter, base, and collector regions separated by two *p-n* junctions.

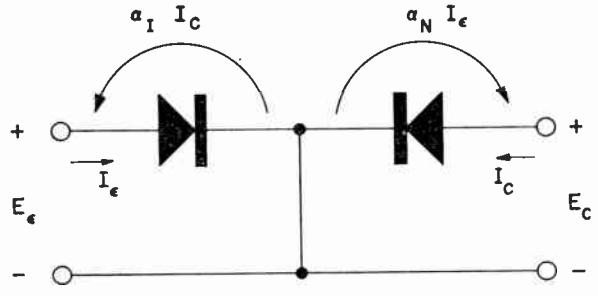


Fig. 3—Large-signal two-diode equivalent circuit for junction transistor.

Alternatively, the transistor can be operated with both junctions reversed biased, in which case the emitter and collector currents are quite small, of the order of the reverse currents or less, or with both junctions forward biased, in which case the voltage drops between the three terminals can be quite small. This type of operation makes the junction triode an attractive switch.

More generally, the performance of the ideal transistor can be described in terms of ideal *p-n* diodes and current generators as shown by the equivalent circuit of Fig. 3, which is based on Shockley's original 1948 analysis of the junction triode.[3] For many applications, this equivalent circuit is quite useful, *e.g.*, as a first approximation in switching circuits. However, it is incomplete in that it fails to account for many of the experimentally observed electric-circuit properties.

Historically, most of the efforts to understand the operation of a transistor were directed towards studying the small-signal parameters. However, at the same time considerable fundamental effort also was directed toward further understanding of the electrical characteristics of the *p-n* junction diode, and as this knowledge became available, it was applied directly to the junction triode. These advances are covered first in the following section on dc characteristics, followed by a review of the small-signal parameters which constitutes the bulk of this paper. Additional sections are devoted to brief discussions of switching characteristics, temperature characteristics, and noise in the junction triode.

[1] In addition to numerous texts on the subject, several excellent review papers exist, *e.g.*, [8].

[2] Or with the help of a drift field, as in some of the recent types of graded-base transistors.

[3] The author believes that the circuit in this form was first shown by Adler in an unpublished report [9].

DC Voltage-Current Characteristics

Forward Characteristics

In an ideal diode, cf. (2), forward current (and hence the emitter current in a triode) varies exponentially with forward voltage V as $\epsilon^{qV/kT}$. In practical transistors at low currents the forward voltage-current characteristic generally is found to be exponential, although in general, the exponent is $qV/\lambda kT$ where λ is a number[4] between 1 and 2. On the other hand, at high currents, the forward voltage-current characteristic tends to become essentially linear, due to the series resistances associated with the emitter and/or the base regions or contacts.

A number of theoretical studies of the *p-n* diode, some of which are reviewed by Henkels [10], have been made in an attempt to evaluate the factor λ.[5] These results may be summarized briefly as follows. The current described by (2) for a simple ideal diode corresponds to a diffusion current of low current *density*. However, under certain conditions, other types of currents which vary with voltage as $qV/2kT$ may flow across the *p-n* junction. Hence, in general the sum of two such currents comprising the total forward current, will have a voltage characteristic with a value of λ that varies from 1 to 2 over the current range of interest.

For example, in a normal germanium diode or transistor at low current densities, diffusion current dominates, and total current varies with voltage as $\epsilon^{qV/kT}$. But, with increasing current density in the base region, Hall [13] has shown that hole current injected into the base ultimately varies as $\epsilon^{qV/2kT}$. Hence, the exponent $\lambda=1$ at low currents, and $\lambda\rightarrow 2$ with increasing current.

Alternatively, in silicon junctions at low current densities, as shown by Sah, Noyce, and Shockley [14], and in germanium junctions at low temperatures, as shown by Bernard [15], forward current is dominated by current flow due to carrier recombination in the depletion layer. This current varies with voltage as $\epsilon^{qV/2kT}$, while the smaller, normal diffusion current varies as $\epsilon^{qV/kT}$. With increasing forward voltage, the diffusion current overcomes the recombination current, and hence the exponent λ varies from $\lambda=2$ at low currents to $\lambda=1$ at moderate currents. At still higher currents, $\lambda=2$ as in the germanium case. This depletion-layer recombination current also affects the reverse characteristics of the *p-n* junction and the forward current-transfer ratio α, as is described below.

Reverse Characteristics

For reverse collector-base voltages in the range between a few tenths of a volt and the onset of breakdown described below, the ideal diode model indicates a voltage-independent current, which is termed a saturation current (I_{co} for open-circuited emitter). In germanium transistors at normal and at elevated temperatures, the saturation current is largely the result of minority carriers from the collector and base crossing the collector-base junction to recombine in the base and collector respectively, although a portion of the saturation current also may be due to carriers recombining on the surface [16–18].

However, in most *p-n* junctions the reverse current is not completely voltage independent. This generally is attributed to surface leakage [19], [20]. For example, a logarithmic variation of current with voltage can be attributed to a "channel," or *N*-type surface layer on a *P*-type collector region [21], [22]. A more complete discussion of surface leakage is found in an excellent review of surface phenomena on germanium, Kingston [23].

On the other hand, very careful measurements of reverse currents of *p-n* junctions during the past several years have indicated that in some types of junctions an additional voltage dependence exists due to a bulk effect. For example, Pell has shown that in germanium junctions [24] at low temperatures, and in silicon junctions [25] at normal and at low temperatures, the reverse current may be dominated by a current flow due to recombination of minority carriers in the collector depletion layer. Several other writers [26], [14], [15] have also discussed this type of current, and its effect upon the forward junction characteristics has been noted above. The depletion-layer recombination current is proportional to the volume (or the width) of the depletion layer which, in turn, is a function of reverse voltage.[6] For example, in a grown-junction transistor with a graded collector junction, the depletion-layer width, and hence the depletion-layer component of the reverse current, will be proportional to the cube root of the reverse voltage.

With other than an open-circuited emitter (or collector), the reverse voltage-current characteristics of a transistor are somewhat modified by the effects of the current transfer ratios, both normal and inverse. For example, with an open-circuited base ($I_B=I_E+I_C=0$) and with a reverse voltage applied between collector and emitter, the reverse current is $I_{co}/(1-\alpha)$. [Cf. (1).] This current will exhibit a voltage dependence in addition to that of I_{co} because of the bias dependence of α to be described.

Breakdown Voltage

For any *p-n* junction, as the reverse voltage is increased a critical, or breakdown, voltage is approached for which the reverse current increases rapidly. Early studies [27] indicated that this breakdown was due to the Zener, or field-emission effect. For example, diodes designed for voltage-reference service which utilize this

[4] For example, in the transistor of Shockley, Sparks, and Teal [2], experimental data corresponding to $\lambda=1.25$ are presented.

[5] On the other hand, some of the diode studies have been concerned with structures that are not applicable for the emitter-base junction of a triode, *e.g.*, the *p-i-n* diode [11], [12].

[6] The depletion layer width is inversely proportional to the transition capacitance.

breakdown occasionally are still termed "Zener diodes." However, more recent studies by McKay [28] for silicon and by Miller [29] for germanium have shown that the voltage breakdown generally is due to an avalanche mechanism which occurs in the collector depletion layer.[7]

For sufficiently high electric fields in the collector depletion layer (of the order of 2×10^5 volt/cm) injected carriers create hole-electron pairs by impact ionization, similar to the Townsend effect in gas discharges. This causes an effective multiplication m of the collector current relative to the hole current entering the collector depletion layer from the base. Furthermore, at a critical voltage, defined as the body breakdown voltage V_B, the multiplication is self-sustaining, $m\rightarrow\infty$, and the collector current tends to become infinitely large. For alloy junctions on a wafer, the voltage V_B is simply related to the wafer resistivity.[8]

In practice the breakdown characteristic may be either sharp, as the theory for body breakdown would indicate, or "soft," in which case the current increases moderately for a fairly wide range of voltages so that breakdown is not clearly defined. Soft breakdown generally is attributed to surface conditions, although surface breakdown can be sharp and also due to an avalanche process [32]. On the other hand, Pell [33] has presented data suggesting that a bulk effect might be responsible for soft breakdown in some cases.

Even with body breakdown, thermal effects may modify the breakdown characteristics [34], [35]. For example, Matz [36] has shown that the onset of thermal runaway, particularly at elevated temperatures, may reduce the actual breakdown voltage below the avalanche value.

When both the emitter and collector terminals of the transistor are utilized, the breakdown voltage in some cases may be considerably different from that of a simple diode described above. For example, in some of the early transistors it was found that collector-emitter breakdown would occur at a considerably lower voltage than collector-base breakdown. This was attributed to the "punch-through" effect [37]. In a p-n junction the width of the depletion layer increases with increasing reverse voltage. This causes a decrease in the base width as the collector depletion layer extends partially into the base and partially into the collector with increasing collector voltage. For a sufficiently large collector voltage V_{PT} the collector-depletion layer may extend completely through the base region until it merges with the emitter-depletion layer.[9] Increasing voltage beyond V_{PT} cannot further increase the width of the depletion layer, and the emitter voltage then increases with increasing collector voltage. Consequently, the collector-emitter voltage is effectively limited to V_{PT}, and an apparent breakdown occurs.

However, as information on avalanche multiplication became available, it was noted that a lower collector-emitter breakdown voltage (with $I_B=0$) also could be caused by an infinitely large increase in the collector current $I_{CO}/(1-\alpha)$ when the current transfer ratio α became equal to 1 because of the effect of the avalanche multiplication factor m upon α [37]. For example [see (4)] for a p-n-p germanium transistor with a body breakdown voltage $V_B=100$ volts and a value of $\alpha=0.98$ at low collector voltages, with a collector voltage of 25 volts, avalanche multiplication is sufficiently large to cause $m\doteq1.02$, and hence $\alpha\doteq1$. This 25 volts then would be the collector-emitter breakdown. In general, the collector-emitter breakdown voltage will occur at *either* the voltage determined by avalanche multiplication, *or* by punch-through, depending upon which is the lower.

On the other hand, if the base is connected to the emitter through a dc resistance R_B, the collector-emitter breakdown voltage may be one of several different values. For low collector voltages, as R_B is decreased from infinity (open base), the collector current decreases from $I_{CO}/(1-\alpha)$ to a value somewhat larger than I_{CO} for $R_B=0$. Hence, I_{CO} becomes less dependent upon α, and avalanche multiplication becomes less important with decreasing emitter-base resistance, so that the collector-emitter breakdown due to multiplication increases toward the body breakdown voltage V_B. However, if the punch-through voltage $V_{PT}<V_B$, then the collector-emitter breakdown voltage will occur at V_{PT}.

Small-Signal Parameters

The four small-signal parameters to be discussed, all for the common-base configuration, are the short-circuit input resistance h_{ib}, the open-circuit output conductance h_{ob}, the open-circuit reverse voltage transfer ratio h_{rb}, and the short-circuit forward current transfer ratio $h_{fb}=-\alpha$. First, these parameters are discussed for the case of the Shockley ideal-diode model, and then the modifications that have resulted from increased understanding over the past decade are reviewed.

Ideal-Diode Model

For normal transistor operation, for example, an amplifier, with the emitter junction forward biased and the collector junction reversed biased, the ideal equivalent circuit of Fig. 3 reduces to that shown in Fig. 4. This circuit consists principally of the emitter diode resistance r_e' (the incremental resistance of the emitter-base

[7] Except possibly in the case of very narrow (low-breakdown voltage) junctions, when Zener breakdown actually may be observed [30], [31].

[8] For example, for a wafer of n-type germanium of 1 ohm-cm, the body breakdown voltage $V_B=100$ volts, and V_B varies approximately as the 0.7 power of resistivity.

[9] For example, in the p-n-p germanium allow transistor with a 1 ohm-cm base resistivity and a 0.3-mil base width, punch-through would occur at about 70 volts.

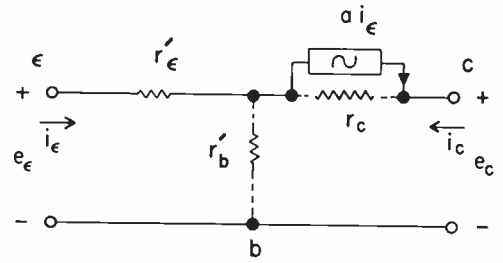


Fig. 4—Small-signal equivalent-Tee circuit for junction transistor.

diode) and a forward-current transfer generator ai_e.[10] In addition, as shown by dotted lines in Fig. 4, two elements are added empirically: a collector resistance r_c to take account of the finite reverse resistance of the collector diode, and a resistance r_b' in series with the base connection to take account of the transverse ohmic resistance of the base region.

The incremental diode resistance r_e' can be calculated from (2) as

$$(1/r_e') \equiv (\partial I_E/\partial V_{EB}) = (q/kT)(I_E + I_{EO}) \doteq (qI_E/kT). \quad (3)$$

In Shockley's analysis [1] the short-circuit current transfer ratio $a \approx \alpha < 1$ can be written as the product of two factors, an emitter efficiency γ and a base transport factor β. The factor $(1-\gamma)$ represents that part of the emitter current that is composed of electrons injected into the emitter from the base which are not available for collection. The factor $(1-\beta)$ represents the small fraction of the injected holes that recombine in the base before reaching the collector. Both of the factors γ and β depend upon the physical parameters of the emitter and base regions, and for decreasing base width, both γ and β approach 1 as a limit.[11]

In theory the reverse resistance of the collector-base diode is infinitely large for collector-base voltages exceeding a few tenths of a volt [cf. (2)]. However, in practice the reverse resistance of a diode generally is dominated by a leakage resistance, which may be of the order of tens of megohms. The base resistance r_b' depends upon the nature of the base contact, but for a good contact r_b' should be of the order of magnitude of the sheet resistance (base resistivity ρ_B divided by base width w) of the base region.

Input Resistance

In practice the short-circuit input resistance h_{ib} of a junction transistor was found to decrease with increasing emitter current I_E, in general agreement with the theoretical result of (3). However, with increasing current, as r_e' becomes small, the input resistance is dominated by the effective[12] ohmic base resistance $r_b'(1-\alpha)$. In some transistors an ohmic resistance of the emitter region also contributes directly to the input resistance.[13]

As noted above, studies of the *p-n* diode have shown that for high current *densities* the forward voltage-current characteristic of (2) must be modified; the exponent kT must be replaced by $2kT$. Accordingly for the triode, (3) also must be modified in general[14] so that $r_e' = (\lambda kT/qI_E)$ with $1 \leq \lambda \leq 2$, and with λ tending to 2 for high current densities [38]. Unfortunately, the increase in input resistance due to this effective increase in diode resistance also can be attributed to emitter lead resistance.[15]

Output Resistance

In early junction transistors, experimental values of collector resistance r_c were considerably lower than the tens of megohms that could be explained by leakage. Furthermore, r_c was found to decrease with increasing dc emitter current. These results were shown by Early [41] to be due to depletion-layer (space-charge-layer) widening. As noted above, an increase in collector voltage causes an increase in the width of the depletion layer, with a consequent decrease in the width of the base. This decrease in base width brings about an increase in both γ and β, and hence in the current transfer ratio α, which in turn causes the collector current to increase. The ratio of the increase in collector current to the corresponding increase in collector voltage represents a collector-base admittance, which is considerably larger than the leakage conductance of the diode.[16]

In addition, as the collector voltage is increased toward the breakdown voltage, the dc collector reverse current increases, giving rise to collector conductance in addition to that due to depletion-layer widening. For example, the collector conductance due to the onset of avalanche breakdown has been calculated explicitly by Matz [36] from the equation given by Miller [29] for the multiplication factor m.

Feedback Factor

For the ideal circuit of Fig. 4 the voltage feedback factor is simply the ratio (r_b'/r_c). However, in early junction transistors, measured values of this feedback factor were considerably larger than could be attributed to the ohmic base resistance alone. This discrepancy also was shown by Early [41] to be the result of deple-

[10] Strickly speaking, the internal current-transfer process is represented by the transfer ratio a, which is different from the *terminal* short-circuit forward current transfer ratio α because of the presence of the base resistance r_b'. At low frequencies the difference is negligible, but at high frequencies a and α may differ considerably when r_b' is of comparable magnitude with the reactance of the collector capacitance (not shown in Fig. 4) in shunt with the collector resistance r_c.

[11] For example, in an early alloy structure, typical values might be $\gamma = 0.998$. $\beta = 0.993$, although measured values of α generally were lower, *e.g.*, 0.98; this is attributed to surface recombination to be described below.

[12] The voltage drop across r_b' is due to base current, which is less than the input emitter current by the factor $(1-\alpha)$.

[13] For example, in the diffused-base transistor of Lee [7], emitter contact resistance contributed 10–20 ohms.

[14] This also should be true at *low* current densities in silicon transistors when the depletion-layer recombination current dominates as noted above, but this result apparently has not been reported as yet.

[15] *E.g.*, see [39], p. 535, and [40].

[16] For example, in a typical alloy junction transistor having a value of $\alpha = 0.98$, the output conductance due to depletion-layer widening would be of the order of 0.5 μmho for $I_E = 1$ ma.

tion-layer widening. As a result of the decrease in base width with increase in collector voltage noted above, a given dc emitter current can be maintained with fewer holes injected into the base. This in turn brings about a slight, but significant, decrease in the emitter junction voltage. The ratio of this change in emitter voltage to the corresponding change in collector voltage is the Early voltage feedback ratio μ_{eo}. Alternatively, this feedback can be expressed in terms of a base resistance r_b'' in the base lead of the equivalent circuit, where r_b'' is equal to the quotient of the emitter voltage to the corresponding collector current, or $r_b'' \doteq \mu_{ec} r_c$. Expressed in these terms, the base resistance r_b'' can be many times greater than the ohmic resistance r_b', especially at lower values of dc emitter current for which the collector conductance is small.[17]

Depletion-layer widening also can contribute additional feedback within the transistor by modulation of of the ohmic resistance [41]. Generally, this effect is not especially important, but in some transistors—especially grown-junction units in which the base connection is extremely close to the collector depletion layer—this base-contact modulation can be appreciable and moreover can be of the proper polarity to cause a *negative* input resistance [42].

In some of the early junction triodes at low values of dc emitter current, measured values of the voltage feedback factor were extremely large and, in some cases, approached 1. This behavior was attributed to the "channel effect," described first by Brown [43], in which a P-type surface layer forms on the N-type base and creates a high resistance path, or "bridge" between emitter and collector.[18] Hence, the emitter voltage tends to rise toward the collector voltage, and the voltage feedback approaches one in magnitude.

A feedback factor of one at higher collector voltages also can be obtained as the result of the "punch-through" effect described above [37]. In this case, however, collector and emitter are effectively short-circuited internally, and transistor action ceases.

Forward Current Transfer Ratio

Surface Effects: In Shockley's analysis the short-circuit forward current transfer ratio α is essentially a constant for a given transistor structure. Experimental verification of the relation between α and structural parameters was provided by Goucher, *et al.* [45]. On the other hand, fairly early in the development of the transistor, it was found that α was considerably affected by surface conditions. Inasmuch as surface recombination is a three-dimensional effect, exact analysis is difficult, and only a few calculations have been carried out. For example, Laplume [46] has evaluated the base transport factor for a cylindrical geometry, and has effectively shown that the lifetime is decreased by increasing surface recombination velocity s. An approximate result used frequently (for example, by Webster [52]) consists of expressing $(1-\alpha)$ as the sum of $(1-\gamma)$ and $(1-\beta)$ and of adding to this sum a surface-recombination term $(1-\beta_s)$ that is proportional to surface recombination velocity s and to an effective recombination area A_s.

In order to evaluate β_s for the specific geometry normally used for alloy transistors, Moore and Pankove and Stripp and Moore [47] have carried out very extensive analyses using electrical analog techniques. Also see [48]. Practical results of this study showed, for example, that the collector area should be approximately two to three times as large as the emitter area and that the emitter junction should penetrate less deeply into the germanium wafer than the collector junction in order to maximize β_s for a given surface recombination velocity.

This dependence upon surface conditions of α and of other parameters, *e.g.*, I_{CO}, coupled with the difficulties of maintaining constant surface conditions over a period of time, had given rise to rather serious reliability problems in some of the early transistors. In the meantime, however, considerable effort has been devoted to the study of surface conditions [23]. Some of the factors that cause a change in the surface with time now are understood [49–51], and the results indicate that a high degree of surface cleaning and/or preparation and protection is required for units of high reliability.

Emitter-Current Dependence: In Shockley's analysis [1], α also is bias independent. A relatively small increase in α with increasing collector voltage could be expected as a result of the decrease in base width due to the depletion-layer widening effect noted above. However, in practical transistors α was found to be quite dependent upon the dc emitter current, in particular, decreasing with increasing current. This behavior was shown by Webster [52] to be the result of high-level injection; it has been commented on since by numerous other writers [53–57]. In the type of transistor generally available at that time, for emitter currents of several milliamperes, the concentration of holes injected into the base from the emitter is comparable with the normal concentration of electrons in the base region. Consequently, in order to preserve charge neutrality in the base, additional electrons are drawn into the base region (from the base contact), which increases the base conductivity. This increased base conductivity with increasing emitter current lowers the emitter efficiency γ, and hence the over-all α. To offset the decrease in base conductivity, a higher emitter conductivity is required. In the case of the alloy-junction transistor, this increased understanding, coupled with technological improvements, led to the introduction of high-emitter efficiency materials [58], which resulted in improved transistor performance at high emitter currents.

The theoretical aspects of high-level injection phe-

[17] For example, for the typical alloy transistor noted above, this feedback factor would be of the order of 2.5×10^{-4}. At the dc current of 0.5 ma, this could give rise to an additional base resistance r_b'' of the order of 1000 ohms, whereas ohmic base resistance r_b' might be of the order of 200 ohms.

[18] Although channels can exist on p-n-p transistors [44], the effect was more pronounced on the early n-p-n grown-junction transistors.

nomena also have been discussed by several other writers, including Rittner [59], Misawa [54], [60] and Fletcher [61]. For example, Misawa has described some of the effects of high-level injection on high-frequency parameters of the transistor.

In practical transistors, the short-circuit forward current transfer ratio α exhibits a decrease at *low* values of dc emitter current, as well as at high currents. One explanation for this low-current α decrease was included in Webster's analysis. Even at low current densities, electrons are drawn into the base to preserve charge neutrality. However, to prevent a flow of *electrons* between emitter and collector, an internal emitter-collector electric field is developed, which is proportional to dc current and which tends to reduce the number of holes that are drawn to the surface (where they can recombine). This effectively increases the factor β_s, and hence the current transfer ratio α, with increasing current, up to the point at which high-level injection effects begin to dominate, as described above.

In some of the early silicon transistors [63], it was found experimentally that α decreased with decreasing emitter current to a greater extent than could be attributed to the surface-recombination and electric-field effect previously described. This behavior remained unexplained for a time, during which successful negative resistance four-region two electrode switching devices (*p-n-p-n* diodes) were built utilizing the behavior [62]. Such a device can be considered as a *p-n-p* transistor with a hook collector, which normally has an over-all current-transfer ratio considerably greater than one. However, at low values of dc emitter current, the alpha of the *p-n-p* transistor is so low that over-all $\alpha<1$, whereas at some intermediate value of current over-all $\alpha=1$, and the device becomes unstable and can be used as a switch.

The substantial decrease in α at low current *densities* in such cases was shown by Sah, Noyce, and Shockley [14] to be due to a decrease in the emitter efficiency resulting from the depletion-layer recombination current already noted above. This current, which represents a significant portion of the total emitter current at low current densities, is not available for collection, and hence the over-all current transfer ratio α is low. With increasing current, however, proportionately more of the emitter current consists of hole current into the base which can be collected, and hence α increases with increasing current.

Very recently an increase of α with increasing emitter current has been observed in some types of transistors [96] and has been attributed to an injecting collector contact. Certain types of collector contacts that normally collect holes (for a *p-n-p* transistor) tend to inject electrons into the collector with increasing dc hole current. This gives rise to an over-all increase in collector current, which can be sufficiently large to offset the effects of β and γ and produce an over-all transistor current transfer ratio α 1. In this case the device can be used as a bistable switching element, and the name "thyristor" has been applied to the unit.

Collector-Voltage Dependence: Experiments also indicated that in some transistors a substantial increase in α could be obtained with increasing collector voltage, to the extent that α could exceed 1. This effect can be attributed to avalanche multiplication in the collector depletion layer as described above, which causes an effective multiplication m of the collector current relative to the current entering the collector depletion layer from the emitter. If the multiplication factor m is large enough to offset the effects of γ and β described above, the over-all transistor α can exceed 1. This effect of current multiplication in transistors was studied extensively by Miller [29], who showed for example that for alloy junctions the multiplication factor m is

$$m = 1/[1 - (V/V_B)^n], \tag{4}$$

where $n=5$ and 3 for germanium *n-p-n* and *p-n-p* transistors respectively and $n \doteq 3$ for silicon transistors.

In many cases, a value of $\alpha>1$ can lead to instability and possible destruction of the transistor. For example, in the common-emitter configuration with the base open-circuited, the current between collector and emitter $I_{CO}/(1-\alpha)$ becomes infinite when $\alpha=1$. On the other hand, the transistor *can* be successfully operated with collector-base voltages exceeding that for which $\alpha=1$, in the so-called "avalanche" mode [64] or "delayed-conduction" mode [65] provided that $V_{CB}<V_B$ [66]. In this type of operation, for example, extremely fast switching times have been reported [67].

Actually, values of $\alpha>1$ had been observed in some of the earliest experimental junction transistors but for a different cause from that described above. For those particular units, which were fabricated by the double-doped grown process [2], the collector resistivity was high, and α was observed to exceed 1 at increased temperatures. This effect also was analyzed by Early[19] and found to be the result of collector-body multiplication. Normally, with a reversed biased collector junction, there is a small diffusion current of electrons from the collector into the base which contributes to the reverse current I_{CO}. Associated with this electron diffusion current is an electron diffusion *gradient*, and the space-charge neutrality condition requires a corresponding *hole* diffusion gradient. However, inasmuch as the holes cannot flow back into the base through the reversed biased collector junction, the diffusion gradient must be balanced by an equivalent electric field in the collector body. This field is proportional to the hole current injected from the emitter and is of the proper polarity to increase the electron current back into the base. Hence, the total collector current, which is the sum of hole and electron currents, is increased in proportion to the emitter current. Normally, this effect is slight, but when the collector material becomes essentially an in-

[19] See [42], pp. 1305–1307. An alternative approximate analysis of collector body multiplication is given by Hunter [68].

trinsic semiconductor, *e.g.*, with increasing temperature, the electron current becomes comparable with the hole current, and the total collector current can be appreciably greater than the emitter current, leading to values of $\alpha > 1$ with consequent instability.

Frequency Variations

In addition to the various bias dependences of small-signal parameters noted above, each of the parameters in general depends upon frequency. The analysis of Shockley, Sparks, and Teal [2] showed that the base transport factor β decreases in magnitude with increasing frequency as a result of dispersion in the transit times of holes through the base, which is inherent in the diffusion process. Assuming that of the factors comprising a,[10] only β is frequency dependent,[20] it is convenient to define an a-cutoff frequency, f_a, as the frequency for which $|a|^2$ has decreased to one-half its low-frequency value. In terms of physical qualities f_a is directly proportional to the mobility of holes in the base and is inversely proportional to the square of the base width. There also is a phase shift associated with a due to the transit time of holes through the base, which in most applications is a more serious limitation than the decrease in the amplitude of a [69], [70].

The finite transit time of carriers through the base also causes the other three transistor parameters to become complex and frequency dependent, as was shown by several writers [42], [73], [76] who, almost simultaneously, extended the analysis of the Shockley model to include frequency effects, along with the effects of depletion-layer widening. However, the phase shift associated with the input and output admittances also can be interpreted equally well in terms of diffusion capacitances representing stored charge in the base region. This caused some confusion for a time when a given transistor could be approximated either by an equivalent circuit employing a frequency-dependent current generator or by a circuit in which the current generator was frequency *in*dependent, but in which a large capacitance representing stored charge limited the high-frequency performance.[21]

Soon after the first practical transistors were announced, considerable effort was directed toward attempts to improve their high-frequency performance. However, it also was recognized that factors in addition to the transit time of carriers through the base were limiting the high-frequency performance. In particular, as several writers [72], [39], [75] have shown, to a first approximation, the high-frequency gain is determined by three parameters—the cutoff frequency f_a already described, the ohmic base resistance r_b', and the collector transition capacitance C_{TC}. For example, the collector transition capacitance imposes an upper limit on the output impedance, while the ohmic base resistance imposes a lower limit on the input impedance, whereas the capacitance and/or the base resistance increases internal feedback. Efforts to improve transistor high-frequency performance led to development of structures other than the simple triode transistor, including the junction tetrode of Wallace [74], the *p-n-i-p* transistor of Early [75] and the drift transistor of Krömer [76]. For a summary of the subject, see [97], [98].

Graded-Base Transistors

With the introduction of solid-state diffusion techniques, the assumption of a uniform base region, which was used for the studies described above and which was satisfactory for the earlier alloy and grown-junction transistors, no longer could be considered valid. Nevertheless, measurements of small-signal parameters of diffused transistors, which generally have a *graded* base, did not indicate any significant *qualitative* differences from the results obtained with earlier uniform-base structures. Quantitative differences do exist, however. Because of the variable impurity distribution in the base region, less of the depletion layer extends into the base. Hence, punch-through is less likely to occur, and the Early effect is less pronounced in graded-base transistors than in uniform-base units of comparable base widths.

The classic theoretical study of graded-base transistors is that of Krömer [76] who evaluated the complex small-signal parameters for a structure having an exponential impurity distribution through the base. For low frequencies, these parameters assume forms similar to the corresponding parameters of a uniform-base structure. High-frequency parameters for a general graded-base structure also have been calculated by Moll and Ross [77].

Non-One-Dimensional Current Flow

As new experimental results were obtained during the preceding decade, efforts generally were made to explain them on the basis of a one-dimensional model as described above. At low frequencies and at low currents one-dimensional flow is a good approximation for most transistors, except for the case of surface-recombination previously noted. However, as current and frequency were increased, some of the experimental results indicated clearly that additional non-one-dimensional effects were becoming quite important.

As the dc current input to a transistor is increased, the transverse flow of base current causes an IR drop in the base which generally is in such a direction as to *reverse bias* those portions of the emitter junction furthest from the base contact. This tends to concentrate most of the active emission area close to the base contact.

In small grown-junction structures this internal

[20] Normally this is a good assumption, but if the low-frequency emitter efficiency γ_0 is appreciably different from 1, the variation of γ with frequency may significantly change the nature of the normal a-frequency variation ("slow-drool" effect) [73].

[21] This subject of equivalent circuits of transistors was reviewed earlier by the author [71].

biasing, or "internal tetrode action" [78], causes a reduction in the effective ac base resistance similar to the effect achieved by Wallace [74] in devising the tetrode transistor. (In the tetrode an additional dc current is passed between two base contacts to bring about a reduction in the active area, causing a decrease in base resistance and a resulting improvement in high-frequency performance.)

On the other hand, in power transistors this transverse IR drop in the base is most undesirable [79], [80]. This causes a reduction in the active emitter area at high currents when the large area is especially necessary to avoid a high current density, with its resulting decrease in current-transfer ratio α. In order to minimize the degradation in performance from this internal biasing effect, transistors designed for operation at high currents must have a small transverse emitter dimension. For example, as discussed in more detail by Clark [81], power transistors frequently employ a parallel configuration of long, narrow emitter and base electrodes.

The internal biasing effect also has a high-frequency counterpart, even for the case in which the dc biasing effect is negligible. In this case, the ac base current increases with increasing frequency and biases off portions of the emitter junction, with a resulting reduction in active emitter area. The net result is a reduction in high-frequency base resistance, plus an associated reactive component [82]. This effect is most pronounced in the grown-junction type of transistor, where it was first observed and used to explain a parallel resonance which occurred in the measured high-frequency input impedance of these devices and which could *not* be explained on the basis of a one-dimensional model [78].

Switching Characteristics

For switching applications of a transistor it is common practice to discuss the behavior in three distinct regions of operation, which are the OFF, the active, and the ON, or saturation, regions. Ebers and Moll [83] have analyzed the junction triode in this manner, and they have shown that this device may have very desirable properties as a switch: a low ON impedance (*e.g.*, a few ohms), a high OFF impedance (*e.g.*, many megohms), and fast switching times.

In the OFF region both emitter-base and collector-base diodes are reverse biased, and the OFF current is determined primarily by the reverse currents I_{CO}, I_{EO}. On the other hand, in the ON region, both diodes are forward biased, and the ON voltage generally is only a fraction of a volt, *provided that* series resistances in emitter and collector regions and/or leads are negligibly small. In *some* types of transistors made by diffusion techniques, series resistance, particularly in the collector,[22] *may* limit the ON impedance.

In the active region, which is the transition region between OFF and ON conditions, to a first approximation the transistor is described in terms of the usual small-signal parameters (at some intermediate bias condition). For example, Moll [84] has shown that the time required to switch ON can be calculated fairly accurately from the active-region high-frequency parameters. In particular, the switching time is inversely proportional to the internal cutoff frequency and proportional to the circuit current gain, although if the load resistance is large, collector capacitance also can influence switching time [85]. The actual switching time also is influenced by the bias dependence of the small-signal parameters, but this can be taken into account by resorting to nonlinear analysis methods, *e.g.*, see [86].

On the other hand, when the transistor is initially in the ON, or saturation, condition, the time required to switch to the OFF condition may be significantly larger than the ON switching time. This is due to the so-called "hole-storage" (or more generally, "minority-carrier storage") effect noted first in diodes, in which the OFF switching time is limited by the time required to sweep carriers out of the base region. Moll [84] has shown that this storage time is a function of the input driving current, the *inverse* current transfer ratio α_I and the *inverse* cutoff frequency $f_{\alpha I}$ of the device as well as of the the α and cutoff frequency for normal operation. Additional storage time also may be exhibited by some transistors, due to storage of carriers in the collector region.

In order to avoid these storage-time effects, other types of switching circuits (particularly for high speeds) have been developed, in which the transistor is switched from OFF to a point in the active region, and is never allowed to enter the ON, or saturation, region. This subject is described here in more detail by Henle and Walsh [87].

Temperature Characteristics

Because of the fundamental nature of a semiconductor, many of the transistor parameters are temperature dependent. Furthermore, for each semiconductor material, there is a range of temperatures over which the device will function satisfactorily. An upper limit on temperature is imposed by the semiconductor tending to act as a conductor, whereas a lower limit is set by the semiconductor tending to act as an insulator. For example, in germanium the upper limit is generally accepted to be of the order of 100°C, although devices have been operated at temperatures somewhat above this, while the lower limit is of the order of −150°C.

Probably the most well-known transistor-parameter variation with temperature is that of the reverse current I_{CO}. The saturation component of the reverse current in theory varies exponentially with temperature, to the extent that it doubles for each 8° increase in temperature. This saturation current generally is observed in germanium transistors at normal and at elevated temperatures, with an exponential temperature variation in general agreement with the theoretical value. At lower temperatures, and at normal temperatures in silicon transistors, the diffusion current generally is obscured by the depletion-layer recombination current

[22] For example, in the diffused-base units of Lee [7] collector series (or saturation) resistance is of the order of 100 ohms.

noted above, which has a different rate of temperature variation, or by a surface leakage current which may have a somewhat unpredictable (and perhaps unrepeatable) temperature variation.

Other parameters of the transistor are observed (*e.g.* see [88], [89]) to vary with temperature in the manner expected from the temperature variation of the corresponding physical parameters, as analyzed in some detail by Gärtner [90]. An additional temperature dependence of note is the decrease in the forward current transfer ratio α with decreasing temperature observed in silicon transistors (and to a lesser extent in germanium transistors). This again is attributed [14] to the increase in the depletion-layer recombination current relative to the normal diffusion current with decreasing temperature, with a consequent decrease in the emitter efficiency.

In practical transistor structures the variation of transistor parameters with temperature cannot be separated from the internal variation of temperature with power dissipation [91], although the latter is essentially a thermodynamic problem. However, the importance of this thermodynamic problem was appreciated rather early in the development of power transistors with the discovery of thermal runaway. In particular, with transistors operating at moderately high power levels, the junction temperature increases considerably above the ambient temperature due to the internal power dissipation. This in turn causes an increase in the exponentially-temperature-dependent reverse current, which will further increase the internal power dissipation and increase the junction temperature still further. If the increase in junction temperature due to increase in I_{CO} is sufficiently large, the process may be cumulative, and the current and junction temperature both will increase indefinitely until the device is destroyed. To prevent this behavior, the device must be capable of conducting heat away at a greater rate than heat is generated internally. In mathematical terms, as Saby [92] has indicated, (see also [93]) this requires $V_C(dI_C/dT) > 1/\theta$ where θ is the thermal resistance of the device.

In pulse operation of a transistor, the junction temperature may be proportional to the *instantaneous* power dissipation, if the pulse repetition rate is slow, or proportional to the *average* power dissipation for fast repetition rates. Whether a given rate is fast or slow for a particular transistor depends upon the value of a characteristic time of the device, called the "thermal time constant." However, as Mortenson [94] has shown, because of the distributed nature of the heat-flow problem in a transistor, the concept of a *simple, single* time constant is *not* necessarily valid in all cases, and the maximum junction temperature may be considerably larger than that predicted on the basis of a single-time-constant model.

Noise

As the old-timers in the transistor field will recall, the first transistors (point-contact variety) were noisy little devices. Although the first junction transistors were not as noisy as the corresponding point-contact units, considerable improvement has been made in quieting them and in understanding the origin of noise in transistors. The subject is discussed in considerably more detail by van der Ziel [95].

Conclusion

An attempt has been made to present the highlights of the progress made in the understanding of the junction triode over the past decade. Even though space limitations do not permit a discussion of *all* of the advances, the number of topics that were able to be included represents an imposing list of accomplishments. Moreover, it is expected that considerable effort will be continued to improve the understanding of present triode devices and their future counterparts, so that in another decade an equally imposing list may be prepared.

Bibliography

[1] Shockley, W. "The Theory of *P-N* Junctions in Semiconductors and *P-N* Junction Transistors," *Bell System Technical Journal*, Vol. 28 (July, 1949), pp. 435–489.

[2] Shockley, W., Sparks, M., and Teal, G. K. "*P-N* Junction Transistors," *Physical Review*, Vol. 83 (July, 1951), pp. 151–162.

[3] Wallace, R. L., Jr., and Pietenpol, W. J. "Some Circuit Properties and Applications of *N-P-N* Transistors," *Bell System Technical Journal*, Vol. 30 (July, 1951), pp. 530–563. Also Proceedings of the IRE, Vol. 39 (July, 1951), pp. 753–767.

[4] Saby, J. S. "Recent Developments in Transistors and Related Devices," *Tele-Tech*, Vol. 10 (December, 1951), pp. 32–34, 58.
———. "Fused Impurity *P-N-P* Junction Transistors," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1358–1360.

[5] Bradley, W. E. "Principles of the Surface-Barrier Transistor," Proceedings of the IRE, Vol. 41 (December, 1953), pp. 1702–1706.

[6] Smits, F. M. "Formation of Junction Structures by Solid-State Diffusion," Proceedings of the IRE, this issue, p. 1049.

[7] Lee, C.A. "A High-Frequency Diffused Base Germanium Transistor," *Bell System Technical Journal*, Vol. 35 (January, 1956), pp. 23–34.

[8] Kircher, R. J. "Properties of Junction Transistors," IRE Transactions on Audio, Vol. AU-3 (July–August, 1955), pp. 107–124.

[9] Adler, R. B. "A Large Signal Equivalent Circuit for Transistor Static Characteristics." Unpublished report, Massachusetts Institute of Technology, Cambridge, Mass., 1951.
———. "Analysis and Equivalent Circuits." Oral paper, Transistor Circuits Conference, University of Pennsylvania, Philadelphia, Pa., February 16, 1956.

[10] Henkels, H. "Germanium and Silicon Rectifiers," Proceedings of the IRE, this issue, p. 1086.

[11] Herlet, A., and Spenke, E. "Gleichrichter mit *p-i-n* bezw. mit *p-s-n* Struktur unter Gleichstrombelastung," *Zeitschrift für angewandte Physik*, Vol. 7 (February, 1955), pp. 99–107 (March, 1955), pp. 149–163, and (April, 1955), pp. 195–212.

[12] Prince, M. B. "Diffused *P-N* Junction Silicon Rectifiers," *Bell System Technical Journal*, Vol. 35 (May, 1956), pp. 661–684.

[13] Hall, R. N. "Power Rectifiers and Transistors," Proceedings of the IRE, Vol. 40 (November, 1952), pp. 1512–1518.

[14] Sah, C. T., Noyce, R. N., and Shockley, W. "Carrier Generation and Recombination in *P-N* Juctions and *P-N* Junction Characteristics," Proceedings of the IRE, Vol. 45 (September, 1957), pp. 1228–1242.

[15] Bernard, M. "Mesures en fonction de la temperature du courant dans jonctions de germanium *n-p*," *Journal Electronics*, Vol. 2 (May, 1957), pp. 579–596. (In French.)

[16] Laplume, J. "Calcul du Courant de Recombinaison en Surface dans le Transistor à Jonction Obtenu par Fusion," *Comptes rendus Academie des Sciences, Paris*, Vol. 238 (March 8, 1954), pp. 1107–1109.
———. "Sur le Courant Inverse de Saturation dans un Type de Redresseur à Jonction *P-N*," *Comptes rendus Academie des Sciences, Paris*, Vol. 239 (November 3, 1954), pp. 1126–1128.

[17] Webster, W. M. "Saturation Current in Alloy Junctions," Proceedings of the IRE, Vol. 43 (March, 1955), pp. 277–280.

[18] Rzhanov, A. V. "Influence of Recombination at Contact on the

Volt/Ampere Characteristics of a Rectifier," *Doklady Akademii Nauk SSSR*, Vol. 98 (September 21, 1954), pp. 389–390. (In Russian.)
[19] Plummer, A. R. F. "Observations on the Growth of Excess Current in Germanium *P-N* Junctions," *Proceedings of the Physical Society B*, Vol. 69 (May 1, 1956), pp. 539–547.
[20] Cutler, M., and Bath, H. M. "Surface Leakage Current in Silicon Fused Junction Diodes," PROCEEDINGS OF THE IRE, Vol. 45 (January, 1957), pp. 39–43.
[21] McWhorter, A. L., and Kingston, R. H. "Channels and Excess Reverse Current in Grown Germanium *P-N* Junction Diodes," PROCEEDINGS OF THE IRE, Vol. 42 (September, 1954), pp. 1376–1380.
[22] Eriksen, W. T., Statz, H., and de Mars, G. A. "Excess Surface Currents on Germanium and Silicon Diodes," *Journal of Applied Physics*, Vol. 28 (January, 1957), pp. 133–139.
[23] Kingston, R. H. "Review of Germanium Surface Phenomena," *Journal of Applied Physics*, Vol. 27 (February, 1956), pp. 101–114.
[24] Pell, E. M. "Reverse Current and Carrier Lifetime as a Function of Temperature in Germanium Junction Diodes," *Journal of Applied Physics*, Vol. 26 (June, 1955), pp. 658–665.
[25] Pell, E. M., and Roe, G. M. "Reverse Current and Carrier Lifetime as a Function of Temperature in Silicon Junction Diodes," *Journal of Applied Physics*, Vol. 27 (July, 1956), pp. 768–777.
[26] Rashba, E. I., and Tolpygo, K. B. "The Static Reverse Voltage-Current Characteristics of the Barrier Layer Formed at the Boundary between an *N*-Type Semiconductor and a *P*-Type Semiconductor," *Zhurnal Tekhnicheskoi Fiziki*, Vol. 25 (July, 1955), pp. 1335–1338. (In Russian.)
[27] McAfee, K. B., Ryder, E. J., Shockley, W., and Sparks, M. "Observations of Zener Current in Germanium *P-N* Junctions," *Physical Review*, Vol. 83 (August 1, 1951), pp. 650–651.
[28] McKay, K. G. "Avalanche Breakdown in Silicon," *Physical Review*, Vol. 94 (May 15, 1954), pp. 877–884.
[29] Miller, S. L. "Avalanche Breakdown in Germanium," *Physical Review*, Vol. 99 (August 15, 1955), pp. 1234–1241.
[30] Knott, R. D., Colson, I. D., and Young, M. R. P. "Breakdown Effect in *P-N* Alloy Germanium Junctions," *Proceedings of the Physical Society B, London*, Vol. 68 (March 1, 1955), pp. 182–185.
[31] Chynoweth, A. G., and McKay, K. G. "Internal Field Emission in Silicon *P-N* Junctions," *Physical Review*, Vol. 106 (May 1, 1957), pp. 418–426.
[32] Garrett, C. G. B., and Brattain, W. H. "Some Experiments on a Theory of Surface Breakdown," *Journal of Applied Physics*, Vol. 27 (March, 1956), pp. 299–306.
[33] Pell, E. M. "Influence of Electric Field in Diffusion Region upon Breakdown in Germanium *N-P* Junctions," *Journal of Applied Physics*, Vol. 28 (April, 1957), pp. 459–466.
[34] Shotov, A. P. "The Breakdown of *P-N* Junctions in Germanium by a Voltage Impulse," *Zhurnal Tekhnicheskoĭ Fiziki*, Vol. 26 (August, 1956), pp. 1634–1645. (In Russian.)
[35] Vul, B. M. "Breakdown of Transition Layers in Semiconductors," *Zhurnal Tekhnicheskoi Fiziki*, Vol. 26 (November, 1956), pp. 2403–2416. (In Russian.)
[36] Matz, A. W. "Thermal Turnover in Germanium *P-N* Junctions," *Proceedings of the IEE*, Pt. B, Vol. 104 (November, 1957), pp. 555–564.
[37] Schenkel, H., and Statz, H. "Voltage Punch-Through and Avalanche Breakdown and Their Effect on the Maximum Operating Voltages for Junction Transistors," *Proceedings of the National Electronics Conference*, Vol. 10 (1954), pp. 614–625.
[38] Misawa, T. "A Note on the Extended Theory of the Junction Transistor," *Journal of the Physical Society, Japan*, Vol. 11 (July, 1956), pp. 728–739.
[39] Giacoletto, L. J. "Study of *P-N-P* Alloy Junction Transistors from D-C through Medium Frequencies," *RCA Review*, Vol. 15 (December, 1954), pp. 506–562.
[40] Baldinger, E., Czaja, W., and Nicolet, M. "Der Einfluss Nichtidealer Emitterübergange auf das Verhalten von Flächentransistoren," *Helvetica Physica Acta*, Vol. 29 (December 15, 1956), pp. 428–430.
[41] Early, J. M. "Effects of Space-Charge Layer Widening in Junction Transistors," PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1401–1406.
[42] Early, J. M. "Design Theory of Junction Transistors," *Bell System Technical Journal*, Vol. 32 (November, 1953), pp. 1271–1312.
[43] Brown, W. L. "*N*-Type Surface Conductivity on *P*-Type Germanium," *Physical Review*, Vol. 91 (August 1, 1953), pp. 518–527.
[44] de Mars, G. A., Statz, H., and Davis, L., Jr. "Measurement and Interpretation of Conductivity of *P*-Type Inversion Layers on Germanium," *Physical Review*, Vol. 98 (April 15, 1955), pp. 539–540.
[45] Goucher, F. S., and Prince, M. B. "Interpretation of α Values in *P-N* Junction Transistors," *Physical Review*, Vol. 89 (February 1, 1953), pp. 651–653.
[46] Laplume, J. "Evaluation du Gain en Courant dans le Transistor à Jonction Obtenu par Fusion," *Comptes rendus Academie des Sciences, Paris*, Vol. 238 (March 22, 1954), pp. 1300–1301.
———. "Sur l'Influence de la Recombinaison en Surface sur le Gain en Courant du Transistor à Jonction," *Comptes rendus Academie des Sciences, Paris*, Vol. 239 (November 15, 1954), pp. 1274–1276.
[47] Moore, A. R., and Pankove, J. I. "The Effect of Junction Shape and Surface Recombination on Transistor Current Gain," PROCEEDINGS OF THE IRE, Vol. 42 (June, 1954), pp. 907–913.
Stripp, K. F., and Moore, A. R. "The Effects of Junction Shape and Surface Recombination on Transistor Current Gain—Part II," PROCEEDINGS OF THE IRE, Vol. 43 (July, 1955), pp. 856–866.
[48] Yanai, H., and Sugano, T. "Effects of Surface Recombination on the Current Amplification Factor of Alloy Junction and Surface Barrier Transistors," *Journal of the Institute of Electrical Communication Engineers, Japan*, Vol. 40 (August, 1957), pp. 883–892.
[49] Wahl, A. J., and Kleimack, J. J. "Factors Affecting Reliability of Alloy Junction Transistors," PROCEEDINGS OF THE IRE, Vol. 44 (April, 1956), pp. 494–502.
[50] Wallmark, J. T. "Influence of Surface Oxidation on Alpha-CB of Germanium *P-N-P* Transistors," *RCA Review*, Vol. 18 (June, 1957), pp. 255–271.
[51] Wallmark, J. T., and Johnson, R. R. "Influence of Hydration-Dehydration of the Germanium Oxide Layer on the Characteristics of *P-N-P* Transistors," *RCA Review*, Vol. 18 (December, 1957), pp. 512–524.
[52] Webster, W. M. "On the Variation of Junction-Transistor Current-Amplification Factor with Emitter Current," PROCEEDINGS OF THE IRE, Vol. 42 (June, 1954), pp. 914–920.
[53] Giacoletto, L. J. "Variation of Junction-Transistor Current-Amplification Factor with Emitter Current," PROCEEDINGS OF THE IRE, Vol. 43 (October, 1955), p. 1529.
[54] Misawa, T. "Emitter Efficiency of Junction Transistor," *Journal of the Physical Society, Japan*, Vol. 10 (May, 1955), pp. 362–367.
[55] Fletcher, N. H. "Note on 'The Variation of Junction Transistor Current-Amplification Factor with Emitter Current,'" PROCEEDINGS OF THE IRE, Vol. 44 (October, 1956), pp. 1475–1476.
[56] Hauri, E. R. "Zur Frage der Abhängigkeit der Stromverstärkung von Flächentransistoren vom Emitterstrom," *Technische Mitteilungen von der Schweizerischen Post-Telegraphen-und Telephonverwaltung*, Vol. 34 (November, 1956), pp. 441–451.
[57] Matz, A. W. "Variation of Junction Transistor Current Amplification Factor with Emitter Current," PROCEEDINGS OF THE IRE, Vol. 46 (March, 1958), pp. 616–617.
[58] Armstrong, L. D., Carlson, C. L., and Bentivegna, M. "*PNP* Transistors Using High-Emitter-Efficiency Alloy Materials," *RCA Review*, Vol. 17 (March, 1956), pp. 37–45. Also in *Transistors I*. Princeton, N. J.: RCA Laboratories, 1956, pp. 144–152.
[59] Rittner, E. S. "Extension of the Theory of the Junction Transistor," *Physical Review*, Vol. 94 (June 1, 1954), pp. 1161–1171.
[60] Misawa, T. "Diffusion Capacitances and High-Injection Level Operation of Junction Transistor," PROCEEDINGS OF THE IRE, Vol. 43 (June, 1955), pp. 749–750.
[61] Fletcher, N. H. "The High Current Limit for Semiconductor Junction Devices," PROCEEDINGS OF THE IRE, Vol. 45 (June, 1957), pp. 862–872.
[62] Moll, J. L., Tanenbaum, M., Goldey, J. M., and Holonyak, N. "*P-N-P-N* Transistor Switches," PROCEEDINGS OF THE IRE, Vol. 44 (September, 1956), pp. 1174–1182.
[63] Tanenbaum, M., and Thomas, D. E. "Diffused Emitter and Base Silicon Transistors," *Bell System Technical Journal*, Vol. 35 (January, 1956), pp. 1–22.
[64] Miller, S. L., and Ebers, J. J. "Alloyed Junction Avalanche Transistors," *Bell System Technical Journal*, Vol. 34 (September, 1955), pp. 883–902.
[65] Kidd, M. C., Hasenberg, W., and Webster, W. M. "Delayed Collector Conduction, a New Effect in Junction Transistors," *RCA Review*, Vol. 16 (March, 1955), pp. 16–33.
[66] Schenkel, H., and Statz, H. "Junction Transistors with Alpha Greater Than Unity," PROCEEDINGS OF THE IRE, Vol. 44 (March, 1956), pp. 360–371.
[67] Beale, J. R. A., Stephenson, W. L., and Wolfendale, E. "A Study of High-Speed Avalanche Transistors," *Proceedings of the IEE*, Pt. B, Vol. 104 (July, 1957), pp. 394–402.
[68] Hunter, L. P. "Transistor Action" in *Handbook of Semiconductor Electronics*. New York: McGraw-Hill Book Co., Inc., 1956, Sec. 4, pp. 4-6 to 4-7.
[69] Thomas, D. E. "Transistor Amplifier—Cut-Off Frequency," PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1481–1483.
[70] Pritchard, R. L. "Frequency Variations of Current-Amplification Factor for Junction Transistors," PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1476–1481.
[71] Pritchard, R. L. "Electric-Network Representation of Transistors—A Survey," IRE TRANSACTIONS ON CIRCUIT THEORY, Vol. CT-3 (March, 1956), pp. 5–21.
[72] Pritchard, R. L. "High-Frequency Power Gain of Junction

Transistors," PROCEEDINGS OF THE IRE, Vol. 43 (September, 1955), pp. 1075–1085.
[73] Pritchard, R. L. "Frequency Variations of Junction-Transistor Parameters," PROCEEDINGS OF THE IRE, Vol. 42 (May, 1954), pp. 787–789.
[74] Wallace, R. L., Jr., Schimpf, L. G., and Dickten, E. "A Junction Transistor Tetrode for High-Frequency Use," PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1395–1400.
[75] Early, J. M. "*P-N-I-P* and *N-P-I-N* Junction Transistor Triodes," *Bell System Technical Journal*, Vol. 33 (May, 1954), p. 519.
[76] Krömer, H. "Zur Theorie des Diffusions- und des Drifttransistors," (Parts I, II, and III), *Archiv der elektrischen Übertragung*, Vol. 8 (May, 1954), pp. 223–228, (August, 1954), pp. 363–369, (November, 1954), pp. 499–504.
———. "Der Drifttransistor," *Naturwissenschaften*, Vol. 40 (November, 1953), pp. 578–579.
———. "The Drift Transistor" in *Transistors I*. Princeton, N. J.: RCA Laboratories, 1956, pp. 202–220.
[77] Moll, J. L., and Ross, I. M. "The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity," PROCEEDINGS OF THE IRE, Vol. 44 (January, 1956), pp. 72–78.
[78] Pritchard, R. L., and Coffey, W. N. "Small-Signal Parameters of Grown-Junction Transistors at High-Frequencies," 1954 IRE CONVENTION RECORD, Pt. 3, pp. 89–98.
[79] Fletcher, N. H. "Some Aspects of the Design of Power Transistors," PROCEEDINGS OF THE IRE, Vol. 43 (May, 1955), pp. 551–559.
———. "Self-Bias Cutoff Effect in Power Transistors," PROCEEDINGS OF THE IRE, Vol. 43 (November, 1955), p. 1669.
[80] Gudmundsen, R. A. "A New Type Transistor Tetrode" (abstract), IRE TRANSACTIONS ON ELECTRON DEVICES, Vol. ED-4 (April, 1957), p. 195.
[81] Clark, M. A. "Power Transistors," this issue, p. 1185.
[82] Pritchard, R. L. "Two-Dimensional Current Flow in Junction Transistors at High Frequencies," this issue, p. 1152.
[83] Ebers, J. J., and Moll, J. L. "Large signal Behavior of Junction Transistors," PROCEEDINGS OF THE IRE, Vol. 42 (December, 1954), pp. 1761–1772.
[84] Moll, J. L. "Large-Signal Transient Response of Junction Transistors," PROCEEDINGS OF THE IRE, Vol. 42 (December, 1954), pp. 1773–1784.
[85] Easley, J. W. "The Effect of Collector Capacity on the Transient Response of Junction Transistors," IRE TRANSACTIONS ON ELECTRON DEVICES, Vol. ED-4 (January, 1957), pp. 6–14.
[86] Bashkow, T. R. "Effect of Nonlinear Collector Capacitance on Collector Current Rise Time," IRE TRANSACTIONS ON ELECTRON DEVICES, Vol. ED-3 (October, 1956), pp. 167–172.
[87] Henle, R. A., and Walsh, J. L. "Application of Transistors to Computers," this issue, p. 1240.
[88] Lin, H. C., and Barco, A. A. "Temperature Effects in Circuits Using Junction Transistors" in *Transistors I*. Princeton, N. J.: RCA Laboratories, 1956, pp. 369–402.
[89] Guggenbühl, W., and Schneider, B. "Zur Stabilisierung des Gleichstromarbeitsspunktes von Flächentransistoren," *Archiv der elektrischen Übertragung*, Vol. 10 (September, 1956), pp. 361–375.
[90] Gärtner, W. "Temperature Dependence of Junction Transistor Parameters," PROCEEDINGS OF THE IRE, Vol. 45 (May, 1957), pp. 662–680.
[91] Vasseur, J. P. "Puissances Maxima des Elementes Semiconducteurs à Jonctions," *Annales de Radioélectricité*, Vol. 11 (January, 1956), pp. 1–28.
[92] Saby, J. S. "Transistors for High Power Applications," 1954 IRE CONVENTION RECORD, Pt. 3, pp. 80–83.
[93] Lin, H. C. "Thermal Stability of Junction Transistors and Its Effect on Maximum Power Dissipation," IRE TRANSACTIONS ON CIRCUIT THEORY, Vol. CT-4 (September, 1957), pp. 202–210.
[94] Mortenson, K. E. "Transistor Junction Temperature as a Function of Time," PROCEEDINGS OF THE IRE, Vol. 45 (April, 1957), pp. 504–513.
[95] van der Ziel, A. "Noise in Junction Transistors," this issue, p. 1019.
[96] Mueller, C. W., and Hilibrand, J. "The 'Thyristor'—A New High-Speed Switching Transistor." IRE TRANSACTIONS ON ELECTRON DEVICES, Vol. ED-5 (January, 1958), pp. 2–5.
[97] Pritchard, R. L. "Modern High-Frequency Transistors—A Survey," *Fortschritte der Hochfrequenztechnik*, to be published in 1958.
[98] Hall, R. N. "Fabrication Techniques for High-Frequency Transistors," *Fortschritte der Hochfrequenztechnik*, to be published in 1958.

Lumped Models of Transistors and Diodes*

JOHN G. LINVILL†

Summary—**Lumped models are shown which can be used to approximate the properties of transistors and diodes over a wide range of conditions and applications.**

Two analytical procedures lead to a lumped model approximating a distributed system. In the familiar one, a differential analysis of the distributed system is made, subsequently a rational approximation is made to the transcendental functions resulting from the differential analysis. In the other, one makes a lumped model at the outset to approximate the distributed system and analyzes it. The lumping approximation at the outset generally simplifies the analysis and permits consideration of phenomena which would be prohibitive to analyze on a differential basis. Throughout, it provides a close tie of analysis to the physical phenomena involved.

A lumped model for diffusion transistors is shown here which is analogous in end result to the Ebers-Moll model. The simple model is subsequently amended to account for the drift phenomenon, photo effects and avalanche multiplication.

* Original manuscript received by the IRE, December 9, 1957; revised manuscript received, March 25, 1958. This work was supported by a grant from the National Science Foundation, NSF G-2426, and by the U. S. Army Signal Corps, U. S. Air Force, and the U. S. Navy through ONR contract Nonr 225(24) at Stanford Electronics Labs.

† Dept. of Electrical Engineering, Stanford University, Stanford, Calif.

INTRODUCTION

SEMICONDUCTOR devices in various forms bring a new range of useful phenomena to electronics. There are two kinds of carriers, and both diffusion and drift flow are significant in many cases. Other physical phenomena, photogeneration of carriers, ionization through the field and the ultimate avalanche, the depletion capacitance with its nonlinear performance, are new tools for performance of useful functions. At the same time, one must account in most applications for the transit time of carriers in the devices. The distributed nature of the transistor and diode are apparent in many applications. Thus the transistor simultaneously brings new opportunities and presents new problems to the circuit designer.

Circuit design depends heavily upon the possibility of approximating the electrical performance of the transistor by circuit models which are simple enough to facilitate invention and design and sufficiently accurate in predicting the performance which is experimentally obtained. A number of circuit models of transistors have

been developed for small-signal applications by Early, Pritchard, Giacoletto, Zawels and others; numerous refinements of these have been employed.[1] Ebers and Moll have shown how diffusion transistors can be satisfactorily represented for a range of large-signal applications.

The lumped circuit models presented below are selected to be mathematically tractable. The laws established between their parts approximate corresponding ones in the physical device. The advantage brought by use of the lumped models is the possibility of treating phenomena in addition to diffusion, and of providing additional insight in the simple diffusion model.

An outline of the mode of analysis used to obtain the earlier representations is useful as an introduction to the present approach. First, equations are written expressing the laws of flow, generation and recombination. Using these laws, differential equations are written to express continuity of hole and electron flow. Partial differential equations arise because of the distributed nature of the processes in the transistor. Second, the partial differential equations, simplified to particular cases and boundary conditions, are solved to give relationships between the terminal variables. The results of the second step are terminal relationships between current and voltage which are transcendental functions of frequency. For convenience, the transcendental functions, in a third step, are approximated by rational functions and these are used by the circuit designer. It is interesting to observe that the analysis has, in the second step, made some compromises in special cases and in geometry to give partial differential equations which are soluble. Then, in a third step, further approximations are made which mask the fact that the system being represented is distributed. The analysis used in this paper to lead to the lumped models has three steps as well; the first is the same as that described above. The second step involves approximation to the equilibrium conditions by writing corresponding equations for finite elements instead of differential elements. This gives ordinary differential equations rather than partial differential equations. The third step is the solution of the differential equations written in the second step.

The procedures are compared conveniently through Fig. 1 which represents them in block form. In simple cases the two procedures lead to the same results, though the form is different. The lumped analog of the Ebers-Moll model is a case in point. In cases where more complications arise in the form of additional significant phenomena, for instance photo-multiplication, the drift mechanism and avalanche processes, new results are obtained in simple form. The lumped representation has the advantage of relating terminal behavior to internal processes on a one-to-one basis retaining

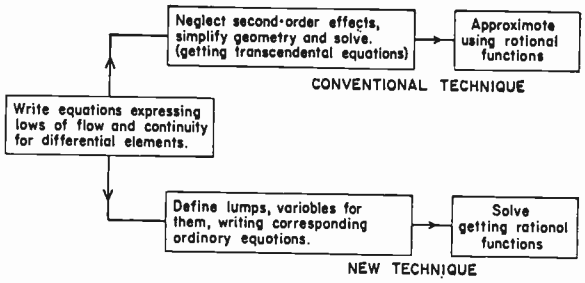


Fig. 1—Comparison of conventional analysis technique with that used for lumped models.

simplicity through early rather than final use of the lumping approximation.

Lumped models have often been used to approximate distributed elements in other systems. Electric transmission lines, cavity resonators, thermal systems and acoustic systems have frequently been approximated in this way. The method is particularly advantageous for the transistor case because of the larger number of the kinds of processes occurring in the device.

Physical Processes and Their Representation

In a volume element of semiconductor material, Fig. 2, which is sufficiently small, the densities of holes, electrons and charge density can be considered constant throughout the volume having values of p, n and ρ, respectively. Accounting for the donor and acceptor densities, N_D and N_A, the densities are related according to (1), where q is the magnitude of charge on the electron:

$$\rho = (p - n + N_D - N_A)q. \tag{1}$$

In uniform material in equilibrium, the net charge density is zero. The density of holes is related to the electron density in the case of uniformity and equilibrium by

$$pn = n_i^2, \tag{2}$$

where n_i is the density of electrons or holes in intrinsic or undoped material. This equation arises through equality of the rates of generation and recombination of carriers. Consideration of (1) and (2) for uniform material reveals that for doping which increases hole population ten-fold or more, the hole density is about $N_A - N_D$. A corresponding equation can be written for n material.

Frequently, it is convenient to work in terms of the equilibrium densities which are functions only of the doping of the material at fixed temperatures. The symbols used for equilibrium densities of p and n materials, respectively, are p_p, n_p and p_n, n_n. This procedure makes it possible to deal in terms of excess densities as variables for p material using,

$$p_e = p - p_p, \tag{3}$$

$$n_e = n - n_p. \tag{4}$$

The analogous expressions for n material are

[1] Readers are referred to an excellent review paper by R. L. Pritchard, "Electric-network representation of transistors—a survey," IRE Trans. on Circuit Theory, vol. 3, pp. 5–21; March, 1956.

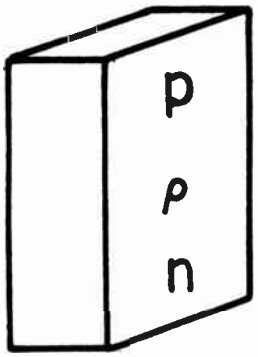


Fig. 2—Variables in a volume element of semiconductor.

$$p_e = p - p_n, \tag{5}$$

$$n_e = n - n_n. \tag{6}$$

Flow Relationships

Carriers flow between two contiguous volume elements of material, by two processes, drift and diffusion. (See Fig. 3). Drift flow is proportional to field strength and mobility of carriers; diffusion flow is proportional to density gradient and diffusion constant. In terms of the volume elements shown in Fig. 3, one deals with electrostatic potentials of the two volume elements v_1 and v_2, and densities of carriers p_1, n_1 and p_2, n_2. Actually the potential and the densities are functions of space inside the volume element, and the values that are indicated by the subscripts 1 and 2 can be taken at a particular point, perhaps at the middle of the volume element. If the volume elements are small to the point of vanishing, the expressions given below become precise. When they are finite, the expressions are approximations. Finally, when one deals with a lumped model of a transistor, correspondence between variables at the internal points of the transistor and at certain points in the model become vague. The model, as with any model or so-called equivalent circuit, is never equivalent to a physical device. It is a simpler mathematical structure having the same terminal variables as a physical device, these being related to some internal variables selected by the analyst. The relationships between these variables are also selected by the analyst, usually to approximate physical relationships, but also selected for compromise between accuracy and mathematical convenience.

One can identify the drift and diffusion components of the currents between volume elements of Fig. 3 as follows:

$$i_p = i_{p\mu} + i_{pD} \tag{7}$$

$$i_n = i_{n\mu} + i_{nD} \tag{8}$$

The drift components of current can be written as

$$i_{p\mu} = \frac{(p_1 + p_2)}{2}(v_1 - v_2)\mu_p F, \text{ and} \tag{9}$$

$$i_{n\mu} = \frac{(n_1 + n_2)}{2}(v_1 - v_2)\mu_n F. \tag{10}$$

where μ_p and μ_n are the mobilities of holes and electrons

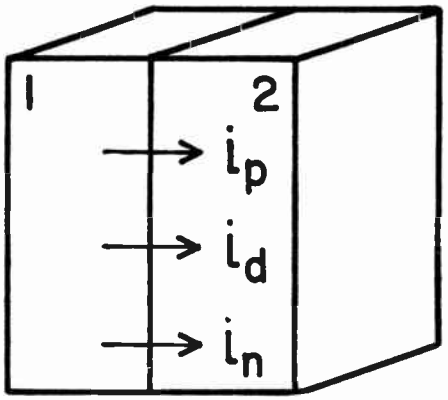


Fig. 3—Carrier flow between contiguous volume elements.

respectively, and F is a function depending only upon the geometry of the volume elements.

The diffusion components of current correspondingly are written as

$$i_{pD} = (p_1 - p_2)D_p F, \text{ and} \tag{11}$$

$$i_{nD} = (n_2 - n_1)D_n F \tag{12}$$

where D_p and D_n are diffusion constants of holes and electrons. Mobility and diffusion constants are related by the Einstein relationship,

$$\frac{\mu}{D} = \frac{q}{KT}, \tag{13}$$

K being Boltzmann's constant, and T being the absolute temperature.

The differential analogs of (7) through (10) are familiar. In them F reduces to the ratio of the area of the interface of the volume elements divided by the thickness of the elements.

In addition to the physical flow of carriers across the interface between volume elements, there can be a flow of displacement current across the interface corresponding to a change in the charge densities of the regions separated by the interface. The law relating to the displacement current involves Poisson's equation and will be established when continuity equations are written.

For purposes of visualization, it is convenient to think of the hole, the displacement and the electron currents as having separate points of assembly, and to associate with these points the flows of holes current, displacement current or electron current to and from the volume element. Thus one is led to the representation of Fig. 4.

Continuity of Hole, Electron and Displacement Currents

If the flow of holes into a volume element from its interfaces with other volume elements exceeds the flow of holes out of the volume element, then (neglecting generation and recombination for the moment) the hole population of the element increases by the integral of the excess of flow in over flow out. If this is the case in volume 1 in Fig. 4, the node labelled h_1 receives more hole current than leaves it. One can attach a storage circuit element to h_1, and then prescribe a continuity of

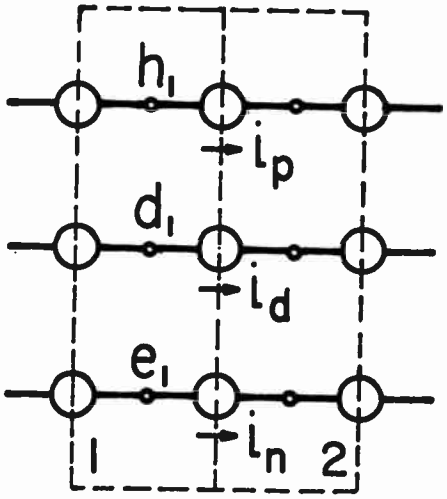


Fig. 4—Use of nodes to represent assembly points of holes, displacement current and electrons.

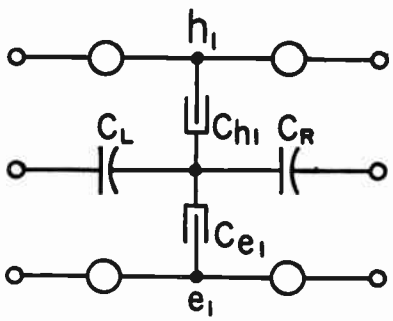


Fig. 5—Storage elements for holes, electrons and charge.

current relationship for the node (like Kirchhoff's current law). This is shown in Fig. 5.

The charge on the hole capacitor, C_{h1}, is the hole charge in the volume element. The flow into this element is governed by a continuity equation of hole current for the volume element.

In a precisely similar way one can deal with the storage of electrons through attachment of a storage circuit element at the node e_1.

If the current from node h_1 through C_{n1} is not equal to the current into the node e_1 through C_{e1}, there is a change in the charge density of the volume element. When there is a rate of change of stored charge, by Poisson's equation a displacement current must emanate from the volume element. Thus the storage circuit elements for electrons and holes must connect to the line carrying displacement current.[2] In the case of a one-dimensional geometry as shown in Fig. 5, the displacement current can flow either to the right or to the left to the contiguous elements. The circuit element carrying the current is an electric capacitor.

The charge on the electron capacitor, C_{e1}, is the electron charge in the volume element. The current flow into it is determined by the continuity equation at the electron node for the volume element. The electric capacitor elements C_L and C_R are charged to the difference

[2] A suggestion made by D. S. Gage.

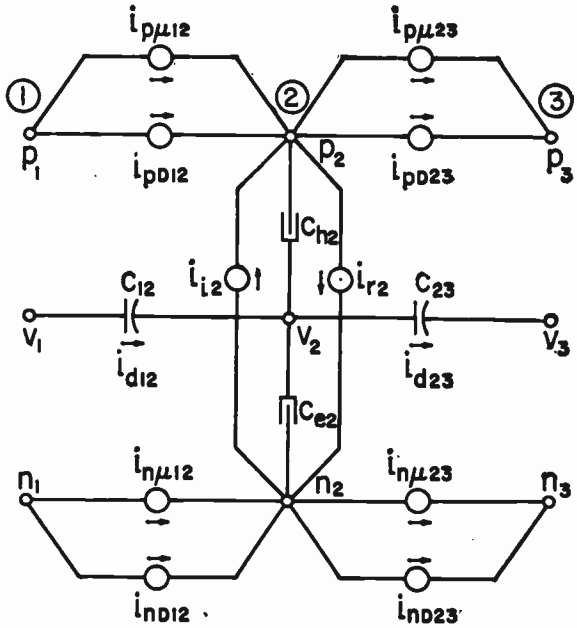


Fig. 6—Addition of generation-recombination sources to model.

in electrostatic potential of the contiguous volume elements. The net charge on them can change only by the development of a space charge in the volume element. A space charge develops when there is a difference in the current flow in the hole and electron capacitors.

Generation and Recombination

The representation for flow and storage in the volume element shown in Fig. 5 is complete except for the generation and recombination of hole-electron pairs inside the volume element. Generation of hole-electron pairs occurs through absorption of light energy, radiation or thermal energy. In the presence of flows of carriers in sufficiently high fields, hole-electron pairs are formed by ionization through collision. Generation and recombination occur throughout the volume element, but in terms of the model shown in Fig. 5 the effect can be represented by a current source between e_1 and h_1 which, over a period of time, carries a charge equal to the *net* generation of carriers in the volume over that period of time. Such a representation is shown in Fig. 6. The current generators shown there are separated to illustrate the phenomena of generation and recombination, and also to show the components of conduction of holes and electrons from volume element to volume element.

In the absence of external sources of generation, and for the simplest mechanism of recombination, the net recombination current for a volume element is proportional to the excess density of minority carriers for that element and inversely proportional to the lifetime of minority carriers. Thus, if the lump of Fig. 6 is in n material,

$$i_{r2} = (p_2 - p_{n2})\frac{H}{\tau_p} = \frac{p_{e2}H}{\tau_p}, \qquad (14)$$

where τ_p is the lifetime of holes in the region and H is

proportional to the volume of the element.

The external ionization current i_{i2} may be the result of a number of different causes. It may be photo-ionization current in which case the current is proportional to the illumination. It may be due to ionization by collisions. In this case, which is particularly significant in the case of avalanche breakdown,

$$i_{i2} = i_{p12} f_h(v_1 - v_2) + i_{n12} f_e(v_1 - v_2), \tag{15}$$

where f_h and f_e are functions of the electric field between volume elements 1 and 2 and the geometry of the element. The quantities f_h and f_e are observed to be proportional to the ionization coefficients for holes and electrons, respectively.[3]

In correspondence to (9)–(12),

$$i_{p\mu 12} = (v_1 - v_2) \frac{(p_1 + p_2)}{2} G_{p\mu 12} \tag{16}$$

where $G_{p\mu 12}$ is proportional to hole mobility and dependent upon the geometry of the volume element.

$$i_{pD12} = (p_1 - p_2) G_{pD12} \tag{17}$$

where G_{pD12} is proportional to the hole diffusion constant and dependent upon the geometry of the volume element.

$$i_{n\mu 12} = (v_1 - v_2) \frac{(n_1 + n_2)}{2} G_{n\mu 12} \tag{18}$$

$$i_{nD12} = (n_2 - n_1) G_{nD12}. \tag{19}$$

Incidentally,

$$\frac{G_{p\mu 12}}{G_{n\mu 12}} = \frac{G_{pD12}}{G_{nD12}} = \frac{\mu_p}{\mu_n}. \tag{20}$$

At this point there are a number of interesting observations in connection with Fig. 6. First, the representation is comprehensive in that all phenomena of interest in a transistor volume element can be accounted for with it. As will be seen presently, in practically all cases, one can, by good approximations, simplify the situation enormously from the complete one shown there. A second observation is that the currents shown by the sources are determined either by independent external quantities, like the amount of light falling on the element, or by the variables stored on the storage elements. The currents into the storage elements are always determined by continuity requirements.

In the figures to this point a one-dimensional geometry has been assumed which is simple but not necessary.

Nothing has been assumed about a homogeneous material, in fact the whole of a transistor from emitter connection to collector connection, junctions and all, could be analytically "sliced" into volume elements and

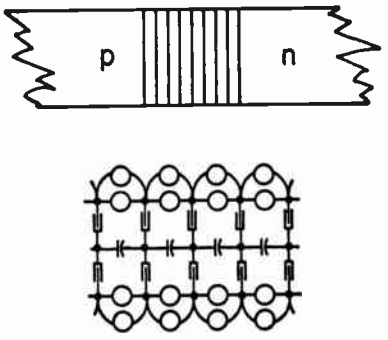


Fig. 7—*P-N* junction as represented by a lumped model.

treated in the same way. In general, such a procedure would not be simplest, but is certainly possible.

Behavior of Junctions

The *pn* junction is the fundamental building block in the operation of diodes and transistors. It is simply a boundary between regions of *p* and *n* material in which the change is quite abrupt from one type to the other. Fig. 7 shows such a junction with adjoining material cut into a large number of slices parallel to the junction. For simplicity, the circuit elements having to do with the generation and recombination processes are omitted from the model. Generally for abrupt junctions these can be neglected, though this is not always possible.

The equilibrium density of holes decreases rapidly from the *p* material to the *n* material, and the equilibrium density of electrons increases rapidly. Reflection upon (11), (12), (17) and (19) indicates that there will be very large components of diffusion currents of holes from left to right and electrons from right to left. The storage elements, by such currents, are charged so that the *p* material becomes negative and the *n* material becomes positive. The result is large drift currents of holes flowing from right to left and of electrons from left to right. The drift and diffusion components of current counterbalance each other and are much larger than the net currents typically flowing through the junction. By virtue of the counterbalancing it is possible to approximate closely the relationships between potential and density of carriers.

The differential forms of (9) and (11) are equated to express the balancing

$$-\mu_p p \frac{dv}{dx} = D_p \frac{dp}{dx}, \text{ or,} \tag{21}$$

$$-\frac{\mu_p}{D_p} dv = \frac{dp}{p}, \tag{22}$$

which can be integrated between points *a* and *b* to yield, with the help of (13),

$$\frac{p_a}{p_b} = \exp \frac{q(v_b - v_a)}{KT}. \tag{23}$$

[3] K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, vol. 94, pp. 877–884; May 15, 1954.

[4] C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in *p-n* junctions and *p-n* junction characteristics," PROC. IRE, vol. 45, pp. 1228–1242; September, 1957.

In an analogous fashion one can obtain,

$$\frac{n_a}{n_b} = \exp \frac{q(v_a - v_b)}{KT} . \tag{24}$$

Eqs. (23) and (24) can be used to point up some very interesting features about the behavior of semiconductor devices. In the first place, a changing density of carriers brought on by a grading in the doping of a semiconductor always results in the development of an internal electrostatic field which is fully committed to counteracting the diffusion of carriers. The field cannot, in addition, be used as a source of power and voltmeters requiring power do not detect it externally. However, an n region, for instance one as in Fig. 7, in which the n region becomes less heavily doped as one moves in the region away from the junction will develop an electrostatic field from left to right which "speeds" immigrant holes from the p region toward the right. This phenomenon is the basis of operation of drift transistors.

The application of an external voltage to the junction upsets the balance between drift and diffusion. The result is that new densities of the carriers are established, bringing about a balancing again. The application of a voltage v from p to n, as in Fig. 8, results in a change of hole density in the n material and of electron density in the p material in accordance with

$$p = p_n \exp \frac{qv}{KT} \tag{25}$$

and

$$n = n_p \exp \frac{qv}{KT} . \tag{26}$$

The majority carrier densities are not influenced by the application of the voltage being equal to the density of dopant in the region.[5]

It is convenient to represent a junction by a single lump on the basis of the above considerations. Such a representation relates the minority carrier densities to the voltage *applied* across the junction. It, in addition, neglects net recombination in the junction volume with the result that the input hole (or electron) current equals the output hole (or electron) current. Finally a capacitance is shown across the junction providing a path for the flow of displacement current. The capacitance is a nonlinear one equal to the derivative of charge put on the junction with junction voltage. Fig. 8 shows such a model for the junction expressing the corresponding "laws of the junction" given in (27)–(30).

[5] I. A. Lesk and J. L. Moll have pointed out that this statement and (27) and (28), which assume insufficient forward bias to modify conductivity significantly, should be generalized to include the case of very large signals. For the very large signal case, (23) and (24) still apply. In addition, the requirement of space-charge neutrality means that n_e on the p side is the same as the excess density of holes there, and p_e on the n side is the same as the excess density of electrons there. Finally:

$$\frac{p(p \text{ side})}{p(n \text{ side})} = \frac{p_p + n_e}{p_n + p_e} \doteq \frac{n(n \text{ side})}{n(p \text{ side})} = \frac{n_n + p_e}{n_p + n_e} = \epsilon^{\frac{q(v_{\text{built-in}} - v_{\text{applied}})}{KT}} .$$

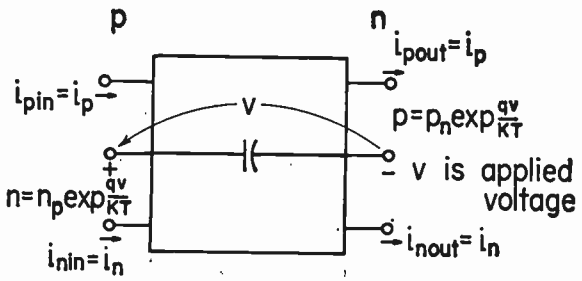


Fig. 8—One-lump model of a junction.

For the n material at the edge of the junction,

$$p = p_n \exp \frac{qv}{KT} . \tag{27}$$

For the p material at the edge of the junction,

$$n = n_p \exp \frac{qv}{KT} . \tag{28}$$

Hole and electron flows are continuous through the junction.

$$i_{p \text{ in}} = i_{p \text{ out}} \tag{29}$$

$$i_{n \text{ in}} = i_{n \text{ out}} . \tag{30}$$

Eqs. (27) and (28) represent the most significant nonlinearities in semiconductor devices. The balancing of drift and diffusion phenomena is the foundation of this nonlinearity.

Models of Particular Semiconductor Devices

In the preceding section, the phenomena of interest have been described with indications made of models which are useful to approximate the relationships between the variables. In most practical situations many of the phenomena described will not be of central interest, and one can employ simplified models which are good approximations and at the same time are much easier to deal with. A sequence of models will illustrate applications involving different phenomena.

In connection with the applications of models it is important to distinguish *device* models and *circuit* models.[1] A model which has internal variables that represent or are analogous to corresponding ones in a device is a device model. The engineer working with the design of transistor devices is principally interested in device models. The circuit model has external variables which correspond to the terminal variables applying to the device; the internal parameters of device and model have possibly no direct relationships to each other at all. Some device models are useful as circuit models, though they contain information which is really not relevant to the circuit application of the transistor. Typically, for circuit applications, the usefulness goes down rapidly as the complication goes up. Lumped models are generally device models in the sense that they have internal variables which are analogous to the internal variables of the

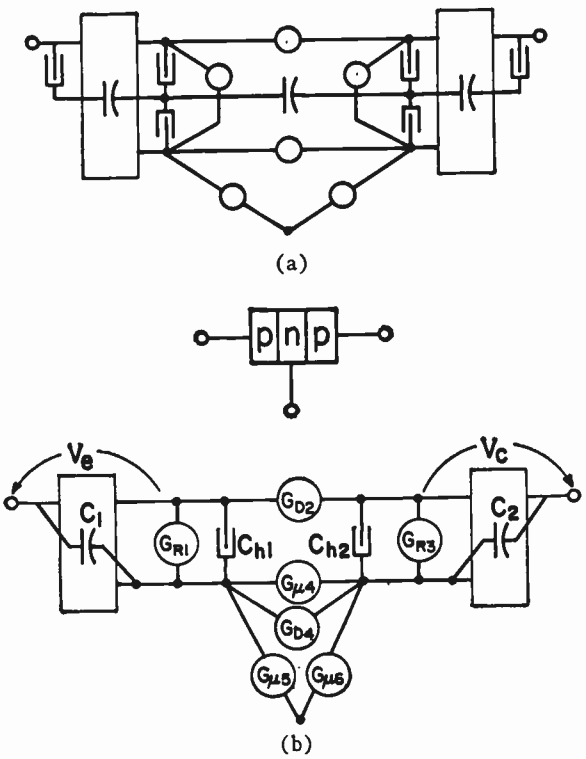


Fig. 9—Lumped diffusion models of a transistor.

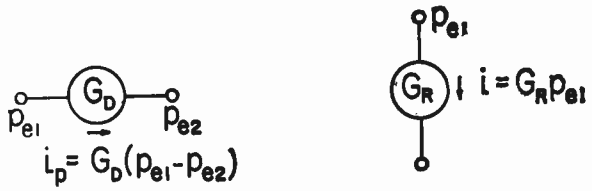


Fig. 10—Linearity of diffusion and recombination conductances.

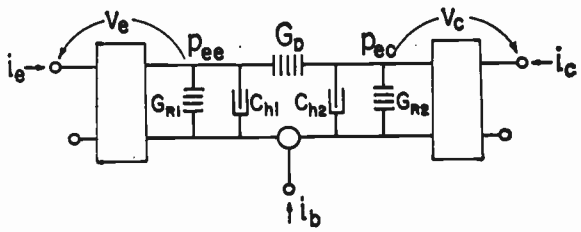


Fig. 11—The reduced lumped diffusion model of a transistor.

device. However, for circuit applications the *form* of the model may be chosen from the physical principles, and the parameter *values* may be chosen from external measurements on the transistor. In such a model, it is possible for one indirectly to account for phenomena which are not represented in the form of the model. For example, surface recombination is an important component of recombination, but the volume recombination parameter may be adjusted in value to account for some of the effects of surface recombinations without introducing an additional lump in the model to make the representation better from a device point of view.

The Diffusion Model of Transistors and Diodes

In many transistors and diodes in which the p and n regions are uniform and signal levels are moderate, the flows of minority carriers are largely by diffusion. Very small fields are required to provide the necessary flow of majority carriers because of their high density, and these fields, because of the much smaller density of minority carriers, provide insignificant amounts of minority carrier flow. One can make the approximation that minority carriers flow only by diffusion and neglect displacement current in the p and n regions. The corresponding simplified model of the device is a diffusion model.

A lumped model of a p-n-p transistor in which the n region is represented by two lumps is shown in Fig. 9. The emitter-base and collector-base junctions are assumed to have unity emitter efficiency; all of the current crossing the junctions is assumed to be holes. (In the physical transistor at the emitter and collector terminals, which are remote from the junction, there are ohmic contacts and the current is again carried by electrons.) The elements of the model of Fig. 9(b) are simplified in a number of ways. The fact that there is no displacement current in the n region leads to the omission of the electric capacitor between the lumps and the collapsing of the majority carrier storage elements which, with a signal, store the negative of the charge stored by the minority storage elements and have no interaction with the rest of the circuit.

In the model of Fig. 9(b),

G_u's are drift current generators,

G_D's are diffusion current generators,

G_R's are net recombination current generators.

The diffusion and recombination elements applying to minority carriers have currents which are linear in the excess density of minority carriers as is shown in Fig. 10. Diffusion conductances are apparently analogous to electrical conductances except that the variable excess density is substituted for electrical potential. In addition, recombination conductance bears a somewhat analogous relationship to an electrical conductance, one terminal of which is grounded. The current through the recombination conductance is directly proportional to the excess density of minority carriers at one terminal. One can emphasize this relationship by use of a symbol as shown in Fig. 11 which omits the depletion capacitance of the junctions for simplicity and combines the majority carrier generators in the base. The neglecting of the electric field in the base, and the potential drop across it in comparison to the junction voltages, reduces the role of the current generators of majority carriers to the function of satisfying Kirchhoff's current law.

At first sight, the lumped model of Fig. 11 appears to be a crude approximation to a transistor. However, a simple analysis reveals that it is equivalent to the Ebers-Moll model for large-signal behavior.[6] It is simple

[6] J. J. Ebers and J. L. Moll, "Large-signal behavior of junction transistors," PROC. IRE, vol. 42, pp. 1761–1772; December, 1954; J. L. Moll, "Large-signal transient response of junction transistors," PROC. IRE, vol. 42, pp. 1773–1784; December, 1954.

to relate the parameters of the model to the normal and inverted α's, the normal and inverted alpha cutoff frequencies and the collector and emitter saturation currents.

The normal alpha, α_N, is defined as the ratio of collected current to emitted current when the collector is held to zero volts. For zero volts on the collector, p_{ec} is zero and

$$\alpha_N = -\frac{i_c}{i_e} = \frac{p_{ee}G_D}{p_{ee}(G_D + G_{R1})} = \frac{G_D}{G_D + G_{R1}} . \quad (31)$$

Similarly,

$$\alpha_I = \frac{G_D}{G_D + G_{R2}} . \quad (32)$$

The collector saturation current, I_{co}, is the collector current which flows when the emitter is open and the collector is heavily reverse biased. In this condition,

$$p_{cc} = -p_n \quad (33)$$

$$I_{co} = \frac{-p_n(G_{R1}G_D + G_{R1}G_{R2} + G_{R2}G_D)}{(G_{R1} + G_D)} . \quad (34)$$

In a similar manner the emitter saturation current, I_{eo}, is given as follows.

$$I_{eo} = \frac{-p_n(G_{R1}G_D + G_{R1}G_{R2} + G_{R2}G_D)}{(G_{R2} + G_D)} . \quad (35)$$

Simultaneous consideration of (31), (32), (34), and (35) reveals that

$$\frac{\alpha_N}{\alpha_I} = \frac{I_{co}}{I_{eo}} , \quad (36)$$

as Ebers and Moll show.

The alpha cutoff frequencies, ω_N and ω_I, can be evaluated from the model as follows. When the collector is held at a fixed potential, p_{ec} is substantially constant. When a unit step of current is injected into the emitter, the excess density approaches a steady value with a transient having a time constant which is the reciprocal of ω_N.

$$\omega_N = \frac{G_{R1} + G_D}{C_{h1}} \quad (37)$$

$$\omega_I = \frac{G_{R2} + G_D}{C_{h2}} . \quad (38)$$

Fig. 11 illustrates the fact that the nonlinearity in the simple diffusion model is totally restricted to the junctions. It emphasizes the advantage for large-signal analysis, pointed out by Ebers and Moll, of dealing with excess densities and currents in the base region. Other treatments leading to the ladder configuration somewhat like Fig. 11 have used the variables voltage and current, and have included a small-motions linearization of the junction relationships in the models. This procedure spoils the model for large-signal applications.

Numerous refinements on the representation given by Fig. 11 can be made at the expense of additional complication. Increasing the number of lumps in the base region from two to three permits a closer approximation to the terminal performance of what is a distributed system. This improvement is the analog of using a two-pole approximation for the transcendental expression for alpha. If the emitter efficiency of the transistor is not unity, then one can account for the behavior of the emitter body in which it is important to consider the flows of minority carriers, which are electrons there. Techniques similar to this will be illustrated in connection with diodes from which the extension to the transistor case is clear. If the recombination of carriers is due principally to surface recombination at a point away from the inter-electrode area, it is possible to add additional lumps with recombination characteristics matched to the surface recombination phenomenon. Further, one can use more adequate laws, the Shockley-Read-Hall one, for example, to represent the recombination phenomenon rather than to use the simple kind used in Fig. 11. Generally, in order to simplify the analysis, one will use the simplest model which can be tolerated for a given application.

The five parameters of Fig. 11 can be evaluated within an undetermined factor p_n, which doesn't influence terminal performance, using the same five measured quantities defined by Ebers and Moll, namely, I_{co}, α_N, α_I, ω_N and ω_I. These serve to determine the lumped parameters G_{R1}, G_{R2}, G_D, C_{h1} and C_{h2} through (31), (32), (34), (37) and (38). Thus the parameters of the model of Fig. 11 are easily determined from circuit performance data.

A lumped diffusion model of a p-n diode employing two lumps in each of the regions is shown in Fig. 12. In this model, the displacement currents have been neglected and the usual three lines are accordingly reduced to two. The conduction mechanism for majority carriers is illustrated by the current sources in the hole line on the p side and in the electron line on the n side. The mechanism involves both drift and diffusion. Interestingly, the majority currents do not come into simultaneous consideration with diffusion of the minority carriers when one neglects voltage drops in the p and n regions. The hole current on the p side, for instance, is determined by continuity of the hole current in the n side where it is entirely determined by diffusion.

The law of conduction of the model shown in Fig. 12 for the dc case is

$$i = I_s\left(\exp\frac{qv}{KT} - 1\right), \quad (39)$$

where

$$I_s = n_p\left(G_{R2} + \frac{G_{R1}G_{D1}}{G_{R1} + G_{D1}}\right) + p_n\left(G_{R3} + \frac{G_{D2}G_{R4}}{G_{D2} + G_{R4}}\right). \quad (40)$$

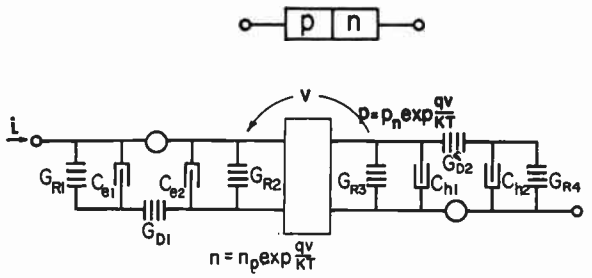


Fig. 12—A lumped diffusion model of a diode.

The division of current crossing the junction between hole and electron flow is determined by the properties of the two regions. The electron current is the first term of (40); hole current is the second term. It is clear that high conductivity on a particular side predisposes that side to correspondingly low minority carrier density. When the elements on the opposite sides (G's of Fig. 12) are about equal, the high conductivity material emits its carriers across the junction more than it receives the majority carriers from the other side.

Transients in the diode resulting from switching operations can be calculated readily using the model shown in Fig. 12. In many interesting cases, particularly those of heavy storage, the external circuit supplying the diode appears as a current source since the voltage across the diode is small compared to the external circuit voltage. In such a case, computation of the transients of excess density of carriers in the diode is very simple and is purely a linear problem. It is precisely the same as computation of transients in an RC network. In a subsequent and independent step one finds the variation of the junction voltage with time through use of the exponential law of the junction.

The following computation of transients illustrates the use of a lumped model of a diode. The diode is assumed to have unity emitter efficiency from the p to the n material. Thus the first component of (40) is negligible in comparison to the second component. The diode has the following parameters:

$$p_n = 10^{12} \text{ per cc},$$

$$G_{R3} = G_{R4} = \frac{G_{D2}}{2} = 10^{-17} \text{ amperes cm}^3,$$

$$C_{h1} = C_{h2} = 10^{-22} \text{ coulombs cm}^3. \quad (41)$$

The diode is carrying a forward current of 10 ma when the current is reversed. The current will remain at -10 ma until the density at the junction decays to p_n. It is of interest to determine how long it takes for the density at the junction to be reduced to p_n. When this occurs the voltage on the junction reverses in accordance with (27). The model employed is shown in Fig. 13(c).

At the beginning,

$$p_{c1} = 6 \times 10^{14} \text{ cm}^{-3},$$
$$p_{e2} = 4 \times 10^{14} \text{ cm}^{-3}, \quad (42)$$

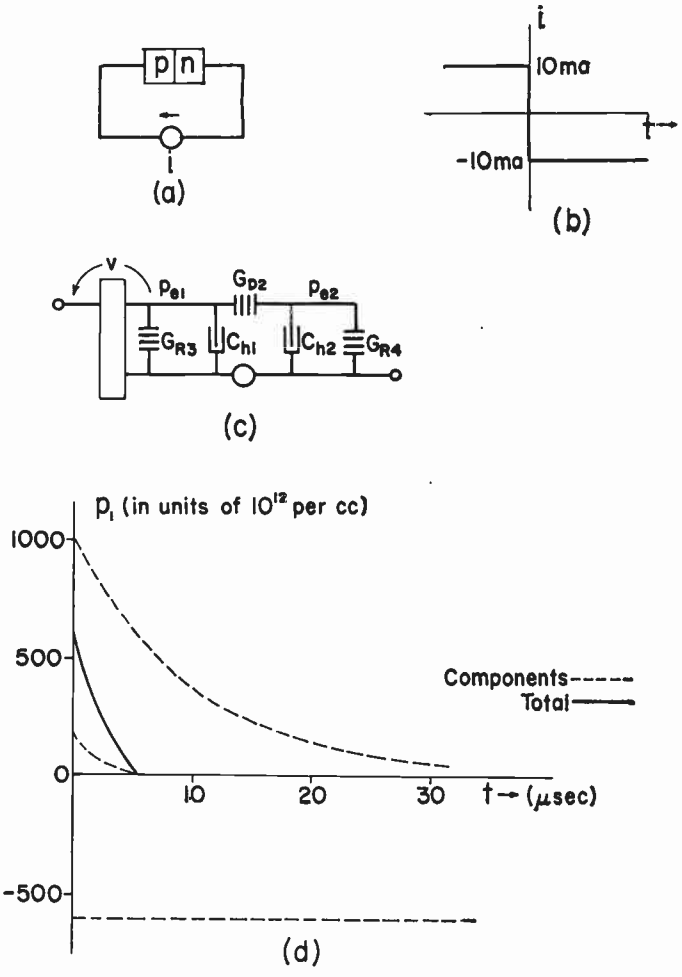


Fig. 13—Calculation of transient in a diode.

and these are the negatives of the steady values toward which the excess densities tend as the current reverses. At the outset, the junction voltage is

$$v = \frac{KT}{q} \ln \frac{p_{e1}(0) + p_n}{p_n} = 0.026 \ln 601 = 166 \text{ mv}. \quad (43)$$

A straightforward calculation reveals that $p_1(t)$ is given by

$$p_1(t) = -599 \times 10^{12} + 1000 \times 10^{12} \exp(-10^5 t) + 200 \times 10^{12} \exp(-5 \times 10^5 t). \quad (44)$$

The components and sum of p_1 are shown in Fig. 13(d). The time at which p_1 goes to p_n is seen to be 5 μsec after the reversal of current.

Photodiodes and Phototransistors

Terminal characteristics of photodiodes and phototransistors are easily determined through the use of the diffusion models just discussed. When light falls on a semiconductor, it creates hole-electron pairs. In a language appropriate to the lumped models just discussed, illumination has the influence of placing a current source from the electron line to the hole line. The placement of the current source with respect to the junction corre-

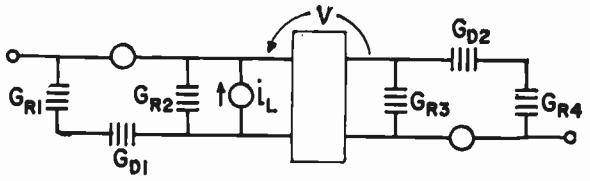


Fig. 14—Lumped model as in Fig. 12 for illumination of p material near the junction.

sponds to the location at which light falls in the physical device. Light falling near the junction in the p material on a diode as represented in Fig. 12 provides a representation as shown in Fig. 14 for the dc case. The current source increases the hole and electron densities in the n and p materials with the result that a voltage appears at the junction in correspondence with the law of the junction.

Where the diode is open circuited, the voltage is easily calculated as follows:

$$n_a = \frac{i_L}{G_{R2} + \dfrac{G_{D1}G_{R1}}{G_{D1} + G_{R1}} + \dfrac{p_n}{n_p}\left(G_{R3} + \dfrac{G_{D2}G_{R4}}{G_{D2} + G_{R4}}\right)} \tag{45}$$

$$v = \frac{KT}{q} \ln\left(\frac{n_c}{n_p} + 1\right). \tag{46}$$

Illumination on either side of the junction which falls close to the junction has the same influence. Illumination which falls at a point farther from the junction has less influence through the attenuation effect of recombination which occurs locally. This is observed through considering i_L of Fig. 14 moved to a point paralleling either G_{R1} or G_{R4}.

In connection with (45) and (46), one must remember that the laws of the junction apply only for voltages lower than the built-in voltage of the junction. Hence, increasing the illumination indefinitely does not result in indefinite increase of the junction voltage.

Eqs. (45) and (46) apply to the open-circuit case. In the short-circuit case, v remains at zero and the terminal current becomes i_L. Intermediate terminations require the solution of a circuit equation involving the nonlinear junction relationship which can be done graphically for the dc case.

The operation of the diffusion model of a phototransistor can be interpreted in a simple extension of the argument just presented. In particular, if the transistor has a heavily reverse-biased collector and a floating base, which is simple but not necessarily typical, the situation is that shown in Fig. 15. The illumination is assumed to fall near the emitter junction, but other cases can be dealt with easily.

The excess density at the collector is held substantially constant at $-p_n$ by the reverse voltage on the collector. Consequently, the ionization current from the light source must flow through G_{R1} and for this to occur the excess density changes at the emitter by the amount

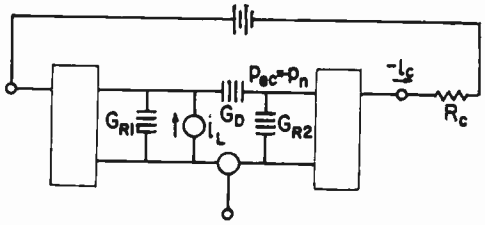


Fig. 15—Lumped model of a phototransistor.

$$\Delta p_{ec} = \frac{i_L}{G_{R1}}. \tag{47}$$

The excess density change induced at the emitter-base junction by the illumination results in a component of load current through the transistor which is proportional to i_L.

$$\Delta(-i_c) = \frac{i_L G_D}{G_{R1}}. \tag{48}$$

Use of (31) permits rewriting (48) in the following form:

$$\Delta(-i)_c = \frac{i_L \alpha_N}{1 - \alpha_N}. \tag{49}$$

The use of a biasing circuit which is ordinarily employed for the base terminal slightly reduces the current amplification of the phototransistor, but still the current amplification improvement over a photodiode is approximately by the common-emitter current gain, as is shown in (49).

Drift Transistors

In p-n-p drift transistors, the doping of the base region is heavy near the emitter and decreases toward the collector in the base region. The result is that there is a built-in field which tends to speed the flow of injected holes from the emitter to the collector. Thus there are two components of flow of holes in the base region as illustrated in Fig. 6. Eq. 16 illustrates the law of drift flow. The electrostatic potential difference between sections in the base when no biases are impressed is determined only by the doping of the base region. When biases are impressed no significant additional field is set up in the base; space-charge neutrality exists on an incremental basis. Because of this fact the drift transistor is still a linear device as regards its signals since v in (16) is not a variable.

A three-lump model of a drift transistor is shown in Fig. 16. The displacement current line in the base region is omitted as it was for the diffusion model. In the case of the drift transistor, the equilibrium density of holes in the base region is a function of the distance through the base region. Accordingly, at the nodes, one can designate the equilibrium densities as p_{n1}, p_{n2} and p_{n3}. The model is shown with incremental variables p_{c1}, p_{c2} and p_{c3}, indicated for the nodes. In terms of the discussion of p-n junctions it is clear that

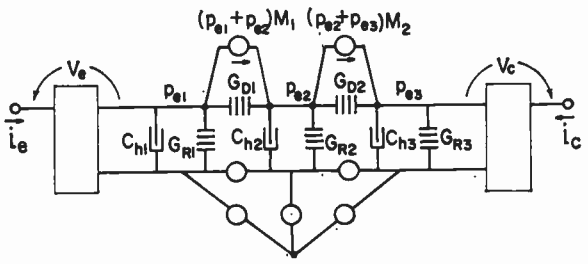


Fig. 16—Three-lump model of a drift transistor.

$$\frac{p_{n2}}{p_{n1}} = \exp \frac{q}{KT}(v_1 - v_2), \tag{50}$$

where v_1 and v_2 are the electrostatic potentials at nodes 1 and 2, respectively. A similar expression can be written for nodes 2 and 3. The *applied* voltages in the model are assumed to be entirely across the junctions.

The presence of the built-in field in drift transistors increases the alpha cutoff frequency by several times over the value in a corresponding diffusion transistor. Diffusion transistors operate with the emitter and collector in exchanged roles with moderate changes in performance. The behavior of the drift transistor is entirely different when the usual emitter acts as collector and vice-versa since the drift field is built into a device, aiding hole flow in one direction and opposing it in the other.

The following numerical example illustrates use of the lumped model of a drift transistor. The parameters are:

$$\begin{aligned} G_{D1} &= G_{D2} = 2 \times 10^{-16} \text{ amp cm}^3 \\ M_1 &= M_2 = 8 \times 10^{-16} \text{ amp cm}^3 \text{ [see (9)]} \\ G_{R1} &= G_{R2} = G_{R3} = 4 \times 10^{-18} \text{ amp cm}^3 \\ C_{h1} &= C_{h2} = C_{h3} = 10^{-24} \text{ coulombs cm}^3 \end{aligned} \tag{51}$$

In terms of these parameters, the α and the cutoff characteristic of α are determined. With the collector heavily reverse biased, p_{c3} is fixed at $-p_{n3}$.

The ratio of incremental collector to incremental emitter current (or α) can be found from the following equilibrium equations. Using complex densities and currents and abbreviating $G_{R1}+sC_{h1}$ as Y_1 and $G_{R2}+sC_{h2}$ as Y_2,

$$I_e = P_{e1}(Y_1 + G_{D1} + M_1) + P_{e2}(-G_{D1} + M_1), \tag{52}$$

$$0 = P_{e1}(-M_1 - G_{D1}) + P_{e2}(Y_2 + G_{D1} + G_{D2} - M_1 + M_2), \tag{53}$$

$$-I_c = P_{e2}(M_2 + G_{D2}). \tag{54}$$

Solving directly for $-I_c/I_e$ yields

$$\frac{-I_c}{I_e} = \frac{(M_2 + G_{D2})(M_1 + G_{D1})}{(Y_1 + M_1 + G_{D1})(G_{D1} + G_{D2} - M_1 + M_2 + Y_2) + M_1^2 - G_{D1}^2}. \tag{55}$$

Using the parameters from (51) and determining the dc α, one obtains

$$\alpha_0 = 0.994. \tag{56}$$

Substituting the parameters given in (51) leads to the following polynomial for the denominator of (55):

$$s^2 + 14.08 \times 10^8 s + 100.6 \times 10^{16}. \tag{57}$$

Factoring the polynomial reveals that the poles of α provide a cutoff characteristic with an asymptotic cutoff rate of 12db/octave for the three-lump model. Moreover, the break frequency is about 1000×10^6 rad/sec. Thus this model indicates a cutoff frequency of about 140 mc.

Avalanche Multiplication

When a *p-n* junction is heavily reverse biased, the field in the junction goes to correspondingly high values. Under these conditions, carriers flowing through have nonzero probability of ionizing hole-electron pairs on collision as indicated in (15). As a matter of fact, the ionization coefficients for holes and electrons have been determined as a function of field strength. When sufficiently high field strengths are obtained, an avalanche breakdown occurs, and additional reverse current requires insignificant increase in the reverse voltage. In a transistor avalanche, multiplication occurs as in a diode with the result that the alpha of the transistor becomes larger than unity. The variation of the multiplication is highly nonlinear and has been empirically observed by Miller[7] to be of the form

$$m = \frac{1}{1 - \left(\frac{v_c}{v_b}\right)^n} \tag{58}$$

where m is the factor by which collector current is multiplied at a fixed emitter current as a function of collector-to-base voltage v_c; v_b is the avalanche breakdown voltage and n is a constant between 3 and 5 depending upon the transistor.

To the extent that avalanche multiplication involves only steady conditions, the phenomenon is quite simple. However, in the dynamic case where space charge changes in the transistor are a part of the action, the situation is considerably more complicated. In the dynamic case, the lumped model of Fig. 17 appears helpful for both qualitative understanding and quantitative evaluation. To illustrate this point we consider, qualitatively, the development of an avalanche in the blocking oscillator circuit of Fig. 18.[8]

At the outset, we assume that there is sufficient reverse voltage on the collector so that the ionization can occur, but the alpha is still finite. The voltage across v_{c1}

[7] S. L. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, vol. 99, pp. 1234–1242; August, 1955.

[8] H. Schenkel and H. Statz, "Transistors with alpha greater than unity," PROC. IRE, vol. 44, pp. 360–371; March, 1956.

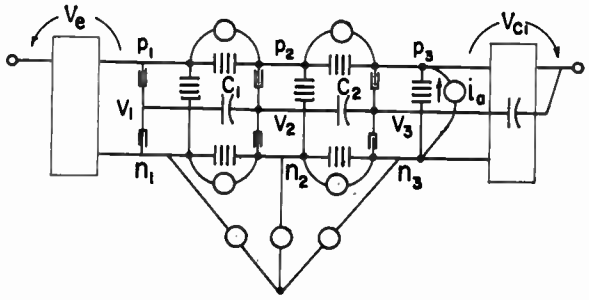


Fig. 17—Lumped model of an avalanche transistor.

is not the total collector-to-base voltage because of the depletion of part of the base evidenced by a charge on C_2 in Fig. 17. We begin consideration of the blocking oscillator at the point in the cycle where the collector voltage is increasing again following a discharge. At this point we assume the emitter junction is reverse biased. As the condenser voltage increases, the emitter begins to conduct bringing holes across the emitter-base junction. These flow toward the collector both by drift and diffusion. At the collector, in accordance with an equation

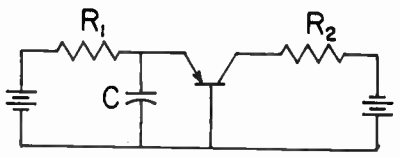


Fig. 18—Schenkel-Statz blocking oscillator.

like (15), they produce hole-electron pairs, a current in the generator i_a in dependence upon the junction current and v_{c1}. The holes produced are swept across the collector junction, the electrons by drift and diffusion move back toward the emitter junction. The model assumes unity emitter efficiency of the p-to-n regions at both emitter and collector junctions. As the electrons move back toward the emitter junction, they decrease the potential at the emitter and induce further emission at the junction there. This enhanced emission increases the extent of the avalanche, but ultimately decreases the field and ionization at the collector. Thus the discharge of C builds up and then down. After multiplication ceases, the electrons produced are disposed of through the base leads and by recombination.

Two-Dimensional Current Flow in Junction Transistors at High Frequencies*

R. L. PRITCHARD†, SENIOR MEMBER, IRE

Summary—**The effect of two-dimensional current flow in a junction transistor at high frequencies is analyzed, with particular emphasis on the rectangular geometry employed for grown-junction transistors. At high frequencies, the distributed nature of the base region must be taken into account in general. For example, in the usual equivalent circuit for the transistor, the ohmic base resistance r_b' must be replaced in general by a complex base impedance, which decreases in magnitude with increasing frequency as a result of a high-frequency internal biasing effect. The effect of this modification upon circuit performance is discussed, and transistor design considerations are outlined briefly. An approximate distributed model is used for the analysis, but it is shown in Appendix II that the solutions thus obtained are consistent with the equations governing two-dimensional current flow in a semiconductor region.**

* Original manuscript received by the IRE, March 5, 1958. Portions of this paper were presented at the 1954 IRE-AIEE Conference on Semiconductor Devices Research, Minneapolis, Minn., June 29, 1954. The work described here was performed at General Electric Res. Lab., Schenectady, N. Y.

† Texas Instruments, Inc., Dallas, Texas; formerly at General Electric Res. Lab., Schenectady, N. Y.

Introduction

MOST transistor equivalent circuits which are in common use today are based upon a one-dimensional analysis of the junction transistor.[1–3] (This is the model introduced by Shockley[1] and subsequently modified by numerous authors; *e.g.*. Early[2] and Pritchard[3] to include high-frequency effects.) Although this is a good approximation for many types of transistor, in some cases the effect of current flow in the transverse direction at high frequencies may modify transistor performance significantly, even though the

[1] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junctions," *Bell Sys. Tech. J.*, vol. 28, pp. 472–474; July, 1949.

[2] J. M. Early, "Design theory of junction transistors," *Bell Sys. Tech. J.*, vol. 32, pp. 1271–1312; November, 1953.

[3] R. L. Pritchard, "Frequency variations of junction transistor parameters," PROC. IRE, vol. 42, pp. 786–799; May, 1954.

dc current flow may be essentially one-dimensional.[4]

This paper describes an analysis of the effect of transverse current flow upon transistor parameters for the particular case of the grown-junction type of transistor geometry. A two-dimensional model is employed which effectively superimposes solutions to two one-dimensional problems. This same approach should be applicable to related transistor-analysis problems.[5] The four-terminal electrical parameters are calculated for the grown-junction triode, and it is shown that the transistor equivalent circuit in general must be modified to include a frequency-dependent complex base impedance. The effect of this modification upon transistor circuit performance is discussed and device design considerations are presented.

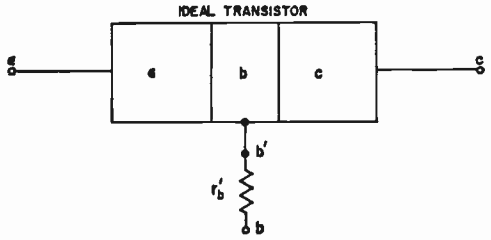


Fig. 1—Conventional model of junction transistor, showing ideal transistor and ohmic base resistance r_b'.

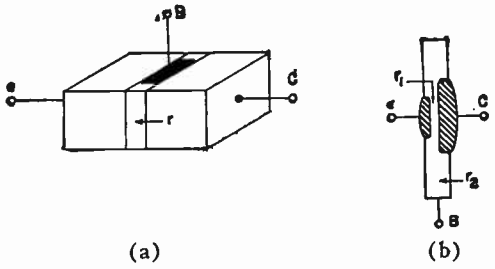


Fig. 2—Types of transistor geometry, (a) grown-junction, (b) alloy, or fused, junction, or surface-barrier transistor.

Grown-Junction Transistors

The effects of high-frequency transverse current flow were observed experimentally several years ago during the course of parameter measurements on grown-junction transistors.[6] A sort of parallel-resonance phenomenon was observed in the common-base input impedance under conditions for which the input impedance should have been inductive reactive. The capacitance required to effect the antiresonance was considerably larger than could be attributed to stray capacitance and appeared to be associated with the base resistance.

The conventional method of taking account of the nonzero resistivity of the base region is to incorporate a lumped resistance r_b' (or r_{bb}') in series with the base lead of an ideal, or intrinsic, transistor, as shown in Fig. 1.[7] However, re-examination of this approach shows that if the base resistance is large relative to the emitter resistance $r_e' = (kT/q_e I_E)$, of Shockley, *et al.*, then the distributed nature of the base resistance must be taken into account.[8] When this is done, it turns out that the junction transistor in general may be represented by the same type of model shown in Fig. 1, but with r_b' replaced by a complex, frequency dependent base impedance z_b' (see Fig. 4). The magnitude of z_b' decreases with increasing frequency, due to a reduction in the effective emitter area, which in turn results from a biasing off of portions of the emitter junction by the increased ac base current. This increase in base current at high frequencies is due to the phase shift in the forward current transfer ratio α, which also causes the phase shift associated with z_b'. Under certain conditions z_b' reduces to a simple constant resistance r_b', but in most grown-junction transistors, the base impedance is complex.

It should be emphasized that the effects described here are a result of the *geometry* of the transistor construction, rather than of its *method* of fabrication. A grown-junction transistor generally is of the form shown in Fig. 2(a), in which virtually all of the base resistance r is distributed through the active base region of the transistor. On the other hand, in an alloy, or fused-junction, transistor, or in a surface-barrier transistor, as shown in Fig. 2(b), a significant part of the effective ac base resistance arises from *nondistributed* resistance r_2 between the base contact and the active region of the base. Furthermore, the effective ac base resistance generally is considerably smaller in the latter types of transistors. For both of these reasons, the alloy-type of transistor shown in Fig. 2(b) generally can be represented quite well by the usual model of Fig. 1 comprising an ideal transistor plus a simple lumped resistance, as numerous experimental studies have indicated, for example, the extensive studies of Giacoletto.[9,10]

[4] The effect of transverse dc current upon transistor performance; *i.e.*, the internal-biasing effect, has been discussed by:
R. A. Gudmundsen, "The reduction of effective emitter voltage due to base current," paper presented at AIEE-IRE Semiconductor Device Research Conference, State College, Pa.; July 7, 1953.
N. H. Fletcher, "Some aspects of the design of power transistors," PROC. IRE, vol. 43, pp. 551–559; May, 1955. See p. 553.

[5] For example, the tetrode transistor, an analysis of which will be described by the writer in a future paper.

[6] R. L. Pritchard and W. N. Coffey, "Small-signal parameters of grown-junction transistors at high frequencies," 1954 IRE CONVENTION RECORD, pt. 3, pp. 89–98.

[7] J. M. Early, "Effects of space-charge layer widening in junction transistors," PROC. IRE, vol. 40, pp. 1401–1406; November, 1952. See p. 1404.

[8] This effect was discussed in a qualitative manner by W. Shockley, M. Sparks, and G. K. Teal, "*p-n* junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; July 1, 1951. See p. 162.

[9] L. J. Giacoletto, "Study of *p-n-p* alloy junction transistor from dc through medium frequencies," 1954 IRE CONVENTION RECORD, pt. 3, pp. 99–103; and *RCA Rev.*, vol. 15, pp. 506–562; December, 1954.

[10] L. J. Giacoletto, "Performance of a radio-frequency alloy junction transistor in different circuits," in "Transistors I," RCA Labs., Princeton, N. J., pp. 431–457; 1956.

Analysis of Distributed Model

The transistor structure to be analyzed is that of Fig. 2(a). It is assumed that the base contact is made along one complete surface of the base region,[11] thus rendering the problem two-dimensional. This two-dimensional problem then is separated into two one-dimensional problems.[12] Such a procedure is equivalent to considering the over-all transistor as comprising an infinite number of infinitesimal, ideal transistors with successive base terminals connected together by elementary resistances rdx as shown by the distributed model of Fig. 3. The validity of this procedure has been investigated by analyzing the complete two-dimensional problem for a somewhat simpler model, details of which are set forth in Appendix I, where it is shown that the separation *is* valid, at least under low-level conditions.

Under the usual linear, small-signal conditions, each ideal transistor is described by

$$di_e = (y_{ee}e_e + y_{ec}e_c)dx, \tag{1}$$

$$di_c = (y_{ce}e_e + y_{cc}e_c)dx, \tag{2}$$

where y_{ee}, y_{ec}, y_{ce}, y_{cc} denote the common-base admittance parameters *per unit length* for the ideal one-dimensional transistor,[13] including transition capacitances at emitter and collector. In general, these parameters are functions of position x. In addition, continuity of current requires that

$$di_b = di_e + di_c, \tag{3}$$

and, finally, inspection of Fig. 3 shows that

$$de_e = de_c = i_b r dx, \tag{4}$$

where r is the per-unit-length transverse resistance of the base region. By integrating (1) through (4) subject to appropriate boundary conditions, the *terminal* parameters can be evaluated for this model.

[11] In some transistors, connection to the base region is more nearly a point, rather than a line type of contact. However, this generally leads to high values of ac base resistance, and a line contact is preferred.

[12] In longitudinal direction (y direction in coordinate system of Fig. 3) only minority-carrier current is considered—the usual one-dimensional problem that has been considered in much detail (*e.g.*, Early, "Design theory of junction transistors," *op. cit.*, and Pritchard, *op. cit.*). In the transverse direction (x direction of Fig. 3) only majority-carrier, or conduction, current; *i.e.*, Ohm's law, is considered.

[13] Explicit forms of these admittances have been presented by numerous writers; *e.g.*, Early, "Design theory of junction transistors," *op. cit.*, Pritchard, *op. cit.*, and Giacoletto, "Study of *p-n-p* alloy junction transistor from dc through medium frequencies," *op. cit.*

Note also that built-in field transistors may be included within the scope of this analysis by employing appropriate expressions for $y_{ee}\ldots$; *e.g.*, the drift transistor of H. Krömer, "Zur theorie des Diffusions- und des drifttransistors," *Arch. elek. Übertragung*, vol. 8, pp. 223–228, May, pp. 363–369; August, 1954.

——, "The drift transistor," in "Transistors I," RCA Labs., Princeton, N. J., pp. 208–214; 1956.

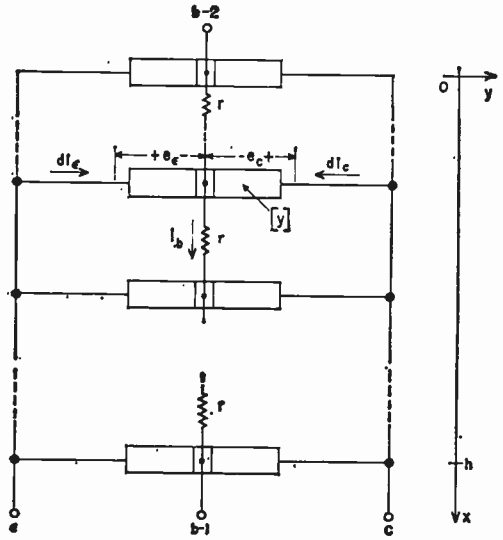


Fig. 3—Distributed model for grown-junction transistor, to take account of two-dimensional current flow in base region.

Triode Transistor

In the simple triode transistor, the admittances $y_{ee}\cdots$ will decrease with increasing distance x from the base connection because of the internal (dc) biasing effect.[4] A transverse IR drop resulting from the dc base current causes a nonuniform emitter-base voltage distribution across the base. The net effect is a sort of internal tetrode action in which the active base region tends to be crowded closer and closer to the external base contact as the dc emitter current I_E is increased. This has been verified experimentally by measurements of low-frequency ac base resistance r_b', and the results show a decrease in r_b' with increasing I_E.[14]

However, to a first approximation the internal biasing effect may be neglected, and the results obtained may be generalized by replacing the actual transverse resistance $R_b \equiv rh$ with an *effective* resistance due to internal biasing. In this case the admittance parameters become independent of position; (1) through (4) may be combined and solved subject to appropriate boundary conditions in a manner analogous to that employed for a conventional transmission line (see Appendix II). Expressions are obtained directly for emitter and collector voltages in terms of three arbitrary constants, and terminal emitter and collector currents may be evaluated by integrating (1) and (2), respectively.

For a simple triode transistor having one base connection at $x=h$, one of the constants is evaluated by requiring that the base current vanish at the other base surface $x=0$. The remaining two constants are eliminated by calculating any of the terminal small-signal parameters.

A particularly useful choice for the latter is that of the common-emitter h parameters. It turns out that

[14] See Pritchard and Coffey, *op. cit.*, p. 92 and Fig. 10, p. 98.

three of these four parameters are independent of the base resistance r and are identical with the corresponding parameters for the ideal transistor[15]

$$\left.\begin{aligned} h_{re} &= h_{12\epsilon} = (y_{12}' + y_{22}')/y_\Sigma \\ h_{oe} &= h_{22\epsilon} = (y_{11}'y_{22}' - y_{21}'y_{12}')/y_\Sigma \\ h_{fe} &= h_{21\epsilon} = -(y_{21}' + y_{22}')/y_\Sigma \end{aligned}\right\} \quad (5)$$

where y_{ij}' denotes a common-base short-circuit output admittance parameter for the one-dimensional transistor; *e.g.*, $y_{11}' \equiv y_{\epsilon\epsilon}h$, and

$$y_\Sigma \equiv (y_{11}' + y_{12}' + y_{21}' + y_{22}') \quad (6)$$

is the total ideal transistor admittance.

On the other hand, the fourth parameter, the short-circuit input impedance h_{ie} *is* a function of r:

$$h_{ie} = h_{11\epsilon} = (R_b/\Gamma h)\coth \Gamma h, \quad (7)$$

where

$$\Gamma^2 \equiv (R_b y_\Sigma)/h^2 \quad (8)$$

is a propagation constant, and

$$R_b = rh \quad (9)$$

is the transverse resistance of the base. In terms of base resistivity ρ_B and base width w, $R_b = (\rho_B/w)$ is the sheet resistance of the base region in a transistor having a square cross section, or $R_b = (\rho_B/w)(h/d)$ for a rectangular cross section, with a base depth d.

Complex Base Impedance

The impedance h_{ie} may be separated into two impedances, one of which is the short-circuit input impedance $(1/y_\Sigma)$ of the ideal transistor, while the remainder is defined as an impedance z_b'; note, however, that this is a purely arbitrary procedure:

$$z_b' \equiv h_{ie} - (1/y_\Sigma) = R_b(\coth \Gamma h/\Gamma h) - (1/y_\Sigma). \quad (10)$$

If this is done, the distributed model may be represented *exactly*, so far as *three-terminal measurements* are concerned, by the ideal theoretical model plus the impedance z_b' between the (inaccessible) base terminal b' and the external base terminal b, as shown[16] in Fig. 4.

Inasmuch as the frequency variation of the ideal transistor parameters is well known by now, only the

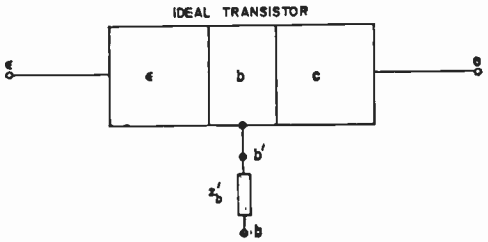


Fig. 4—Modified model of junction transistor, showing ideal transistor and complex frequency-dependent base impedance z_b'.

nature of the base impedance need be described here. This will illustrate the differences in performance which can be expected between the fused-junction and grown-junction types of transistors. To a fair approximation the total transistor admittance y_Σ is $(1-a)$ times the emitter conductance $(1/r_\epsilon')$, where a is the internal current transfer ratio, including the effects of barrier capacitances.[17] Alternatively, y_Σ is approximately equal to the admittance $y_{b'e}$ of the hybrid-pi equivalent circuit.[18] Hence, from (8), the propagation constant becomes

$$(\Gamma h)^2 \doteq (R_b/r_\epsilon')(1 - a) \approx R_b y_{b'e}. \quad (11)$$

At low frequencies,

$$(\Gamma h)^2 \doteq (R_b/r_\epsilon')(1 - a_0) \doteq (R_b/r_\epsilon')(1 - \alpha_0),$$

which is required to be small relative to 1 for negligible internal dc biasing of the emitter. Under this condition, in (10) for z_b' the hyperbolic cotangent may be replaced by the first few terms in its power-series expansion, yielding

$$z_b' = r_b' = (R_b/3), \qquad \omega = 0. \quad (10a)$$

Thus, at low frequencies the ac base impedance of the grown-junction transistor is purely resistive and is equal to one-third of the total transverse base resistance; see (9).

At high frequencies the phase shift of a (or $y_{b'e}$) causes $(\Gamma h)^2$ to become complex. This may be taken into account by relating $(\Gamma h)^2$ to the a-cutoff frequency ω_a, or by introducing the $b'-e$ diffusion capacitance $C_{b'e}$. Alternatively, especially for the newer types of built-in field transistors, it is convenient to define a cutoff frequency ω_T at which the magnitude of the common-emitter short-circuit forward current-transfer

[15] All three of these results (5) could have been anticipated. For the first two parameters, inasmuch as the transistor parameters $y_{\epsilon\epsilon}$—are uniform over the base region, the base current is zero everywhere in the base when terminal base current is zero. Hence, the resistance r does not influence h_{re} and h_{oe}. In the case of h_{fe}, the base current and collector current vary in magnitude and phase through the base, but for the case of zero collector-emitter voltage, the ratio of the two currents is exactly the same as that of an incremental transistor; *i.e.*, independent of r. On the other hand, for h_{ie}, the emitter-base voltage is fixed, and the variation in i_b does affect this parameter.

[16] Note that this model is also valid for common-base and common-collector configurations, even though it was derived for the common-emitter configuration.

[17] R. L. Pritchard, "High-frequency power gain of junction transistors," PROC. IRE, vol. 43, pp. 1075–1085; September, 1955. See p. 1083.

[18] For example, Giacoletto, "Study of p-n-p junction transistor from dc through medium frequencies," 1954 IRE CONVENTION RECORD, *op. cit.*, p. 99; and *RCA Rev.*, pp. 531–533. Note, however, that when $y_{b'e}$ is expressed in terms of base width, emitter current, etc., the result is *only* valid for the special case of a uniform-base transistor; *e.g.*, the conventional alloy transistor. More generally, $y_{b'e}$ can be expressed simply as a parallel combination of conductance $g_{b'e}$ and capacitance $C_{b'e}$.

ratio h_{fe} has decreased to 1.[19] If this is done, $(1-a)$ becomes approximately $j(\omega/\omega_T)$, and the propagation constant from (11) becomes

$$(\Gamma h)^2 \approx j\omega(R_b/\omega_T r_e'), \quad (1-\alpha_0)^2 \ll (\omega/\omega_T)^2 < 1. \quad (11a)$$

Two limiting cases can arise. If (R_b/r_e') is small, then $|\Gamma h|^2$ may be small, and the series solution employed at low frequencies also applies at high frequencies. In this case, again

$$z_b' = r_b' = (R_b/3), \quad (R_b/r_e')(\omega/\omega_T) < 1, \quad (10b)$$

is *real*, independent of frequency. On the other hand, if $(R_b/r_e')(\omega/\omega_T)$ is large, then at high frequencies $|\Gamma h|^2 \gg 1$ may obtain, and (10) yields

$$z_b' \approx [R_b r_e'/j(\omega/\omega_T)]^{1/2} - (1/y_\Sigma), \quad (\omega R_b/\omega_T r_e') > 2. \quad (10c)$$

For very large values of the dimensionless frequency variable $(\omega R_b/\omega_T r_e')$, the base impedance z_b' comprises equal resistive and reactive components and decreases in magnitude as the square root of increasing frequency. Hence, a broken-line approximation could be constructed from this $\omega^{-\frac{1}{2}}$ variation and a frequency-independent low-frequency value $(R_b/3)$ to describe z_b' approximately over the entire frequency range. A more accurate graphical description of z_b' is provided by the solid curves of Fig. 5, which show the phase and normalized magnitude of z_b' as a function of $(\omega R_b/\omega_T r_e')$. The high-frequency approximation of (10c) is shown by the dotted curves, while the dashed line for the phase at low frequencies is an approximation obtained by adding a capacitance $(1/5\omega_T r_e')$ in parallel with $(R_b/3)$.[20]

The capacitive reactance associated with the base impedance at high frequencies is sufficient to account for the observed antiresonance in the common-base-short-circuit input impedance.[21] Furthermore, measured terminal parameters for a number of grown-junction transistors have been shown[22] to be in good qualitative agreement with the corresponding parameters calculated from the model of Fig. 4 using (10c) for the high-frequency base impedance z_b'. (See also Fig. 7.)

Transverse AC Voltage Distribution

The physical explanation for the decrease in the magnitude of the base impedance at high frequencies is simply that less of the base region is utilized owing to the by-passing effect of the distributed base resistance by the distributed transistor capacitance $(1/\omega_T r_e')$, or

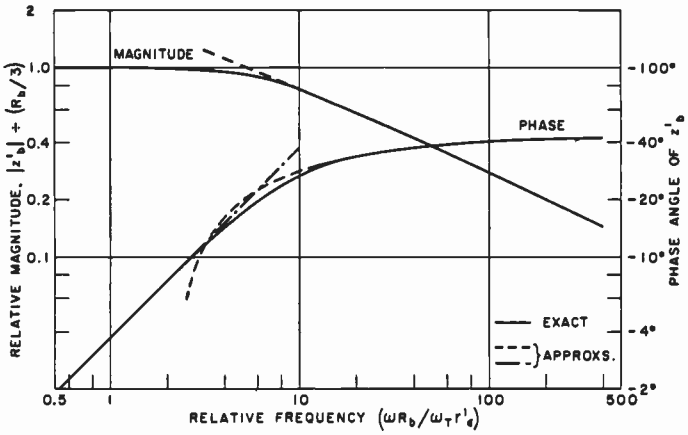


Fig. 5—Variation of complex base impedance z_b', normalized with respect to low-frequency value, as a function of dimensionless frequency variable $(\omega R_b/\omega_T r_e')$.

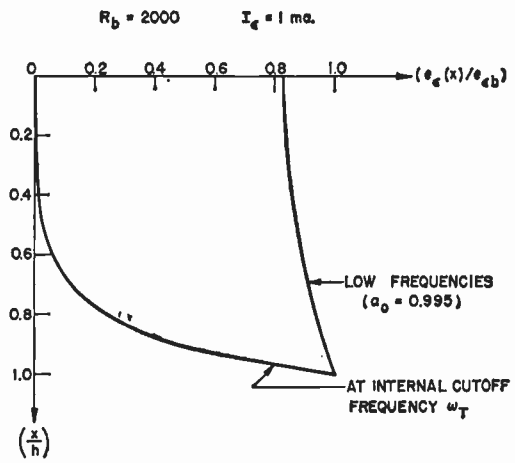


Fig. 6—Transverse variation of emitter-base voltage in theoretical model of distributed transistor, showing high-frequency internal biasing effect.

$C_{b'e}$. This is illustrated in Fig. 6, which shows the calculated normalized emitter-base voltage $e_e(x)/e_{eb}$ for $e_{cb}=0$, as a function of transverse distance x through a distributed-model transistor with $R_b=2500$ ohms, $r_e'=25$ ohms ($I_e=1$ ma), $\alpha_0=0.995$, and with the reasonable assumption[23] that $|y_{12}'+y_{22}'| \ll |y_\Sigma|$. At low frequencies the ac emitter voltage is essentially constant through the base region. However, at a frequency equal to the internal cutoff frequency ω_T, $(\Gamma h)^2 \approx j100$ is not small, and the ac emitter voltage decreases rapidly with increasing distance from the base connection. For example, $|e_e|$ decreases by 50 per cent in the first 10 per cent of the distance across the base, and the phase shift (not shown in Fig. 6) decreases 40 degrees. The effective ac base impedance z_b' in this case has decreased from the low-frequency value of $r_b'=800$ ohms (resistive) to approximately 230 ohms in magnitude with a phase shift of approximately -40 degrees at ω_T.

Two-Base Triode

If a second base connection is made to the upper edge

[19] See Acknowledgments concerning ω_T. This cut-off frequency thus is a measure of the phase shift associated with the inherent current-transfer ratio rather than of its amplitude-frequency characteristic. For a conventional uniform-base transistor, $\omega_T=0.83\omega_a$, where ω_a is the inherent alpha-cutoff frequency, but for the newer high-frequency transistors, ω_T may be significantly less than ω_a. (See, for example, experimental data for a diffused-base transistor; C. A. Lee, "A high-frequency diffused base germanium transistor," *Bell Sys. Tech. J.*, vol. 35, Figs. 1 and 2, pp. 26–27; January, 1956.

[20] Addition of this capacitance also provides a slightly better approximation for $|z_{\nu'}|$ in the region of $(\omega R_b/\omega_T r_e') \doteq 2-6$.

[21] In fact, under simplified conditions, this model predicts an antiresonance at a frequency somewhat in excess of ω_T.

[22] Pritchard and Coffey, *op. cit.*, Fig. 3 through Fig. 7, p. 97.

[23] In which case $e_e(x \doteq e_{eb}[\cosh \Gamma x/\cosh \Gamma h]$.

of the base region, and if the two base contacts are connected together, the analysis may be repeated with different boundary conditions. When this is done, the same type of results are obtained, but with R_b replaced by $R_b/4$. The factor 4 arises from the fact that effectively the original transistor has been cut in half longitudinally, and the two halves now are in parallel. Thus, for example, the low-frequency value of the base impedance becomes

$$z_b' \equiv r_b' = (R_b/12). \tag{12}$$

Circular Cross Section

If a ring type of base contact is made to the base region (assumed now to be circular in cross section of radius a), equations somewhat like (1) through (5) are obtained, but for a cylindrical geometry.[24] In this case the result for the base current may be expressed in terms of the Bessel function of imaginary argument $I_1(\Gamma a) = -jJ_1(j\Gamma a)$, where $(\Gamma a)^2 = (Ry_\Sigma/\pi)$, and R in this case is the sheet resistance ρ_B/w of the base region.[25] As in the case of the rectangular geometry, these results may be manipulated to yield the model shown in Fig. 4, with

$$z_b' \equiv \left[\left(\frac{R}{2\pi}\right)\left(\frac{1}{\Gamma a}\right)\frac{I_0(\Gamma a)}{I_1(\Gamma a)}\right] - \frac{1}{y_\Sigma}. \tag{13}$$

For small values of $(\Gamma a)^2$, as at low frequencies, or at high frequencies when $(\omega R/\pi\omega_T r_e')^2 \ll 1$, power-series expansion yields

$$z_b' = r_b' = (R/8\pi), \tag{14}$$

which is the result given previously by Early.[26] On the other hand, for large values of $(\Gamma a)^2$, as *may* occur at high frequencies, with a large value of R, z_b' is complex and is approximately equal to

$$z_b' \approx [(Rr_e')/4\pi j(\omega/\omega_T)]^{1/2} - (1/y_\Sigma). \tag{15}$$

More generally, z_b' for the circular geometry can be described very closely by the curves of Fig. 5, but with the resistance R_b of that analysis replaced by $(0.6\ R/\pi)$. In other words, for a given departure of z_b' from its low-frequency value, for a transistor having a circular cross section and ring base contact, and a sheet resistance R, the frequency must be 5 times as large as in the case of a unit having the same sheet resistance but with a square cross section and a line contact.

Circuit Considerations

The effective decrease in base resistance which occurs at high frequencies in the grown-junction type of transistor actually improves its over-all circuit performance. This has been demonstrated by calculations of maximum available high-frequency power gain,[27] which show that the gain decreases less rapidly with frequency (15 db/decade) for the grown-junction transistor than for the fused-junction transistor (20 db/decade). Experimental results have been presented for a large number of grown-junction transistors to substantiate these calculations.[27]

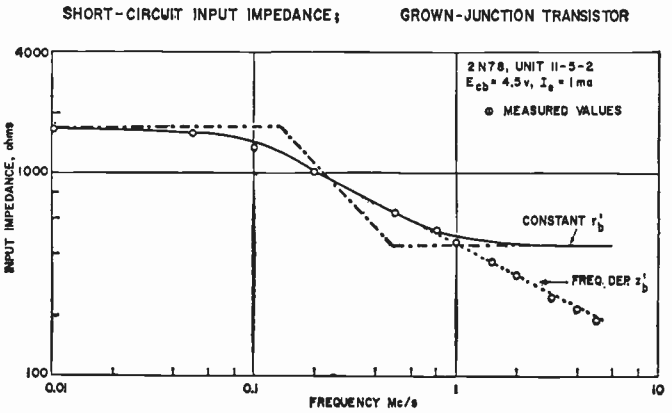


Fig. 7—Variation of magnitude of common-emitter short-circuit input impedance h_{ie} with frequency for typical grown-junction transistor, showing effect of decreasing base impedance at high frequencies.

On the other hand, because of the complex nature of the base impedance, it may be more difficult in some cases to provide satisfactory terminations for the transistor. This probably is of no concern for a tuned narrow-band amplifier, but for a wideband amplifier some difficulty might be experienced in the design of interstage coupling networks. For example, the frequency variation of the input impedance of a common-emitter amplifier stage having a small load resistance, as shown by the solid curve in Fig. 7, can be approximated by a simple lumped network of resistances and capacitances (indicated by the broken-line approximation). On the other hand, for a grown-junction transistor having a relatively high value of base resistance, the frequency variation of the base impedance causes the input impedance at higher frequencies to vary as the square root of the frequency, as shown by the dotted curve in Fig. 7. The points represent measured values for a typical grown-junction unit. Such a variation can *not* be approximated by a simple RC network.

Transistor Design Condensation

For good high-frequency performance it is essential that the low-frequency base resistance r_b' be small, even though the effective high-frequency base resistance may decrease with increasing frequency for a grown-junction transistor. This can be accomplished in transistor design by striving either for a ring type of contact (although this may be difficult to obtain in practice) which would yield a value of $r_b' \doteq R/8\pi$, or for a two-

[24] The calculation for the cylindrical geometry was first carried out by Early, as noted during the discussion following the presentation of Pritchard's paper at the 1954 IRE-AIEE Conference on Semiconductor Devices Research, Minneapolis, Minn., June 29, 1954.

[25] Equivalent to the transverse base resistance R_b of a transistor of square cross section; see text following (9).

[26] Early, "Design theory of junction transistors," *op. cit.*, p. 1311.

[27] Pritchard, "High-frequency power gain of junction transistors," *op. cit.*, pp. 1078–1080.

base connection yielding $r_b' \doteq R/12$ for a square cross section, or $(R/12)\cdot(h/d)$ in general. Either of these cases is considerably better than the single-line contact for which $r_b' = R_b/3$, but even the latter may be several times better than a point contact to the base region.

Furthermore, if r_b' can be made sufficiently small, the base impedance at high frequencies will be frequency *independent*, and any circuit difficulties associated with frequency variation of z_b' would be eliminated. Moreover, as transistor operation is pushed to higher frequencies, it becomes essential to keep r_b' as low as possible. For example, in order to achieve even 3 db of power gain at a frequency of 500 mc with a transistor having an inherent cutoff frequency of this same order of magnitude and a collector capacitance of C_c of 1 $\mu\mu f$, the *total* ac base resistance must not exceed 40 ohms,[28] in which case the base impedance would be essentially independent of frequency up to several hundred megacycles.

Appendix I

Approximate Calculation of Two-Dimensional Current Flow in Junction-Transistor Base Region

Assuming negligible recombination, equations describing current flow in a semiconductor are well-known:[29]

Continuity equation for holes

$$(\partial p/\partial t) = -(1/q_e)(\nabla \cdot J_p) \tag{16}$$

where J_p is the hole current density, defined as

$$J_p = -q_e(D_p \nabla p - \mu_p p \mathbf{E}). \tag{17}$$

Continuity equation for electrons

$$(\partial n/\partial t) = (1/q_e)(\nabla \cdot J_n) \tag{18}$$

where J_n is the electron-current density, defined as

$$J_n = +q_e(D_n \nabla n + \mu_n n \mathbf{E}). \tag{19}$$

In these equations, p and n denote hole and electron density, respectively, $\mathbf{E}$ is the electric field, q_e the electronic charge, and D, and $\mu = (q_e D/kT)$ denote the diffusion constant and mobility, respectively, for holes (subscript p) or for electrons (subscript n). The total current density J is

$$J = J_p + J_n. \tag{20}$$

In addition, Poisson's equation requires that

$$(\nabla \cdot \mathbf{E}) = (q_e/\epsilon_0 \epsilon_r)(p - n + N_d - N_a), \tag{21}$$

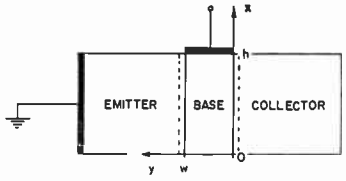


Fig. 8—Coordinate system for two-dimensional current flow in base region of transistor.

where ϵ_0 is the permittivity of free space (8.85×10^{-12} farads per meter), ϵ_r is the relative dielectric constant, and N_d, N_a are the concentrations of donor and acceptor impurities, respectively.

A coordinate system as shown in Fig. 8 is employed with $x = h$ as an equipotential surface.

Subscripts x and y denote respectively, x and y components of vector quantities, and each variable also includes a subscript 0 or 1 to denote dc or ac component, respectively.

For convenience, a p-n-p transistor is considered, although the results for an n-p-n structure are quite similar. In addition, low-level dc operation[30] and small-signal ac operation are assumed, while the transition capacitances and depletion-layer widening effects are neglected, and the emitter efficiency is assumed to equal one.

It follows that the dc base current is zero, and hence dc current flow is one dimensional. Subject to the assumptions noted above, approximate dc solutions to (16) through (19) are[31]

$$p_0(x, y) \doteq p_{B\epsilon}(y/w), \tag{22}$$

where

$$p_{B\epsilon} \equiv p_{B0}\epsilon^{(qV_{EB}/kT)} \tag{23}$$

is the injected hole concentration at the emitter edge of the base region and p_{B0} is the equilibrium hole concentration in the base, while

$$n_0(x, y) \doteq N_d + p_0(x, y), \tag{24}$$

$$\mathrm{E}_{x0} = 0 \tag{25}$$

and

$$\mathrm{E}_{y0} \doteq -(kT/q_e w)(p_{B\epsilon}/N_d). \tag{26}$$

For the ac solution to the two-dimensional problem, a perturbation method is employed. The solutions to the two one-dimensional problems described in the text are used to calculate the currents that are neglected in this separation process, and it is shown that the neglected quantities indeed are negligible.

[28] This conclusion is based on the now well-known equation relating transistor gain and the internal parameters r_b', C_c, and ω_T; see, for example, Pritchard, "High-frequency power gain of junction transistors," *op. cit.*, (8) p. 1078, and footnote 2 of that paper for other references.

[29] For example, W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y., pp. 296–306; 1950.

[30] It should be emphasized that although the success of this analysis depends completely upon neglecting high-level terms, *i.e.*, terms of order (p/N) relative to 1, it is *not* necessarily true that the superposition of two one-dimensional solutions is *not* valid at high level. Unfortunately, elimination of the low-level assumption leads to a difficult problem which has yet to be solved.

[31] W. M. Webster, "On the variation of junction-transistor current-amplification factor with emitter current," Proc. IRE, vol. 42, pp. 914–920; June, 1954.

The solution given in the text must be rewritten here in terms of hole concentration and electric field. This can be done easily by referring back to the one-dimensional solution for the ideal transistor, upon which (1) and (2) are based,[13] and by introducing the simplifying assumptions noted above. In this case,

$$p_1(x, y) = -\left(\frac{q_e p_{B\epsilon}}{kT}\right)\left(\frac{\sinh \Gamma_1 y}{\sinh \Gamma_1 w}\right) v_1(x), \tag{27}$$

with

$$v_1(x) = -e_{\epsilon b}\left(\frac{\cosh \Gamma x}{\cosh \Gamma h}\right) \tag{28}$$

and

$$E_{x1} = -(\partial v_1/\partial x)$$

and where

$$\Gamma_1^2 \equiv j\omega/D_p \tag{29}$$

and

$$\Gamma^2 = \frac{j\omega}{D_n}\left(\frac{p_{B\epsilon}}{N_d}\right)\left[\frac{\cosh \Gamma_1 w - 1}{(\Gamma_1 w)\sinh \Gamma_1 w}\right]. \tag{30}$$

As is demonstrated below, space-charge neutrality exists in the base so that

$$n_1(x, y) = p_1(x, y). \tag{31}$$

The transverse ac electron current density J_{nx1} may be calculated from (19) by employing (31), (27), (24), (22), and (28):

$$J_{nx1} = q_e \mu_n N_d \left\{1 + \left(\frac{p_{B\epsilon}}{N_d}\right)\left[\left(\frac{y}{w}\right) + \frac{\sinh \Gamma_1 y}{\sinh \Gamma_1 w}\right]\right\} E_{x1}. \tag{32}$$

By substituting this result in the continuity equation for electrons (18), together with (31), and (27), the y derivative of the longitudinal ac electron current density $(\partial J_{ny1}/\partial y)$ may be obtained. Integration of this result with respect to y and requiring that the electron current vanish at both the collector and emitter junctions $y=0$ and $y=w$ yields

$$J_{ny1} = \left[\frac{q_e \mu_n p_{B\epsilon} w}{(\Gamma_1 w)\sinh \Gamma_1 w}\right]\left(\frac{j\omega}{D_n}\right)\left[\left(\frac{y}{w}\right)(\cosh \Gamma_1 w - 1) - (\cosh \Gamma_1 y - 1)\right] v_1(x). \tag{33}$$

Note that J_{ny1} is not required to, and in fact does not, vanish *everywhere* in the base region.[32]

The variation of J_{ny1} through the base region may be displayed in an approximate manner for not-too-high frequencies by expanding the hyperbolic functions in the brackets of (33) and retaining only the first two terms

$$J_{ny1} \propto (\Gamma_1 w)^2[(y/w) - (y/w)^2].$$

Thus, the longitudinal electron current reaches a maximum halfway between emitter and collector, although it is zero at both emitter and collector junctions.

The y component of ac electric field E_{y1} is determined from (19), after substituting (33), (24), (31), (27), (22), and (26):

$$E_{y1}(x, y) \doteq \left(\frac{p_{B\epsilon}}{N_d}\right)\frac{\Gamma_1}{\sinh \Gamma_1 w}\left\{\cosh \Gamma_1 y - \left(\frac{\mu_p}{\mu_n}\right)\cdot\left[(\cosh \Gamma_1 y - 1) - \left(\frac{y}{w}\right)(\cosh \Gamma_1 w - 1)\right]\right\} v_1(x). \tag{34}$$

To verify (31), this result (34), together with (28), may be substituted in Poisson's equation (21) to yield

$$(p_1 - n_1) \doteq \left(\frac{\epsilon_0 \epsilon_r}{q_e}\right)\left(\frac{p_{B\epsilon}}{N_d}\right)\left(\frac{\sinh \Gamma_1 y}{\sinh \Gamma_1 w}\right)\Gamma_1^2\left[1 - \frac{\mu_p}{\mu_n}\right] v_1(x),$$

which also may be written

$$\left(\frac{p_1 - n_1}{p_1}\right) = (1 - \mu_n/\mu_p)(\epsilon_0 \epsilon_r/q_e \mu_n N_d) j\omega. \tag{35}$$

The first term on the right-hand side of (35) is of the order of magnitude of 1–2, whereas the second term is the relaxation time of the semiconductor comprising the base region; *e.g.*, approximately 10^{-12} seconds for 1 ohm-cm germanium. Hence, for all reasonable frequencies $|p_1 - n_1| \ll |p_1|$, and ac space-charge neutrality indeed *is* preserved in the base region.[33]

The transverse ac hole current density J_{px1} may be calculated from (17) after substituting from (27), (22), and (28)

$$J_{px1} = +q_e \mu_p p_{B\epsilon}\left[\left(\frac{y}{w}\right) - \frac{\sinh \Gamma_1 y}{\sinh \Gamma_1 w}\right] E_{x1}. \tag{36}$$

Note that at low frequencies, the drift and diffusion components of J_{px1} (first and second terms, respectively) are equal in magnitude but opposite in direction.

Similarly, the longitudinal ac hole current density J_{py1} can be calculated from (17) with the help of (27), (22), (34), and (26),

$$J_{py1} = q_e D_p\left(\frac{\Gamma_1 q_e p_{B\epsilon}}{kT}\right)\left(\frac{\cosh \Gamma_1 y}{\sinh \Gamma_1 w}\right)\left\{1 + \left[\frac{p_{B\epsilon}}{N_d}\right]\cdot\left[\left(\frac{y}{w}\right) + \frac{\tanh \Gamma_1 y}{\Gamma_1 w}\right]\right\} v_1(x). \tag{37}$$

The only unused equation now is the continuity equation for holes (16). If (27), (36), and (37) are substituted in (16) and if terms of order $(p_{B\epsilon}/N_d)$ are neg-

[32] This is in contrast to the analysis of Rittner, who neglected the electric field and the electron currents in a three-dimensional theory with small injection level. See E. S. Rittner, "Extension of the theory of the junction transistor," *Phys. Rev.*, vol. 94, p. 1168; June 1, 1954.

[33] Note also that (35) indicates that n_1 and p_1 have the same x and y dependence; hence, the substitution of derivatives of p_1 for those of n_1 used in some of the above equations also is valid.

lected,[34] then this equation merely verifies the well known result $\Gamma_1^2 = j\omega/D_p$.

The total ac current densities J_{x1} and J_{y1} now can be calculated from (20) and the preceding results. Thus, the total transverse ac current density J_{x1} from (32) and (36), neglecting terms of order (p_{Be}/N_d) is

$$J_{x1} \equiv (J_{nx1} + J_{px1}) \doteq (q_e\mu_n N_d)\mathrm{E}_{x1} \tag{38}$$

which is just the equation for the transverse drift, or conduction, current in the base; *i.e.*, the effects of minority carriers and of diffusion are negligible.

The total longitudinal ac current density J_{y1} from (37) and (33), neglecting terms of order (p_{Be}/N_d) is

$$J_{y1} \equiv (J_{py1} + J_{ny1}) \doteq \frac{\mu_p q_e p_{Be}\Gamma_1}{\sinh \Gamma_1 w}\left\{\cosh \Gamma_1 y - \left[(\cosh \Gamma_1 y - 1) - \left(\frac{y}{w}\right)(\cosh \Gamma_1 w - 1)\right]\right\} v_1(x). \tag{39}$$

The first term in the braces arises from the diffusion of minority carriers through the base region, whereas the second term may be of comparable magnitude and arises from the drift of majority carriers longitudinally through the base. This longitudinal drift current is necessary in order to maintain continuity of electrons in the base in spite of the substantial *transverse* drift current. Hence, the longitudinal distribution of current density $J_{y1}(y)$ in this two-dimensional model is *not* the same as in the usual one-dimensional case, *even* under low-level conditions.[35]

However, since the electron current was required to vanish at emitter and collector, the *terminal* currents $J_{y1}(w)$ and $J_{y1}(0)$ under low-level conditions[30] *are* identical with those of a corresponding one-dimensional transistor.

[34] If terms of order $((p_{Be}/N_d)$ are *not* excluded, (16) is *not* satisfied for all values of y, and it would be necessary to introduce a correction term, also of order (p_{Be}/N_d), in the expression for $p_1(x, y)$ (which also modifies slightly the result for J_{py1}) to obtain approximate satisfaction of (16).

[35] To the first approximation, for not-too-high frequencies, the variation of ac current density J_{y1} from collector to emitter is linear, rather than parabolic as would be the case if electron current were neglected.

Appendix II

Solution of Distributed-Model Transistor Equations

Combination of (1) through (4) yields a second-order differential equation:

$$(d^2 i_b/dx^2) = \Gamma^2 i_b, \tag{40}$$

where Γ^2 has been defined by (8) and (6). The solution to (40) is well known

$$i_b(x) = A \sinh \Gamma x + B \cosh \Gamma x \tag{41}$$

where A and B are arbitrary constants.

Substituting (41) in (4) and integrating yields

$$e_e(x) = (R_b/\Gamma h)[A \cosh \Gamma x + B \sinh \Gamma x] + E, \tag{42}$$

where E is a third arbitrary constant.

As a consequence of (4), $e_c(x)$ is related to $e_e(x)$ of (42) by a constant, which can be evaluated by combining (1) through (3) with (42) and comparing the result with (41) after differentiation. It then follows that

$$e_c(x) = e_e(x) - [y_2/(y_{12}' + y_{22}')]E. \tag{43}$$

For a simple triode transistor $i_b(0) = 0$, and $B = 0$ must obtain. The remaining two constants A and E can be evaluated by calculating terminal small-signal parameters in terms of terminal voltages, *e.g.*, $e_{be} = e_e(h)$, and terminal currents; *e.g.*,

$$i_e = \int_0^h (di_e/dx)dx.$$

Acknowledgement

The assistance of E. Teter in carrying out numerical calculations for Fig. 5 and of J. Lawrence in obtaining the experimental data for Fig. 7 is gratefully acknowledged. The concept of the internal cutoff frequency ω_T introduced above is the result of informal discussions on IRE-AIEE Task Group 28.4.7 on Transistor Internal Parameters with the author's colleagues, R. B. Adler, J. B. Angell, J. M. Early, and W. M. Webster. Also, J. M. Early offered a number of helpful suggestions in reviewing the manuscript, for which the author is very grateful.

Construction and Electrical Properties of a Germanium Alloy-Diffused Transistor*

P. J. W. JOCHEMS†, O. W. MEMELINK†, AND L. J. TUMMERS†

Summary—The fabrication of high-frequency transistors by the alloy-diffusion method is described. A group of transistors was subjected to an extensive series of measurements in order to establish an equivalent circuit characterizing the transistor for small ac amplitudes at a fixed dc bias. The resulting "physical" T-equivalent circuit is valid up to at least 25 mc. The different circuit elements are discussed with respect to their physical background. A translation from the T-circuit into the electrically more convenient π-circuit is also presented.

Introduction

THE manufacture of germanium transistors with an α cutoff frequency of the order of 100 mc or more requires techniques for making base layers of a thickness of a few microns in a controllable way. It is generally agreed that the alloying technique[1] lacks this necessary degree of controllability, due to the spread in penetration depth, which may be a few microns itself.

These thin base layers can be obtained, however, by making use of diffusion of an impurity into germanium.[2] The advantage of the diffusion technique lies in the fact that the diffusion is a rather slow process, which can be controlled accurately. By a proper choice of the temperature and the time of diffusion, a thin base layer can be realized. However a remaining problem, now, is how to make a good emitter contact to this layer, on which only a very shallow depth of alloying can be permitted.

A solution for this problem is provided by combining alloying and diffusion in one step. This method has been described by Longini[3] in connection with the construction of hook transistors. Application of this technique for making high-frequency transistors has been reported by Beale.[4] Early work on the alloy-diffusion technique in our laboratory was carried out by Ploos van Amstel. As the geometry of this transistor is different from that of an alloy transistor, it can be expected that the equivalent circuit describing its electrical properties will also be different.

In the following we first shall describe briefly the process we used for making an alloy-diffused transistor with an α cutoff frequency above 100 mc. After that a "physical" equivalent circuit will be given, which is to some extent related to the structure of the transistor. Finally a π-equivalent circuit will be presented from which some important circuit properties of the transistor can be derived. The equivalent circuits are valid from 0.1 to at least 25 mc, a frequency range for which there will be many applications for this transistor.

The Fabrication of the Alloy-Diffused Transistor

A tiny pellet of bismuth containing arsenic and aluminum is alloyed into a p-type germanium wafer at elevated temperatures. The system of the solid and liquid phase reaches an equilibrium corresponding to a certain penetration depth of the liquid Ge-Bi alloy. Keeping the wafer at a constant temperature over a certain period of time, arsenic will diffuse from the alloy into the solid material making the germanium n type over a distance determined by temperature and time. When the wafer is cooled a low-ohmic p-type germanium layer recrystallizes from the liquid alloy if sufficient aluminum is present to overcompensate the remaining arsenic. Fig. 1 gives a cross section of the wafer with the emitter pellet.

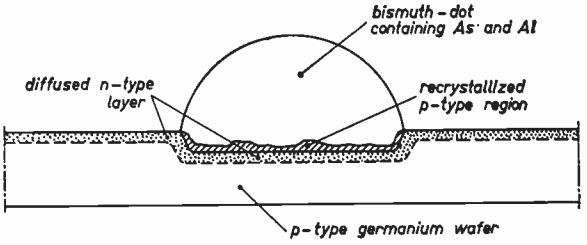


Fig. 1—Cross section of the alloyed emitter pellet together with the diffused base layer.

The alloy-diffusion process is carried out in an atmosphere containing arsenic vapor originating by evaporation from the emitter pellet itself or from a separate arsenic source. The result is the formation of an n-type layer all over the surface of the wafer and not only underneath the emitter pellet. To provide a base contact, an ohmic electrical connection has to be made on this layer, which is in effect an extension of the base region. The choice of the dimensions and the shape of the base contact is governed primarily by the desired values of collector to base capacitance and base resistance. A ring shape, for instance, gives a low base resistance, but a relatively high collector to base capaci-

* Original manuscript received by the IRE, March 17, 1958.
† Philips Res. Labs., N.V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands.
[1] R. R. Law, C. W. Mueller, J. I. Pankove, and L. D. Armstrong, "A developmental germanium p-n-p junction transistor," PROC. IRE, vol. 40, pp. 1352–1357; November, 1952.
[2] C. A. Lee, "A high-frequency diffused base germanium transistor," *Bell Sys. Tech. J.*, vol. 35, pp. 23–34; January, 1956.
[3] Longini, British Patent no. 754404.
[4] J. R. A. Beale, "Alloy-diffusion: a process for making diffused-base junction transistors," *Proc. Phys. Soc.* (London), vol. B 70, pp. 1087–1089; November, 1957.

tance, due to the large collector junction area. In order to minimize the collector to base capacitance, the base contact should be small. We have chosen an arsenic-containing bismuth pellet of about the same dimensions as the emitter pellet. This base contact is alloyed into the germanium surface within a short distance of the emitter. The ohmic collector contact is alloyed into the opposite side of the wafer. Before etching, the surface between emitter pellet and base pellet is masked, so that during etching all germanium around the pellets is dissolved except for a small "bridge" between emitter and base pellet.

Some electrical parameters of this transistor, measured at $I_e = 1$ ma; $V_{cb} = -6$ v; $T = 20°C$ are:

α cutoff frequency f_c: 100–200 mc,
collector to base capacitance C_c: 1–2 $\mu\mu$f,
"feedback" base resistance r_b': 60 ohms,
short circuit emitter to collector current gain α_f:

$$1 - \alpha_f < 0.01,$$

forward transfer admittance at

$$10 \text{ mc}, \ |y_f| : 35 \text{ mmho},$$

emitter breakdown voltage $V_{cb\,max}$: 2–4 v,
collector breakdown voltage $V_{cb\,max}$: -60 v.

The "Physical" Equivalent Circuit

The electrical behavior of the transistor for small signal operation at a certain fixed dc bias can be characterized by an equivalent circuit. Examining the structure of the transistor, Fig. 2, one could, in principle, develop a complete equivalent circuit founded on a physical basis. Some of the factors which would have to be accounted for are:

1) A drift field[5] in the base owing to the graded impurity distribution.
2) A comparatively high emitter depletion-layer capacitance due to the low resistivity of the base region.
3) The distributed nature of transistor parameters due to the three-dimensional structure.
4) A collector series resistance.

The resulting three-dimensional equivalent circuit would be quite involved. Therefore, rather than develop such an equivalent circuit, we chose to derive an equivalent circuit on a basis of experimental analysis.

By comparing the results of a wide variety of ac measurements with theoretical considerations we see that the electrical behavior of the transistor between 0.1 and 25 mc (dc bias: $I_e = 1$ ma; $V_{cb} = -6$ v) is adequately described by the T-equivalent circuit of Fig. 3. This T-circuit can be considered as a "physical" equivalent circuit as its elements are closely related to the physical action of the corresponding parts in the real transistor.

[5] H. Krömer, "Der Drifttransistor," Naturwiss., vol. 40, pp. 578–579; December, 1953.

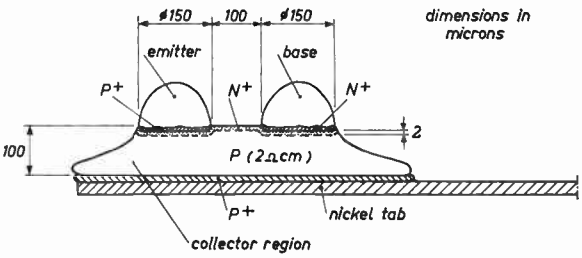


Fig. 2—Cross section of the alloy-diffused transistor.

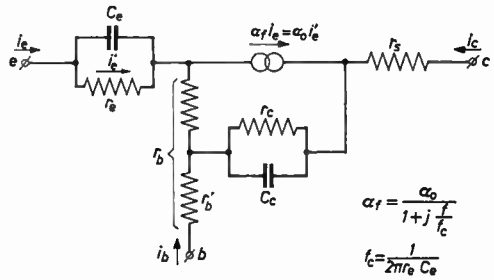


Fig. 3—The "physical" equivalent circuit [0.1–25 mc; $I_e = 1$ ma; $V_{cb} = -6$ v].

In order to fully understand the electrical behavior of this transistor, a detailed discussion of the various elements would seem appropriate. Before discussing the qualitative aspects of the circuit elements, we present Table I for displaying some quantitative information about these elements.

TABLE I
Measured Values of the Elements of the Equivalent Circuit in Fig. 3 for Three Experimental Transistors

Transistor	($I_e = 1$ ma; $V_{cb} = -6$v; $T = 20°C$; Frequency: 0.1 – 25 mc.)						
	r_e ohms	C_e $\mu\mu$f	$1-\alpha_0$	r_b' ohms	r_b ohms	C_c $\mu\mu$f	r_c megohms
Z86–6	26	40	<0.01	61	170	1.2	>1
Z86–19	26	30	<0.01	62	180	1.4	>1
Z86–21	26	33	<0.01	54	170	1.3	>1

The magnitude of the emitter junction differential resistance r_e corresponds closely to the theoretical expression $r_e = kT/qI_e$.

C_e, the capacitance parallel to r_e represents the sum of the "diffusion" or "storage" capacitance—associated with the diffusion of minority carriers into the base region—and the emitter depletion-layer capacitance. As the "diffusion" capacitance is proportional to I_e, and the emitter depletion-layer capacitance is almost independent of I_e—except at very low currents—it is possible to separate these two quantities. From the measurement of C_e vs I_e, Fig. 4, it follows that the emitter depletion-layer capacitance contributes a considerable amount to the total capacitance C_e. This is in contrast to the conventional alloy transistor in which the

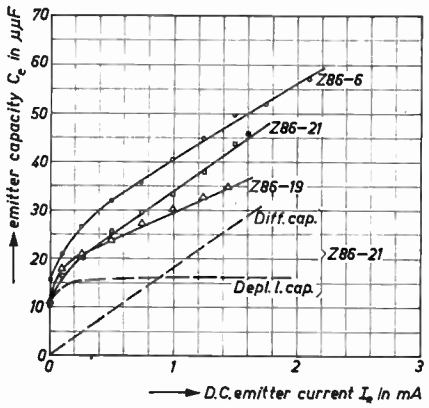


Fig. 4—The emitter capacitance as a function of emitter current [100 kc; $V_{cb} = -6$ v].

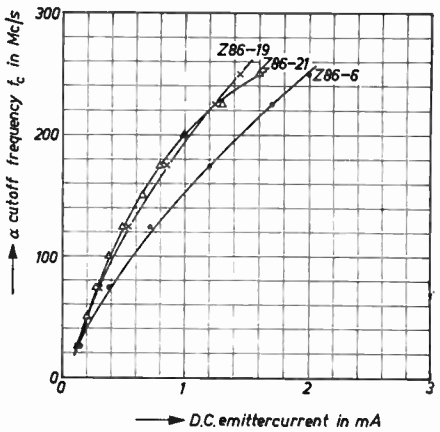


Fig. 5—The α cutoff frequency as a function of emitter current.

emitter depletion-layer capacitance is very small compared to the "diffusion" capacitance.

The current generator $\alpha_f i_e$ represents that part of the emitter current which flows across the collector junction. At low frequencies α_f is substantially unity, because of the high emitter efficiency and the low recombination losses in the thin base layer. At high frequencies α_f decreases because part of the emitter current is used for charging and discharging the capacitance C_c. Hence the α cutoff frequency f_c is given by $f_c = 1/2\pi r_c C_c$.

By virtue of the extremely thin base layer the α cutoff frequency of the "intrinsic" transistor is very high: 300–500 mc. The presence of the emitter depletion-layer capacitance, however, sets a limit to the α cutoff frequency f_c of the actual transistor. This limitation is most pronounced for low emitter currents as in that case the emitter depletion-layer capacitance considerably exceeds the "diffusion" capacitance which roughly determines the f_c of the "intrinsic" transistor.

In Fig. 5 we have plotted the measured f_c of different transistors as a function of I_e. In the neighbourhood of $I_e = 1$ ma the agreement between the f_c-values and the calculated values $1/2\pi r_c C_c$ was found to be quite satisfactory.

Usually the collector junction differential resistance r_c is quite large and has no significance in the megacycle frequency range.

C_c is the collector depletion-layer capacitance. As the collector-base junction has been formed by diffusion of a donor impurity into p-type germanium, we expect this junction to be of the "graded" type.[6] Consequently C_c should be proportional to $V_{cb}^{-1/3}$. This is confirmed by measurements, which also showed that C_c is practically independent of I_c.

The collector series resistance r_s is due to the presence of the p-type material between collector junction and collector contact. Measurements in the near-saturation region of the I_c vs V_{cb} characteristic indicate a collector series resistance value of about 50 ohms. In normal small signal operation below 50 mc, r_s is negligible compared to the transistor-output impedance.

Under conditions of strong saturation—as could happen in pulse applications—the potential difference across the collector junction can change its sign. This leads to an increase in carrier concentrations in the p-type collector material which is accompanied by a modulation of the resistivity and a drop in r_s. The storage capacitance associated with this effect is quite large, because of the wide collector region. Consequently strong saturation seriously affects the speed of response of alloy-diffused transistors in hf switching applications.

One may consider r_b as the effective resistance residing in the base layer between the emitter junction and the base contact. From Fig. 2 it is clear that the collector-junction area covers the whole base region and consequently the collector capacitance will be distributed all over this region and the resistance r_b associated with it. Measurements indicate that at not too high frequencies the distributed collector capacitance can effectively be lumped into a single element C_c which is electrically connected to a tap on the resistance r_b. The part of r_b between this point of connection and the base contact is called r_b'. Both r_b and r_b' depend strongly on transistor-geometry and distribution of resistivity in the base layer.

As r_b' determines the feedback between collector and emitter in the common base circuit, r_b' may be regarded as the "reverse" or "feedback" base resistance, while r_b may be regarded as the "forward" base resistance.

It should be mentioned that the Early effect[7] also contributes to the feedback. In order to incorporate the Early effect into the equivalent circuit of Fig. 3, a voltage generator $\mu v_{c'b'}$ should be inserted as shown in Fig. 6, $v_{c'b'}$ being the ac voltage across the collector junction and $\mu v_{c'b'}$ being the fraction of this voltage appearing across the emitter junction.

[6] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

[7] J. M. Early, "Effects of space-charge layer widening in junction transistors," PROC. IRE, vol. 40, pp. 1401–1406; November, 1952.

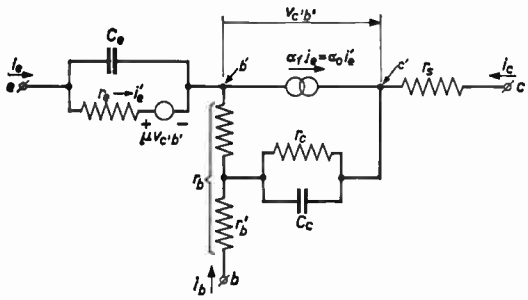


Fig. 6—Introduction of the Early-feedback voltage-generator $\mu v_{c'b'}$ into the equivalent circuit of Fig. 3.

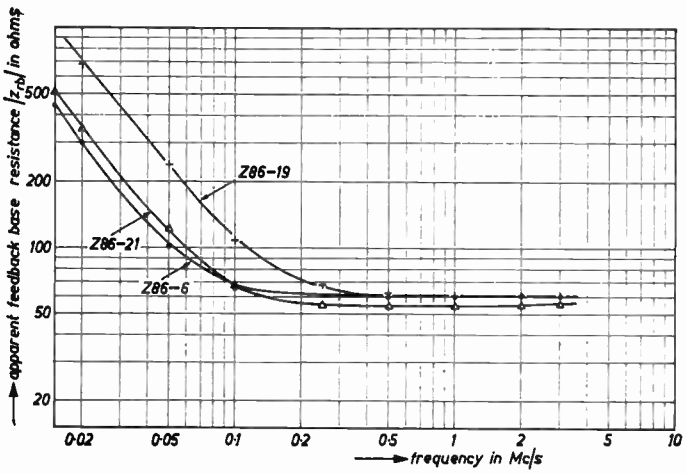


Fig. 7—The apparent feedback base resistance $|z_{rb}|$ as a function of frequency $[I_e = 1$ ma; $V_{cb} = -6$ v$]$.

The Early effect and r_b' together give rise to an apparent feedback base resistance

$$z_{rb} = \left(\frac{v_{eb}}{i_c}\right)_{i_e=0} = r_b' + \frac{\mu r_c}{1 + j\omega r_c C_c}.$$

In Fig. 7 we have drawn the measured $|z_{rb}|$ values vs frequency. Because of the relatively large time constant r_cC_c—of the order of 10 μsec—z_{rb} is constant and the Early effect is negligible above 0.1 mc. From the curves we deduce that the fraction μ lies between 10^{-4} and 10^{-5}.

As we have seen, an essential feature of the alloy-diffused transistor is the occurence of a tapped base resistance in the equivalent circuit. In principle the conventional alloy transistor shows the same feature;[8] but due to the difference in geometry, the tap on r_b is so close to the emitter that effectively one has a single resistance.[9]

The π-Equivalent Circuit

In Fig. 8 we present the hybrid π-equivalent circuit for common emitter obtained by transformation of the

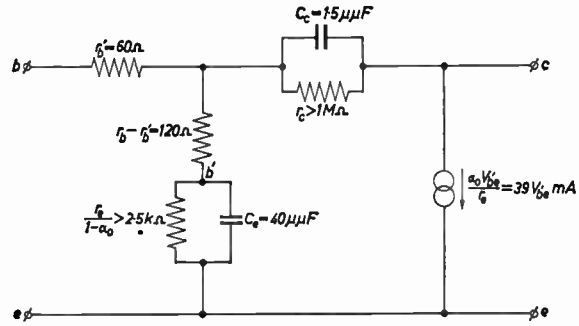


Fig. 8—The hybrid π-equivalent circuit in common emitter as follows from transformation of the circuit of Fig. 3. Numerical values are obtained by averaging over measurements on a large group of transistors.

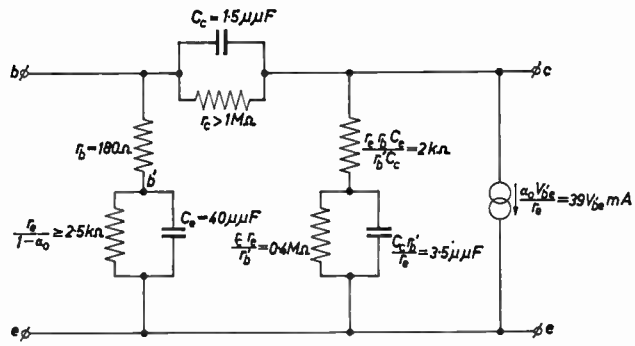


Fig. 9—The π-equivalent circuit in common emitter, as follows from transformation of the circuit of Fig. 3. The numerical values are obtained by averaging over measurements on a large group of transistors.

T-equivalent circuit of Fig. 3. Except for the resistance r_b–r_b' the circuit is quite similar to the one proposed by Giacoletto[9] for the alloy transistor.

In order to facilitate the evaluation of the admittance parameters we present Fig. 9 in which the influence of the "external" resistance r_b' of Fig. 8 has been incorporated in the branches of a genuine π-circuit.

In both the circuits of Figs. 8 and 9 the influence of the collector series resistance as well as the Early effect has been neglected as it is insignificant in the frequency range 0.1–25 mc.

As an illustration of the validity of the equivalent π-circuit of Fig. 9, we finally present measurements of some admittance parameters.

Figs. 10 and 11 show the input and output conductance as a function of frequency, measured on two transistors. The agreement with the π-circuit of Fig. 9 is good to about 50 mc.

Fig. 12 shows the transfer admittance $|y_f|$ as a function of frequency. From Fig. 8 or Fig. 9 it is easily derived that

$$|y_f| \approx \alpha_o / r_e \sqrt{1 + (\omega r_b C_e)^2}.$$

Up to 50 mc the agreement between this expression and the curves of Fig. 12 is good.

[8] J. M. Early, "Design theory of junction transistors," *Bell Sys. Tech. J.*, vol. 32, pp. 1271–1312; November, 1953.

[9] L. J. Giacoletto, "A study of *p-n-p*-alloy junction transistor from d-c through medium frequencies," *RCA Rev.*, vol. 25, pp. 502–562; December, 1954.

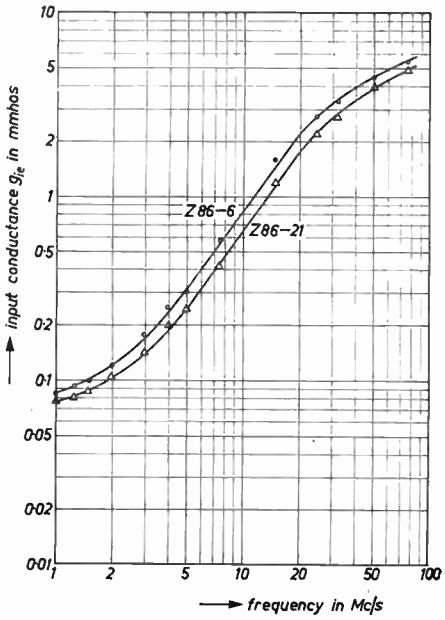


Fig. 10—The input conductance g_{ie} as a function of frequency [$I_e = 1$ ma; $V_{cb} = -6$ v].

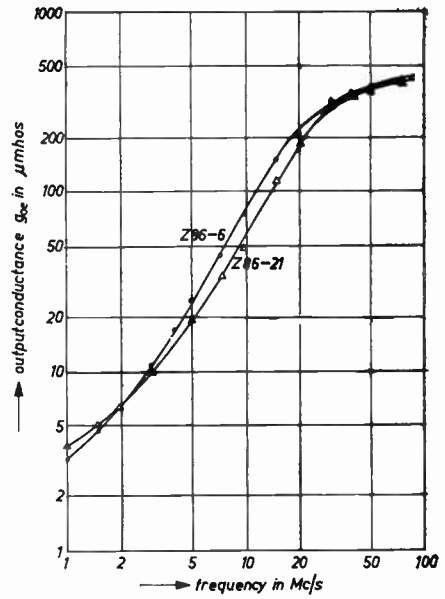


Fig. 11—The output conductance g_{oe} as a function of frequency [$I_e = 1$ ma; $V_{cb} = -6$ v].

Conclusions

The alloy-diffusion process provides an attractive method for the manufacture of high-frequency transistors. In the geometry of Fig. 2 α cutoff frequencies of 150 mc or more are obtainable at $I_e = 1$ ma.

Measurements show that the cutoff frequency at not too high emitter currents is strongly dependent on the emitter depletion-layer capacitance. This means that the cutoff frequency is considerably lower than the value corresponding to the transit time of minority carriers across the base layer.

In the frequency range 0.1 to at least 25 mc the ac behavior of the transistor is characterized by a simple

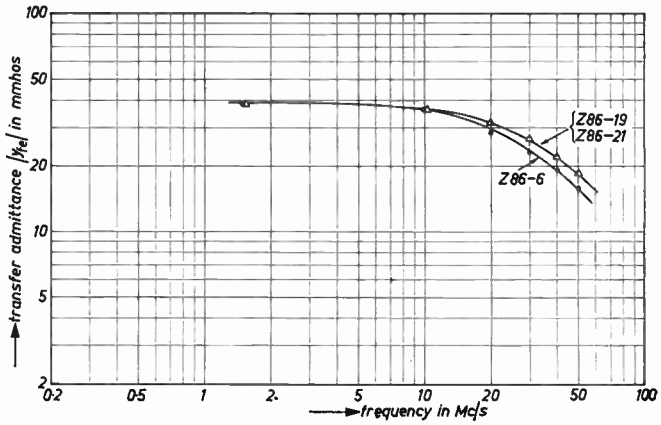


Fig. 12—The transconductance $|y_f|$ as a function of frequency [$I_e = 1$ ma; $V_{cb} = -6$ v; $T = 20$°C].

T-equivalent circuit in which the collector capacitance is connected to a tap on the effective resistance between the emitter junction and the base contact.

List of Symbols

I_e, I_c, I_b, V_{cb}, V_{ce}, V_{eb} = dc currents and voltages.

i_e, i_c, i_b, v_{eb}, v_{ce}, v_{eb} = small amplitude ac currents and voltages.

y_{ie} = input admittance, common emitter circuit, output short-circuited.

y_{fe} = forward-transfer admittance, common emitter circuit, output short-circuited.

y_{oe} = output admittance, common emitter circuit, input short-circuited.

$y = g + jb$ — g: conductance; b: susceptance.

z_{rb} = reverse transfer impedance, common base circuit, input open-circuited.

r_e, r_b, r_b', r_c, r_s, C_e, C_c = elements of the "physical" equivalent circuit of Fig. 3.

α_f = forward-current transfer ratio, common base circuit, output short-circuited.

α_o = low frequency value of α_f.

f_c = cutoff frequency of α_f.

μ = Early-feedback factor $= (v_{eb}/v_{cb})_{i_e=0}$.

Acknowledgment

The authors are indebted to P. W. Haayman for his interest and encouragement and to A. J. W. M. van Overbeek and J. te Winkel for stimulating discussions.

They also wish to acknowledge the care with which J. T. Löbbes has made a number of alloy-diffused transistors and the help of A. H. Dekker and B. C. Bouma in performing the measurements.

Technology of Micro-Alloy Diffused Transistors*

C. G. THORNTON†, SENIOR MEMBER, IRE, AND J. B. ANGELL‡, SENIOR MEMBER, IRE

Summary—Various designs of the micro-alloy diffused transistor (MADT) have been perfected for different classes of service. A high-speed switching version, operating in low-voltage saturating circuits with a current gain of 10, gives turn-off and turn-on times of 5 mμsec. Base-gated scale-of-two counters have operated at input rates as high as 140 megapulses per second with these transistors. A high-frequency, low-noise amplifier provides neutralized gain of 20 db at 100 mc with a noise figure of 4 db. A high-frequency, high-power oscillator-amplifier having a thermal impedance of 100°C per watt yields power outputs as great as 1.0 w at 70 mc. This paper discusses the control of design variables, including base width, impurity gradient, and positioning of emitter and collector junctions in the gradient, and describes some of the new circuit and measurement techniques evolved for these units.

Introduction

CONSIDERABLE interest has developed during recent years in choosing optimum manufacturable designs for very-high-frequency transistors. The majority of the work in this field has been directed towards the development of triode structures using some combination of an emitter, a base having graded resistivity (with or without an intrinsic region), and a collector. The problem of obtaining a closely-spaced reproducible geometry for ultra-high-frequency performance has been approached by fabricating the transistor by repeated diffusion on a wafer or, in the case of a grown junction transistor, by alternate ingot doping and diffusion. With these methods one or both of the junctions may actually be formed during the diffusion process. For operation at a somewhat lower frequency, transistors have been made by a single diffusion cycle on a comparatively thick wafer, followed by the application of alloyed electrodes in a conventional manner.

The approach used to fabricate the devices described in this paper involves the use of a single complex unidirectional diffused layer in which the emitter and collector contacts are precisely positioned at optimum impurity-concentration levels by means of electrochemical etching, plating, and micro-alloying. The advantages of this approach over other methods include:

1) Precise positioning of the emitter and collector junctions at locations in the field which are optimum for the intended circuit application.
2) The elimination of the need for an extremely close spacing between a metal base contact and the emitter to achieve low r_b'.
3) The use of a graded surface layer in the region around the emitter to reduce surface recombination.
4) The possibility of obtaining very low series electrical and thermal resistance between the transistor collector and the terminals or case of the transistor.

The essential tools for making micro-alloy diffused base (MADT) transistors by electrochemical techniques were announced over a year ago.[1] It is the purpose of this paper to show that through proper control of the chemical and physical processes involved and a better understanding of the circuit requirements for the device, it has become possible to optimize individual transistor designs for specific applications and to show that new levels of circuit performance now can be expected. Particular emphasis will be placed on the use of graded-base transistors in new types of switching circuits. The designs discussed herein, though not yet generally available to the circuit engineer, are all considered to be as producible as present commercial types and are intended to indicate present capabilities of the process.

Through proper optimization of the design and careful control of the diffusion and etching processes, it has become possible to fabricate the following three classes of devices:

1) Very high-speed switching transistors having turn-on time and the storage-plus-fall time each in the range of 5 to 10 mμsec. These values are obtained in low-voltage saturating circuits having a current gain of 10. Using special base-gated circuits to be described, scale-of-two counters have been successfully operated with inputs up to 140 megapulses per second.
2) A low-noise, low-level vhf amplifier with a maximum frequency of oscillation greater than 1,000 mc and capable of producing 15 db of power gain at 200 mc. A noise figure of 4 db at 100 mc is typical for this device.
3) Power oscillators or amplifiers capable of delivering rf power levels in the range of 0.5 to 1.0 w at 100 mc.

General

The general geometry for the MADT transistor is shown in Fig. 1. The first step in the fabrication process is the diffusion of the base layer impurity (phosphorus or arsenic in this case) into high resistivity n-type ger-

* Original manuscript received by the IRE, March 11, 1958. The work described was supported in part by the U. S. Army Signal Corps under contract No. DA-36-039-SC-72705, and by the U. S. Navy, Bureau of Ships, under contract No. NObsr 72705.

† Semiconductor Product Development Lab., Lansdale Tube Co., Lansdale, Pa.

‡ Research Div., Philco Corp., Philadelphia, Pa.

[1] R. A. Williams, "Diffused-base surface-barrier transistors," presented at the PGED Annual Technical Meeting, Washington, D. C., October 25–26, 1956.

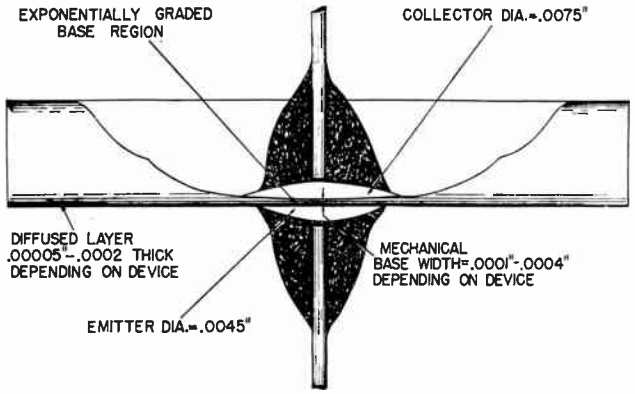


Fig. 1—The geometry of the micro-alloy-diffused base transistor.

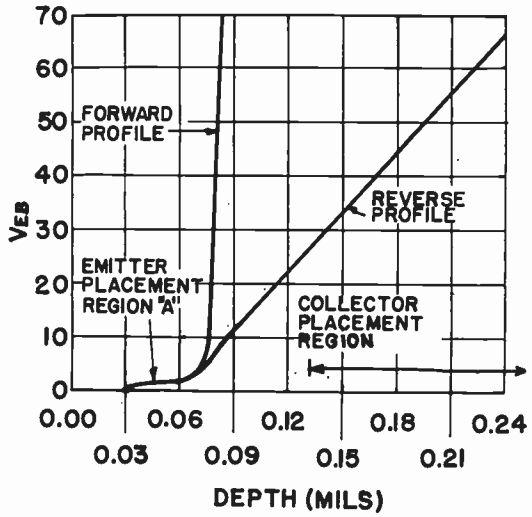


Fig. 2—Typical voltage-breakdown profile.

manium. The process is carried out in an open-tube continuous-flow apparatus with hydrogen as a carrier gas. Time, temperature, and dopant concentration are controlled to produce an impurity gradient extending to a depth of from 0.05 mil to 0.25 mil beneath the surface of the wafer, depending on the application for which the device is intended. Emitter and collector pits are then etched into opposite sides of the wafer. Next, electrodes are plated in these pits. The latter two steps are controlled so as to place the junctions at the points of desired impurity concentrations after micro-alloying. In the micro-alloying process, the electrode penetration is typically less than 0.01 mil. A gallium-indium emitter contact is used to obtain an adequately high injection efficiency.

One method that has been used to explore the impurity distribution of the diffused region is to make a breakdown-voltage or electrode-capacitance profile extending from the surface to the interior of the semiconductor. In this method, a pit is etched into the transistor blank in precise increments. At each new depth, a surface-barrier diode contact is plated onto the bottom of the pit and the breakdown voltage is measured. This process is carried out from both the low and high resistivity sides of the gradient, yielding forward and reverse profiles respectively. The data is plotted with reference to the low resistivity surface, as in Fig. 2. This data is most useful in choosing the region of placement of the emitter and collector contacts for the various classes of the device. Concerning the reverse profile and the problem of collector placement, it is necessary to realize that although the critical avalanche-breakdown field is reached in the small region close to the collector, the potential or breakdown voltage is an integrated function of the field distribution across a region of varying resistivity reaching well in towards the emitter contact.

In the low resistivity region of the forward profile the depletion layer is very shallow, and the breakdown voltage is a close measure of the impurity concentration at that point. As the penetration increases so that the emitter depletion region reaches the high resistivity portion of the gradient, the depletion-layer width increases rapidly and the voltage which can be sustained without reaching a critical field under the emitter becomes large. The curve rises very steeply in this region. The actual depth of the graded-impurity layer is therefore slightly greater than suggested by the forward profile. Actually, the diffusion process is completely definitive and reproducible, and once established, profiling is required only occasionally.

It has been shown[2] that superior high-frequency properties will be obtained with an exponential diffused gradient rather than the error-function distribution obtained with constant source conditions. It can be further demonstrated that an exponential distribution will actually produce the minimum transit time for a given base width. The diffusion equation has, therefore, been analyzed in terms of the required boundary conditions necessary to produce a true exponential impurity distribution. As derived in Appendix I, this result is achieved when boundary conditions are continually adjusted or programmed in such a manner as to cause the source condition itself to increase exponentially during the diffusion time. Since the partial pressure of the source gas is an exponential function of temperature, the temperature of the source is increased linearly throughout the major portion of the diffusion process.

In order to minimize series base resistance, it is highly desirable to produce a heavily doped, degenerate surface layer extending in from the surface to a depth of approximately 0.03 mil. With the proper geometry of the emitter etch pit and accurate placement of the emitter electrode, this diffused surface, which provides a low resistance path for base current, can be maintained to the very edge of the emitter, thus allowing the actual metal base connection to be placed at a convenient distance from the emitter. A second important reason for this

[2] J. L. Moll and I. M. Ross, "The dependence of transistor parameters on the distribution of base layer resistivity," PROC. IRE, vol. 44, pp. 72–78; January, 1956.

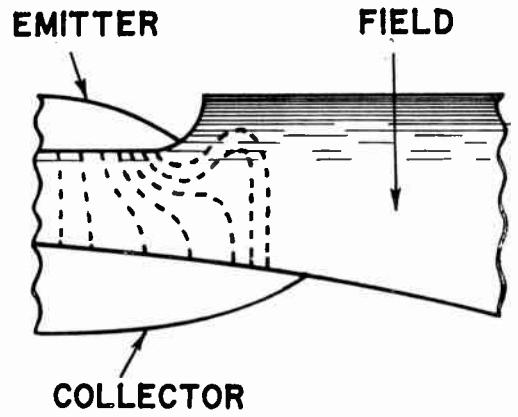


Fig. 3—The effect of heavy surface grading in preventing surface recombination.

layer is to ensure a steep gradient extending back from the depth at which the emitter is placed towards the surface. This gradient tends to prevent the flow of injected carriers to the surface and thereby minimize the effect of surface recombination. This effect is illustrated in Fig. 3. In order to obtain this result and still facilitate placement of the emitter electrode in region A of Fig. 2, the diffusant source temperature is programmed during the latter portion of the diffusion cycle to bring about an abrupt increase in concentration as the surface is approached. This effect can readily be seen in the voltage-breakdown profile.

Specific Devices

High-Speed Switch

The typical characteristics for this class of device are tabulated in Table I.

TABLE I
Characteristics of High-Speed Switching Transistor

Rise time	$t_r = 7$ mμsec
Storage time	$t_s = 5$ mμsec
Fall time	$t_f = 3$ mμsec
Frequency where $h_{fe} = 1$	$f_T = 125$ mc @ 0.5 v
Saturation resistance	$<10\Omega$
Current gain	$h_{FE} = 40$ @ 50 ma
Collector-breakdown voltage	$BV_{CBO} = -20$ v
Emitter-breakdown voltage	$BV_{EBO} = -2$ v
Collector cutoff current @ 5 v	$I_{CO} = 0.5$ μa
Collector depletion-layer capacitance	$C_c = 1.5$ $\mu\mu$f

In choosing performance objectives for a high-speed switching transistor, it is considered essential to have a high dc current gain at a high current density, a high alpha-cutoff frequency at low voltages, and a reasonable value of emitter-breakdown voltage. The values of base spreading resistance and collector capacitance which can be obtained in a graded-base design without compromising other desirable characteristics are still an order of magnitude below those found in homogeneous base transistors and provide no significant limitation on the switching times of this device.

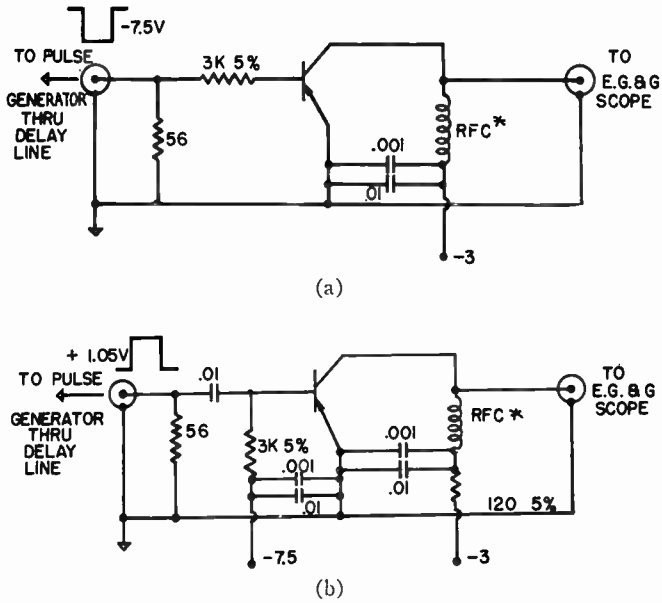


Fig. 4. Rise, storage, and fall time measuring circuits. (a) Rise time test circuit, *RFC—one watt 10 kilohm resistor wound full with No. 30 wire and (b) delay and fall time test circuit.

The above properties are obtained by proper placement of the emitter and collector contacts. An emitter contact placed too near the surface would result in a low emitter-breakdown voltage, low current gain because of poor injection efficiency, and a high emitter capacitance which would cause an unduly long turn-on delay time. Placement of the emitter at too low an impurity concentration level results in a loss of field occurring from a decrease in the collector-emitter impurity ratio and the possibility that the collector space-charge layer may punch through at the operating voltage. Equally important, the collector electrode must be placed at the very edge of the graded region in order to obtain an acceptable transit time at low voltages.

The circuits used for measuring rise, storage, and fall times are shown in Fig. 4. The ratio of collector current to forward base current was set at 10 in order to simulate conditions in the most representative applications. This ratio drives the transistor well into saturation, thereby introducing hole-storage delay time into the turn-off response. The turn-off (storage and fall) times are measured with a reverse (positive) bias of 0.5 v applied to the base. Voltage turn-off bias is used in these tests in preference to other biasing methods for various reasons:

1) Reverse voltage bias gives a good indication of the maximum speed capabilities of the transistor.
2) Most high-speed circuits, including various forms of RC coupling and base gating, are arranged so that the turn-off of one transistor is effected by either direct or capacitive coupling to a transistor entering saturation, thus providing a very low impedance. Hence, voltage turn off is more representative of such circuitry than current turn-off.

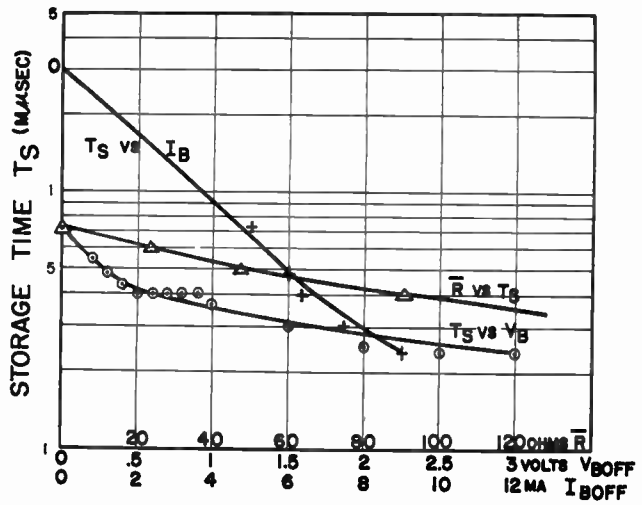


Fig. 5—The dependence of measured hole-storage delay time on reverse bias voltage (V_B Off) and on reverse base-bias current (I_B Off) for a typical 2N501 MADT switching transistor.

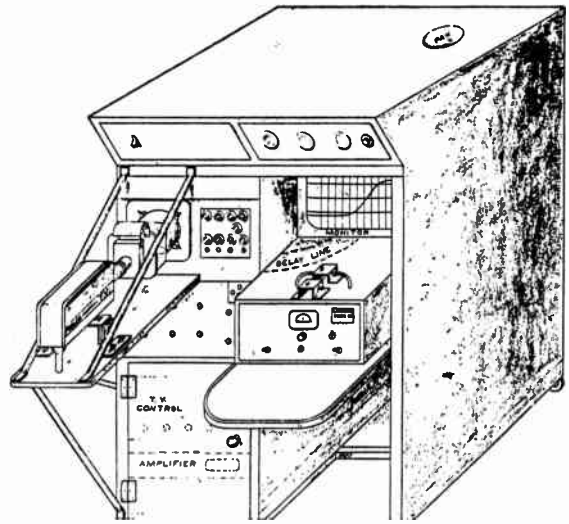


Fig. 6—Closed-circuit tv system and traveling-wave oscilloscope setup employed to measure rise, storage, and fall times.

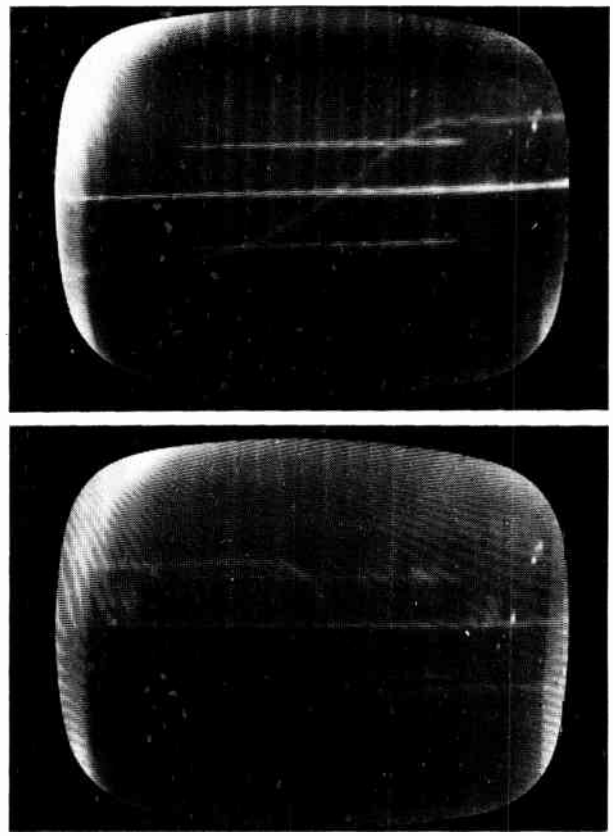


Fig. 7—Typical traces as seen on the tv screen.

3) Because of the low emitter-breakdown voltage of graded-base transistors, it is desirable to limit the reverse bias voltage. However, reverse bias cannot be controlled with current turn-off.
4) A large reverse bias on the base (which would be produced by a continuous reverse current) causes an unnecessarily long turn-on delay time because of the charge developed on the emitter and collector transition capacitances.
5) Voltage turn-off gives an indication of the limitation imposed by large values of base spreading resistance (r_b') on the removal of stored charge, whereas turn-off time with reverse current bias would not be affected by r_b'.

Fig. 5 shows the dependence of measured hole-storage delay time on reverse bias voltage (V_B Off) and on reverse base-bias current (I_B Off) for a typical 2N501 MADT switching transistor. The curve of storage time vs turn-off voltage shows that increasing the bias beyond 0.5 v yields relatively less improvement in delay time than is obtained in going from zero to 0.5 v. The base current corresponding to a 0.5-v bias is typically 6 ma, as can be seen from the plot of storage time vs turn-off current. The ratio of voltage to current for a given storage time might be considered an average turn-off resistance; this resistance (highly nonlinear) vs storage time is plotted in Fig. 5. It is evident that this resistance increases rapidly with a decrease in storage time (increase in turn-off bias), implying that the impedance is relatively unaffected by r_b' (for the low values obtained in the MADT switch) and is governed primarily by the rate at which the stored carriers move in the base region.

The switching speeds encountered with these transistors are too fast to be read accurately with conventional oscilloscopes. Therefore, a traveling-wave oscilloscope[3] is used to measure transistor performance in the test circuits. Because the total vertical deflection of the oscilloscope trace generally encountered in these tests is less than $\frac{1}{8}$ inch, it is difficult to obtain rapid, accurate measurements by direct observation of the trace (even with a strong magnifier). This problem was surmounted through the use of a closed-circuit industrial television system that reproduces the switching-speed waveform on a 17-inch monitor screen. The complete equipment is sketched in Fig. 6. Typical traces are shown in Fig. 7.

[3] Traveling-Wave Oscilloscope, Model No. 2236, manufactured by Edgerton, Germeshausen, and Grier, Inc., Boston, Mass.

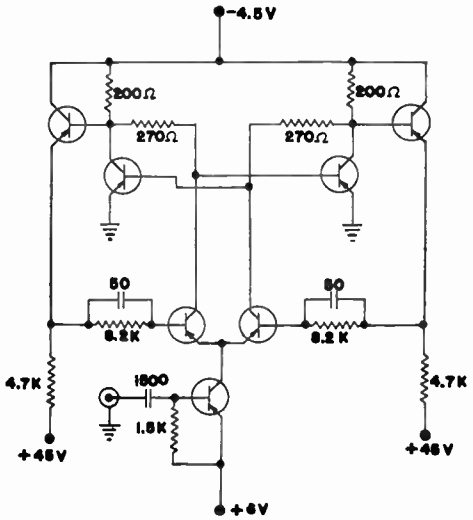


Fig. 8—Base-gated scale-of-two counter.

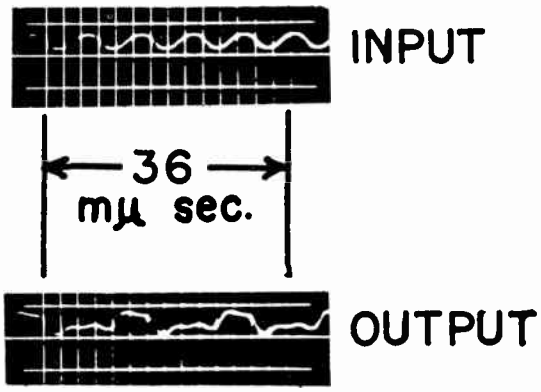


Fig. 9—130 mc counter waveforms. Input and output waveforms obtained from the base-gated scale-of-two counter employing 2N501 MADT transistors.

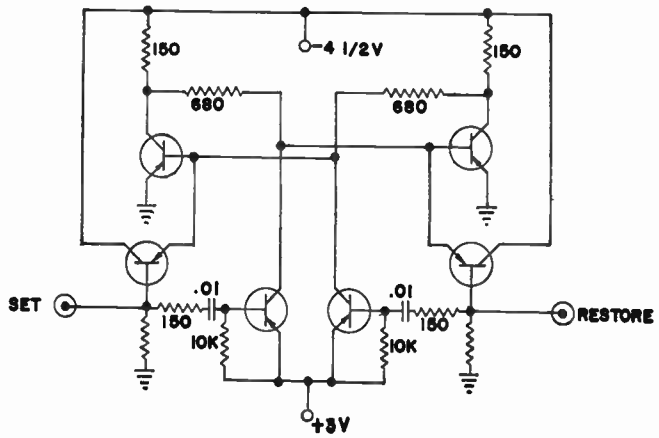


Fig. 10—Schematic diagram of double base-gated flip-flop circuit.

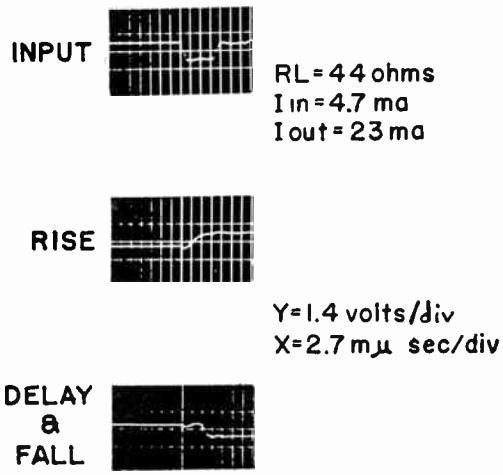


Fig. 11—Waveforms obtained from the set-reset operation of the double base-gated flip-flop.

Values of rise, storage, and fall times can be determined with this equipment to an accuracy better than 0.5 mμsec.

Base-gated circuits have been developed for high-speed switching.[4] In its simplest form, the base-gated circuit uses a saturating transistor gate to connect a positive supply potential to the base of a transistor being turned off. A scale-of-two counter built around a simple base-gated flip-flop is shown schematically in Fig. 8. Input and collector waveforms obtained using MADT switches in this circuit are shown in Fig. 9. This circuit has counted biased, sinusoidal inputs with frequencies up to 140 mc. Double base gating has been applied to flip-flops, in which a second pair of emitter-follower gates is added to provide forward bias to the on-going transistor of the flip-flop, as shown in Fig. 10. Waveforms measured with set-reset operation of this flip-flop are shown in Fig. 11. With 2N501's in this circuit, a complete flip-flop transition is completed in less than 8 mμsec, using only 4.7 ma input current and obtaining an output current swing of 23 ma in each flip-flop transistor.

The most important factor governing rise time of a switching transistor is the time delay of the collector current with respect to the emitter current. This time delay is affected primarily by the minority-carrier transit time from emitter to collector and by the emitter-transition capacitance. The quantity f_T (the frequency where $h_{fe}=1$)[5] can be used to describe the switching-speed capabilities of these units (See Appendix II). It can be determined from the analysis of Moll[6] that, for a transistor being driven well into saturation, the rise time from 10 per cent to 90 per cent of the voltage transition is given by:

$$T_r \approx 0.8 \frac{I_C}{I_B} \times \frac{1}{2\pi f_T} .$$

[4] Base-gating techniques were conceived and developed by A. K. Rapp and M. M. Fortini of the Philco Research Division, and were described in their paper "Higher-Speed Switching Circuits," presented at the IRE-AIEE Transistor and Solid-State Circuits Conference, Philadelphia, Pa.; February 20, 1958.

[5] R. L. Pritchard, "Transistor tests predict high-frequency performance," *Electronic Industries & Tele-Tech*, vol. 16, pp. 62 ff.; March, 1957.

[6] J. L. Moll, "Large-signal transient response of junction transistors," PROC. IRE, vol. 42, pp. 1773–1784; December, 1954.

It is significant that the substitution of a frequently observed rise time of 5 mμsec with a current gain (I_C/I_B) of 10 gives an f_T averaging 250 mc over the voltage range of operation (from $V_{CE} = -3$ v to -0.2 v).

The resistivity grading in the base region does not in itself improve the f_T of the transistor as much as is frequently believed[7] because of limitations on emitter and collector placement, which make it impossible to obtain the maximum possible field. In the fabrication of this device, the emitter is placed at an impurity concentration no higher than 5×10^{16} carriers per cc in order to obtain adequate injection efficiency and minimize emitter-barrier capacity. The mechanical base width is about 0.2 mil with the diffused layer extending throughout this region and falling to a concentration of about 5×10^{13} carriers per cc at the collector. Actually, at any reasonable operating voltage, the collector depletion region reaches through to a more heavily doped region which may be on the order of 5×10^{14} carriers per cc. In this case, the doping ratio between the emitter and collector is only 100. The expected ratio of transit time in a graded-base transistor (τ_n) compared with that of a homogeneous-base transistor (τ_h) can be calculated according to Kroemer:[8]

$$\frac{\tau_g}{\tau_h} = \left(\frac{\Delta V}{2KT}\right)^{3/2} = \left(\frac{\ln N_0/N_1}{2}\right)^{3/2} = \left(\frac{\ln [\rho_1/\rho_0]}{2}\right)^{3/2}$$

$$= \left(\frac{\ln 10^2}{2}\right)^{3/2} \approx 3.5.$$

It is important to note that this small effect of the built-in field is largely lost when the transistor is operated at even moderate current levels, since very high-current densities are reached and the built-in field begins to be swamped out. The major factor responsible for the decrease in transit time of this transistor, as compared with a thin homogeneous-base transistor with equivalent mechanical base width, is the two-to-one reduction in electrical base width caused by the penetration of the depletion layer through the high resistivity portion of the base, even at voltages as low as one-half volt. One result of this fact is that the value of f_T does not fall off drastically with increasing current. This fact is shown in Fig. 12, where f_T is plotted as a function of current out to a value corresponding to 2,000 amperes per cm².

As mentioned above, a *metallic* electrode is placed directly on the edge of the graded region in contact with the semiconductor to form the collector. This step has two very practical results: First, the high-speed operation of the transistor is maintained down to very low-voltage levels. In graded-base transistors produced for high-frequency amplifier applications, the collector is

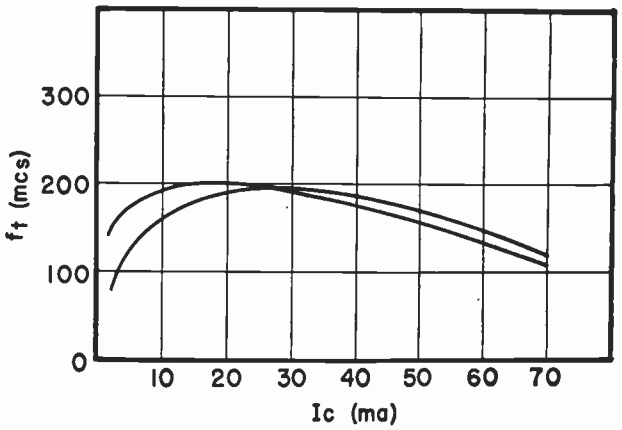


Fig. 12—The dependence of cutoff frequency, f_T, on collector current.

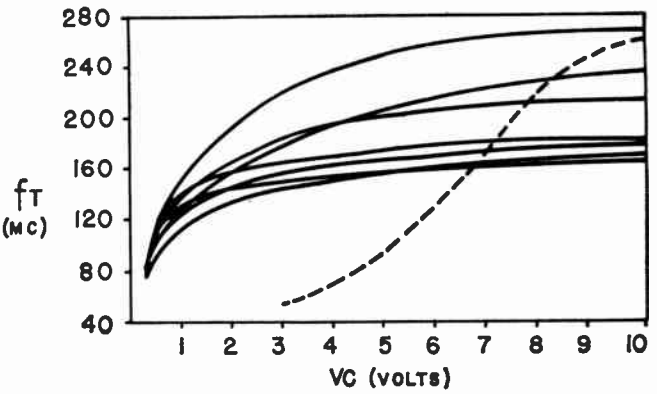


Fig. 13—The dependence of cutoff frequency, f_T, on collector voltage for MADT switching units. The dotted curve illustrates the behavior of a typical graded-base amplifier.

generally placed on intrinsic material, in order to minimize collector capacitance. In such a transistor a minimum collector voltage is required to extend the virtual collector through the intrinsic region to make the transistor operable as a high-frequency device. This voltage across the transistor in the "on" condition results in unnecessarily high-power dissipation, and a possible deleterious effect on life and reliability. The quantity f_T as a function of voltage for a graded-base switching transistor is shown in Fig. 13. The curve of a typical graded-base transistor designed for small-signal amplifier operation at high voltages is shown for comparison. Second, a very small saturation voltage is obtained for this device. In graded-base transistors designed for other than switching applications and produced in such a way as to leave semiconductor material in series with the collector or the emitter, an undesirable amount of series resistance frequently results. The low saturation-voltage characteristics of the MADT switch are seen in Fig. 14.

It is interesting to compare the current-gain linearity of a graded-base transistor with that of a homogeneous-base transistor having a similar geometry. This comparison is given in Fig. 15, for devices having comparable low-current gains and equivalent mechanical geometry. It is noted that the current-gain characteristic does

[7] H. Kroemer, "The drift transistor," *Naturwiss.*, vol. 40, p. 578; November, 1953.
[8] H. Kroemer, "On the theory of the diffusion and the drift transistor," *Arch. elekt. Übertr.*, vol. 8, pp. 223 and 363; 1954.

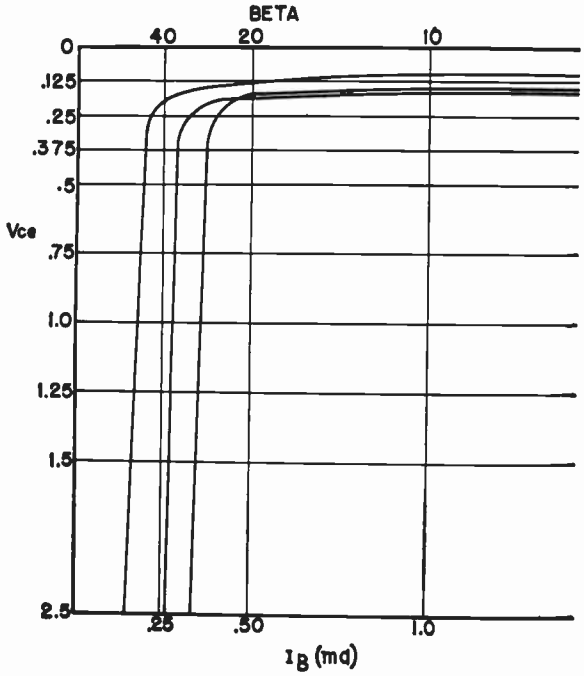


Fig. 14—Saturation voltage characteristics of the MADT switching transistor for $I_C = 10$ ma.

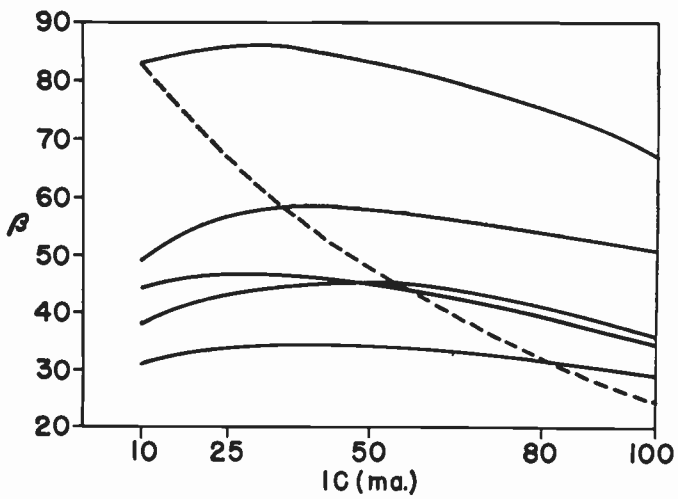


Fig. 15—Comparison of the current-gain linearity of graded-base transistors with that of a homogeneous-base transistor of similar geometry.

not fall off drastically in the graded-base transistor even at the 2,000 ampere per cm² level. This illustration points up the extreme effectiveness of the built-in field and the effective reduction in base width in preventing conductivity modulation under the emitter electrode which would limit the injection efficiency of the device.

High-Frequency, Small-Signal Low Noise Amplifier

The objectives that are of primary importance in the design of a high-frequency, low-level amplifying transistor are adequate gain at the frequency of interest and a noise figure as low as possible. High gain at high frequencies implies the need for a high maximum frequency of oscillation, F_{max}. A good noise figure requires a reasonably high α at the frequency of operation and a low base resistance. Fortunately, the majority of the design variables available for improving the maximum frequency of oscillation also improve the high-frequency noise figure. Typical parameters of a device optimized for this application are shown in Table II.

TABLE II
TYPICAL CHARACTERISTICS OF HIGH-FREQUENCY AMPLIFIER

Maximum frequency of oscillation	$F_{max} = 1{,}000$ mc
Base-resistance collector-capacitance product	$r_b'C_c = 15$ $\mu\mu$sec
Frequency where $h_{fe} = 1$	$f_T = 600$ mc
Low-frequency current gain	$h_{fb} = 0.98$
Collector-voltage breakdown	$BV_{CBO} = -60$v
Emitter-voltage breakdown	$BV_{EBO} = -1$v
Collector cutoff current @ 15 volts	$I_{CO} = 0.5$ μa
Collector depletion-layer capacitance	$C_c = 0.5$ $\mu\mu$f.

In general, the maximization of F_{max} requires the minimization of base spreading resistance, r_b', of collector capacitance, C_c, and of transit time, since $F_{max} \approx (f_T/r_b'C_c)^{1/2}$, even though the low-frequency current gain, h_{fe}, of the device is reduced considerably below that customarily found in homogeneous-base transistors. Therefore, in order to minimize r_b', the emitter of an MADT high-frequency amplifier is generally placed with its periphery at a point of lower resistivity in the diffused layer than is used in the switching transistor. However, a lower limit is set on the average resistivity under the emitter by emitter transition capacitance, C_{te}, which increases as the resistivity is decreased, and by injection efficiency. C_{te} limits the high-frequency current gain, particularly at the low-current levels which are desired for low-noise operation. Low injection efficiency increases the noise figure because of the increase in $1-\alpha_0$, and increases the difficulty of bias-point stabilization. Therefore, a compromise has been effected in this transistor, in which the emitter is placed at an average concentration of 1×10^{16} carriers per cc. In order to minimize r_b', the etch pit in which the emitter is placed is formed with steep sides and no larger at the bottom than the emitter electrode itself, with the result that the spacing between the edge of the emitter and the low-resistivity skin on the wafer is kept very small. Finally, to minimize the collector capacitance, the collector is positioned on high-resistivity material (impurity density $<5\times10^{13}$ carriers per cc) including sufficient thickness of intrinsic material between the electrode and the graded base to permit an inherent breakdown voltage of 100 v or more.

The matched, neutralized power gain vs frequency for this device is shown in Fig. 16. Many experiments have shown that the high-frequency power gain varies inversely with the square of frequency (−6 db per octave) over many octaves, even though gain above 500 mc can only be accurately determined in a coaxial-line test set.[9]

[9] "Transistor testing with type 874 coaxial elements," *Gen. Rad. Exper.*, vol. 32, pp. 3–6; October, 1957.

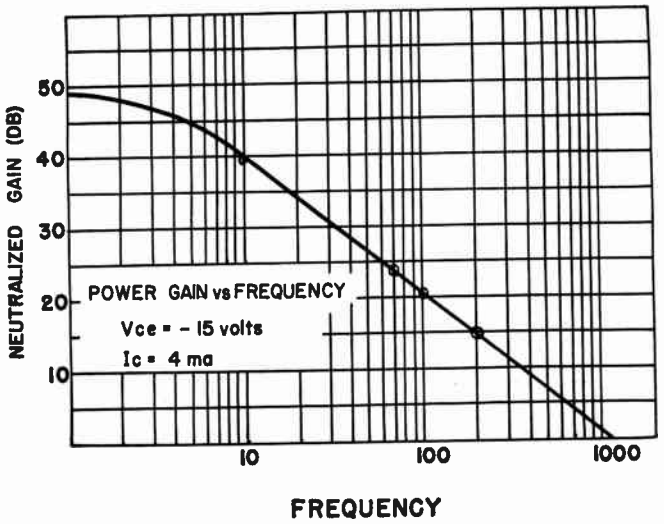


Fig. 16—Matched, neutralized power gain vs frequency data of the high-frequency, small-signal, low-noise amplifier.

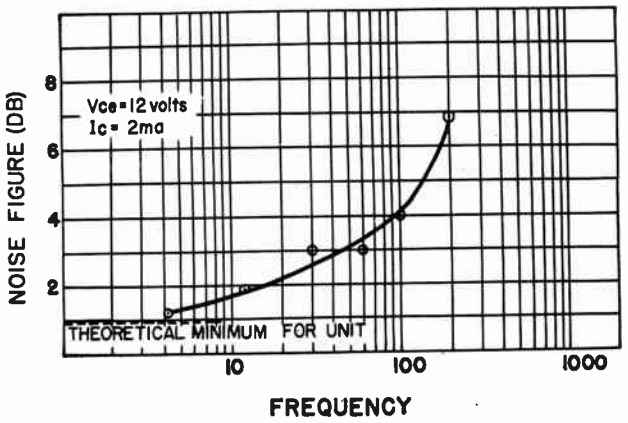


Fig. 17—Noise-figure measurements as a function of frequency obtained with a common-emitter neutralized amplifier. At frequencies above 70 mc, the amplifier was matched at both input and output; below this frequency, the collector circuit was mismatched to limit the gain and avoid regeneration.

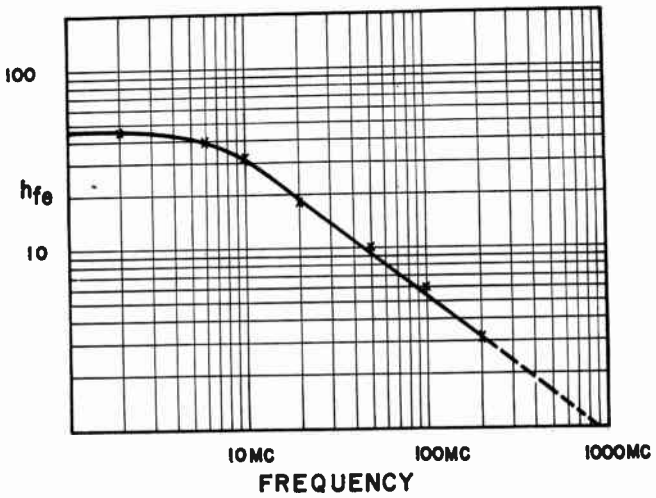


Fig. 18—Common-emitter current-transfer ratio vs frequency for an MADT amplifier with $V_{CE} = -15$ v.

The measured noise figure of a representative unit as a function of frequency is shown in Fig. 17. The transistor bias was adjusted for best noise figure, which corresponded to a collector bias of -2 ma and -12 v. These noise-figure measurements were obtained with a common-emitter neutralized amplifier, using a noise diode as the calibrated source. At frequencies above 70 mc the amplifier was matched at both input and output; below this frequency, where the gain of the transistor becomes extremely high, the collector circuit was mismatched in order to limit the gain and avoid undesired regeneration.

These transistors can also be used as video amplifiers, because of their very high current-gain-bandwidth product. In an iterative amplifier,[10] the current-gain-bandwidth product is given by f_T or by the product of h_{fe} and its frequency of measurement. A plot of h_{fe} vs frequency is shown in Fig. 18. It can be seen that gain-bandwidth products greater than 600 mc can be achieved with these units.[11]

High-Power Oscillators and Amplifiers

The basic design problem which must be solved in order to extend the power-handling capabilities of these devices is to provide a path of suitably low thermal resistance for heat flow from the junction of the transistor to the case, while maintaining the other desirable high-frequency properties. In order to avoid mechanical problems with regard to possible shock and vibration failure, it becomes necessary to alter the internal structure in such a manner as to center the mass of the transistor on a single metal pedestal with no other rigid connections to the transistor body. The design chosen is shown in Fig. 19. It is to be noted that all extraneous germanium previously required for mounting the essential parts of the transistor has been eliminated. The collector pedestal consists of a silver stud tapered to a flattened point approximately 6 mils in diameter, which is separated from the actual collector junction by approximately 1 mil of indium. The stud itself is integrally welded to the transistor case. The latter is designed in such a fashion that it may be affixed directly to an appropriate dissipator. Both calculated and measured values for the internal thermal resistance to the case indicate a value of approximately 110°C per watt. It can therefore be seen that with a mounting base-plate temperature of 30°C and a maximum junction temperature of 100°C, approximately 500 mw can be readily dissipated in the transistor. In circuits operating at greater

[10] L. P. Hunter, "Handbook of Semiconductor Electronics," McGraw-Hill Book Co., Inc., New York, N. Y.; pp. 12-6 to 12-11; 1956.

[11] The straight-line extrapolation of the plot in Fig. 18 to $h_{fe} = 1$ at 800 mc may be slightly optimistic, since the slope over the region of measurement is slightly smaller than the usual -6 db per octave. This unusual slope is thought to be caused by the combined effects of the frequency dependence of emitter-injection efficiency at the lower frequencies and of small measurement inaccuracies toward the upper frequency range. However, using even a conservative $f_T = 600$ mc, corresponding to $h_{fe} = 6$ at 100 mc, the "alpha-cutoff frequency" of this transistor is in the range of 900 to 1200 mc (see Appendix II).

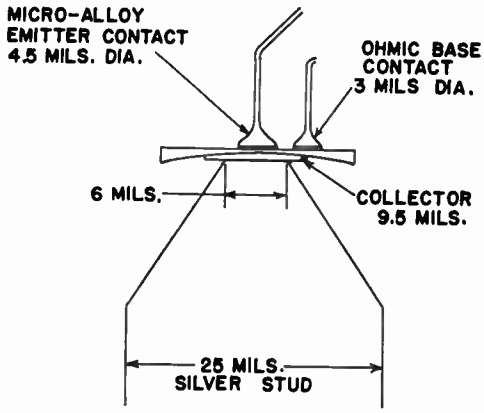


Fig. 19—Geometry of the high-power MADT oscillators and amplifiers.

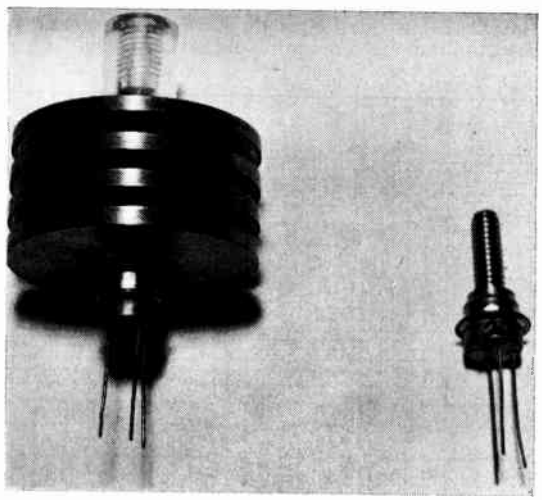


Fig. 20—Completed transistor, with and without dissipator.

than 70 per cent efficiency, it then becomes possible to deliver rf power on the order of 1 w to a load.[12] The completed transistor with and without dissipator is shown in Fig. 20.

The electrical parameters for a typical stud mounted transistor are tabulated in Table III.

TABLE III

TYPICAL CHARACTERISTICS OF HIGH-POWER, HIGH-FREQUENCY TRANSISTOR

Maximum frequency of oscillation	$F_{max} = 600$ mc
Amplifier efficiency @ 70 mc	$= 70 - 80\%$
Oscillator efficiency @ 100 mc	$= 60\%$
Base-resistance collector-capacitance product	$r_b'C_c = 20$ $\mu\mu$sec
Frequency where $h_{fe} = 1$	$f_T = 80$ mc @ 0.5 v
Current gain	$h_{fe} = 15$
Collector-breakdown voltage	$BV_{CBO} = -40$ v
Emitter-breakdown voltage	$BV_{EBO} = -0.5$ v
Collector depletion-layer capacitance	$C_c = 2$ $\mu\mu$f
Collector-cutoff current @ 10 v	$I_{CO} = 2$ μa
Thermal resistance to case	$= 0.11$°C/mw
Thermal resistance to free air	$= 0.28$°C/mw

[12] Such applications, in which the total dc input power far exceeds the dissipation rating of the active element, have been standard practice for many years in high-power vacuum-tube technology. It is necessary that the load be connected at all times to avoid excessive dissipation in these cases.

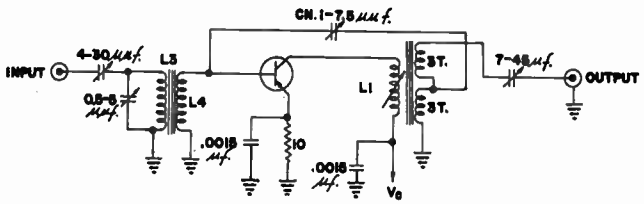


Fig. 21—Schematic diagram of the amplifier-test circuit.

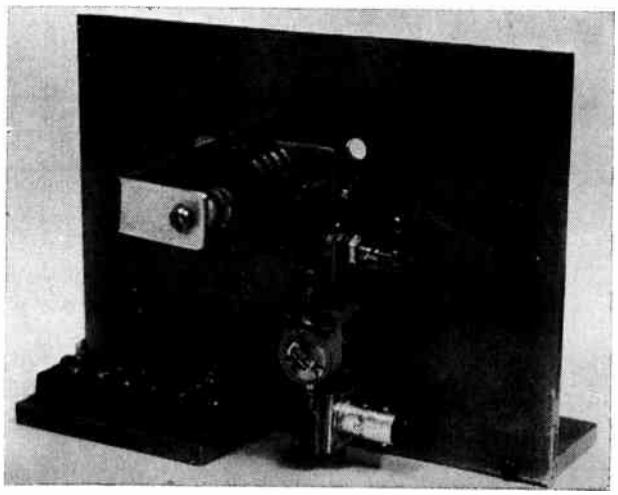


Fig. 22—Amplifier-test chassis.

The parameter requirements for good circuit operation of the MADT amplifier at the high levels made possible by this construction do not differ appreciably from those of an amplifier designed for low-level application, with one exception. The transistor should, if possible, maintain good gain characteristics down to low collector voltages. This requirement is brought about by the fact that for efficient class-B or class-C operation, the collector should swing to near zero voltage so that the peaks of conduction will take place under conditions of lowest collector dissipation and consequently greatest efficiency. Therefore, in the design of this device, the collector is placed somewhat closer to the graded region than in the case of the amplifier, with an unavoidable sacrifice in collector-breakdown voltage and collector capacitance. A typical value for f_T at 0.5 v is about 80 mc. The collector electrode is also made slightly larger for this device than for the previous two devices in order to facilitate the attachment of the silver stud.

One circuit used to test the amplification capabilities of the stud mounted unit is given in Fig. 21. A photograph of the test amplifier is shown in Fig. 22. Using this circuit, it has been possible to obtain an output of 1.0 w at 70 mc with a collector efficiency of 83 per cent (1.20 w dc input) and a gain of 5.2 db. Since the transistor described has a maximum frequency of oscillation in the order of 600 mc, reasonable power-output levels can be expected even at higher frequencies.

The theory of the operation of graded-base transistors and their optimization for use in high-frequency power-oscillator circuits is not at present completely definitive.

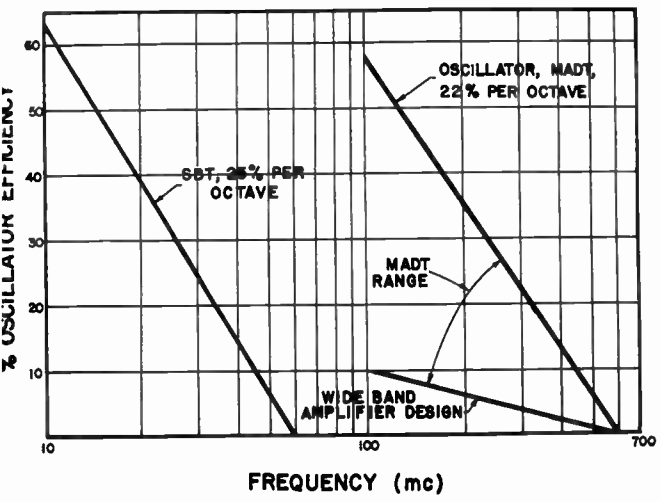


Fig. 23—Oscillator efficiency plotted as a function of frequency.

Their behavior can be examined, however, by plotting oscillator efficiency as a function of frequency. A plot of this characteristic for various transistors is shown in Fig. 23. It is seen that for a given transistor, oscillator efficiency varies approximately linearly with log F, falling to zero at the maximum frequency of oscillation. The 25 per cent per-octave fall-off characteristic shown for a surface-barrier transistor appears to be typical for this class of homogeneous-base transistor. Curves A and B represent two extremes of MADT designs showing a variation of from 4 per cent to 23 per cent per octave for units having the same maximum frequency of oscillation. It has been emphatically demonstrated that units having a minimum r_b' for a given maximum frequency of oscillation tend to have the steeper, more desirable slope. The unit chosen to represent the lower end of the range was designed for a maximum gain-bandwidth product, namely, a high f_T and a comparatively large r_b'. This result is not at all unreasonable when one considers that an appreciable rf base current flows in this application, representing a direct power loss.

Conclusion

Marked improvements in MADT performance have been achieved during the past year by virtue of judicious optimization of the design for specific applications and refinements in control throughout the fabrication process. It is significant that the results described in this paper have been achieved without requiring a major reduction in the size of the active transistor elements, compared with that of many transistor designs produced during the past several years for use in the 5–50 mc range. It is expected that with reasonable reductions in electrode size, the use of thinner diffused gradients, and coaxial package design, gain-bandwidth products and useful power gain in the frequency range of 1,000–10,000 mc can be achieved using the MADT approach. Alternatively, the use of slightly larger collector-dissipator combinations may be expected to extend the power-frequency barrier significantly beyond the 70 w-mc level described herein. Finally, extensions of the techniques described herein are expected to provide 0.3- to 1-mμsec switches and 0.5-a, high-power, memory-drive circuits with transition times in the range of 5 to 20 mμsec.

Appendix I

Programmed Diffusion

The equation governing the process of one-dimensional diffusion from the vapor into a solid crystal is

$$D\frac{\partial^2 n}{\partial x^2} = \frac{\partial n}{\partial t},$$

where n is the impurity concentration in the germanium, x is the distance from the surface, t is the time of diffusion, and D is the diffusion coefficient. If the concentration of diffusing impurity at the surface, n_0, is maintained constant throughout the diffusion process, an error-function distribution of impurity in the crystal will result:

$$n(x) = n_0\left[1 - \operatorname{erf}\left(\frac{X}{2\sqrt{Dt}}\right)\right].$$

As mentioned in the text, a more desirable distribution from the standpoint of device performance would be an exponential one. Another solution to the diffusion equation is

$$n(x) = e^{-(ax - Da^2t)}.$$

If x is set equal to zero, to obtain the desired variation in n_0,

$$n_0 = e^{Da^2t}.$$

It is seen that, by varying the concentration at the surface exponentially with respect to time, an exponential impurity distribution will be obtained within the slab at any fixed time, t.

To meet these conditions, it is necessary to begin the diffusion procedure with a low available impurity concentration and then to increase the concentration with time. The steepness of the exponential distribution, a, varies as the square root of the time rate of increase of concentration, Da^2. The final value of the concentration determines the initial height of the distribution at the surface.

The use of this process, as described in the text, assumes the absence of any rate-determining mechanism at the surface which might affect the rate of solution. Hence, it may not be directly applicable to silicon transistors, where surface oxides are known to affect the solution process.

Appendix II

f_T vs f_α

The frequency f_T, at which $h_{fe}=1$, is used as a high-frequency parameter instead of the alpha-cutoff frequency, f_α, throughout this paper, for two principal reasons. First, the measurement of f_T, as described by Pritchard, is much simpler than that of f_α, particularly for values greater than 100 mc. For the measurements reported herein, a 50-ohm adjustable attenuator, manufactured by Kay Electric Co., was used between a signal generator and a test chassis. The attenuator readings were used as the basis for comparing the current gain of the transistor at any particular frequency with that of a direct connection from base-to-collector terminals of the test socket.

A second, and more important, reason for using f_T is that it is quantitatively more useful than f_α. The former quantity gives the product of current gain and frequency at any high frequency, thus providing the iterative video-amplifier gain-bandwidth product. f_T also serves as a moderately accurate indication of $I_C/I_{B}t_r$ in a saturating switch, if it is measured at a collector voltage corresponding to some weighted average of the voltage range over which the switch is operating. On the other hand, the alpha-cutoff frequency does not suggest the effects of any excess phase in i_c/i_e, beyond that phase associated with a single-pole, low-pass filter. Since most graded-base transistors have substantial excess phase, the alpha-cutoff frequency is unduly optimistic. Because of this excess phase, $1.5f_T<f_\alpha<2.0f_T$ for most graded-base transistors.

To illustrate the importance of the excess phase of α, measurements were made of the complex value of α vs frequency on graded-base transistors. The current gain α was determined from the relation $(1-\alpha)=h_{ib}/h_{ie}$, with the input impedances measured on a Boonton RX meter. MADT's having relatively limited response were chosen, in order to keep the region of greatest interest within the frequency range of the measuring equipment. The results are plotted in Fig. 24 for two typical units. Since $h_{fe}=1$ when Re $(\alpha)=$ Re $(1-\alpha)$, f_T averages slightly less than 200 mc for these units, whereas f_α is appreciably greater than 250 mc.

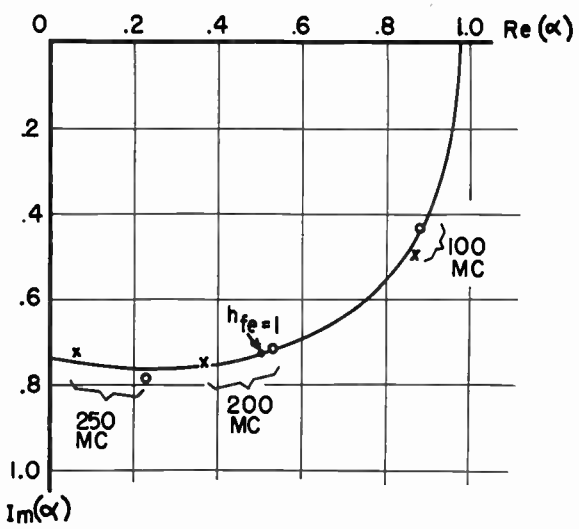


Fig. 24—Complex alpha vs frequency for two MADT's.

Acknowledgment

The authors are deeply indebted to their many colleagues who helped make possible the results described in this paper. In particular, the work of R. L. Williams and J. D. McCotter in advancing the techniques of device fabrication are manifold. F. P. Callahan and G. Lang made significant contributions to the analysis of the diffusion process and the voltage-breakdown mechanism. Finally, the circuit and measurement techniques developed by A. K. Rapp, M. M. Fortini, and W. C. Follmer have proven very valuable.

Junction Transistor Short-Circuit Current Gain and Phase Determination*

D. E. THOMAS†, SENIOR MEMBER, IRE, AND J. L. MOLL†, MEMBER, IRE

Summary—This paper presents a physical analysis which establishes that the phase shift associated with the common-base current gain of a junction transistor is of the type defined by Bode as "minimum." The phase of the common-base current gain is therefore uniquely determined by its magnitude characteristic. Using the minimum phase properties of networks and the empirically observed behavior of transistor current gain, it is shown that the complete common-base and common-emitter current-gain magnitude and phase frequency characteristics can be determined from three amplitude measurements, namely the low frequency α, α_0, the magnitude of the common-emitter current gain at a single frequency in the common-emitter cutoff region, and the common-base cutoff frequency f_α. The paper develops the equations for determining the common-emitter and common-base short-circuit current gains from these three magnitude measurements.

Introduction

SHORTLY after the first experimental junction transistors became available, it was realized that the high common-emitter current gain would be accompanied by a corresponding reduction in cutoff frequency. In order to evaluate this reduction in cutoff frequency, a resistance-capacitance type approximation to the common-base current-gain frequency cutoff characteristic was assumed.[1] The resultant common-emitter current gain also had a resistance-capacitance type frequency cutoff characteristic for which the cutoff frequency was related to the common-base cutoff frequency by the factor $(1-\alpha_0)$.[2,3]

Experimental measurements and physical analysis soon showed that most junction transistors had common-base cutoff amplitude characteristics having asymptotic slopes exceeding the 6 db per octave of the *RC* type cutoff. However, experimental measurements showed that the amplitudes of the common-emitter current gains of these transistors, including diffused base transistors having common-base cutoffs in excess of 500 mc, still had *RC* type frequency cutoffs extending right up to the common-base cutoff frequency.

These observed results led to the analysis presented in this paper of the nature of the common-base and common-emitter current-gain amplitude and phase characteristics of junction transistors as a function of frequency. The analysis covers those junction transistors whose frequency is limited either by diffusion transit time across the base layer or by emitter transition capacitance, and whose direct forward transmission through the base resistance does not appreciably affect the common-base short-circuit current gain at frequencies below the common-base cutoff frequency. This includes most of the known junction transistor types and for these transistors it is shown that:

1) The phase shift associated with the short-circuit common-base current gain is of the type described by Bode as "minimum."[4]
2) The common-base short-circuit current gain, α, can be accurately expressed as:

$$\alpha = \frac{\alpha_0}{1 + j\dfrac{f}{f_\alpha}} e^{j[(K-1)/K]f/f_\alpha} \qquad f < f_\alpha. \tag{1}$$

3) The corresponding common-emitter short-circuit current-gain, β, for transistors having a low frequency α greater than 0.9 can be accurately expressed as:

$$\beta = \frac{\alpha}{1-\alpha} = \frac{\alpha_0 e^{j[(K-1)/\sqrt{K}]f/f_\alpha}}{(1-\alpha_0)\left(1 + j\dfrac{f}{K(1-\alpha_0)f_\alpha}\right)}, \qquad f < f_\alpha \tag{2}$$

where

α_0 = the low frequency magnitude of α,

f_α = the frequency at which the magnitude of α is 3 db below its low frequency magnitude α_0,

and K is given by either

$$\frac{K-1}{K} = \Psi_{(f=f_\alpha)} + \frac{\pi}{4} \tag{3}$$

or

$$K = \frac{f_{\alpha c e}}{(1-\alpha_0)f_\alpha} \tag{4}$$

where $\Psi_{(f=f_\alpha)}$ is the phase of α at f_α and $f_{\alpha c e}$ is the frequency at which the magnitude of β is 3 db below its low frequency value, β_0.

* Original manuscript received by the IRE, February 24, 1958.

† Bell Telephone Labs., Inc., Murray Hill, N. J.

[1] A resistance-capacitance or RC frequency cutoff characteristic is one defined by

$$\frac{1}{1 + j\dfrac{f}{f_0}}.$$

[2] R. L. Pritchard, "Frequency variations of current-amplification factor for junction transistors," PROC. IRE, vol. 40, pp. 1476–1481; November, 1952.

[3] D. E. Thomas, "Transistor amplifier cutoff frequency," PROC. IRE, vol. 40, pp. 1481–1483; November, 1952.

[4] For a discussion of "minimum phase," see H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co., Inc., New York, N. Y.; 1945.

Using (1) and (2), only three magnitude measurements are needed to determine the complete short-circuit current gain and phase characteristics, common emitter or common base for most junction transistors from direct current to f_α. The values needed are the low frequency magnitude of α, α_0, the common-base cutoff frequency, f_α, and the magnitude of the common emitter-current gain at a single frequency f_c in the common-emitter 6 db per octave cutoff region. Since for all transistors of a given type under reasonably close production control, K will be the same, only the single common-emitter current-gain magnitude in the cutoff region is needed for normal production control of junction transistor cutoff frequency. This has been verified in advanced development and pilot production fabrication of germanium diffused base transistors.

The presentation of the analysis substantiating the above conclusions is divided into two sections. The first section covers the physical theory establishing the "minimum" phase characteristic of the common-base short-circuit current gain. The second section develops the expressions for common-base and common-emitter short-circuit current gains given by (1) and (2).

Physical Theory

If passive impedance cutoffs have only a second-order effect on the cutoff frequency of a junction transistor, then its frequency is limited by the diffusion transit time across the base layer. Furthermore, the slope of the amplitude of the common-base cutoff characteristic in the cutoff region is determined by the diffusion and drift of minority carriers across the base layer. In general we cannot assume that the base layer has a uniform density of impurities so that it is necessary to consider the effect of "built-in fields" on the minority carriers. A model which is sufficiently simple to allow for analysis is to assume an exponential[5,6] distribution in a planar transistor. The results for the exponential case give a qualitative picture of what to expect in more general cases. Let the base layer impurity density, N, vary as

$$N = N_0 e^{-x/L} \tag{5}$$

where N_0 is the density at the emitter junction, x is the distance measured into the base layer normal to the plane of the emitter junction, and L is a constant which depends on the fabrication process. Then from solution of the diffusion equations it follows that the transfer function is

$$\frac{\bar{I}_c}{\bar{I}_E} = \frac{e^{w/2L}}{\frac{w}{2L}\frac{\sinh Z}{Z} + \cosh Z} \tag{6}$$

where

$\bar{I}_c$, $\bar{I}_E$ are transforms of collector and emitter current,
w = width of base layer,
$Z^2 = pw^2/D + (w/2L)^2$,
p = complex variable.

Along the real frequency axis, (6) becomes

$$\left|\frac{\bar{I}_c}{\bar{I}_E}\right| = \frac{e^{w/2L}}{\frac{w}{2L}\sinh\left[\left(\frac{w}{2L}\right)^2 + \frac{i\omega w^2}{D}\right]^{1/2}} + \cosh\left[\left(\frac{w}{2L}\right)^2 + \frac{i\omega w^2}{D}\right]^{1/2} \tag{7}$$

The magnitude of (7) has been calculated for various conditions of aiding and retarding fields and is plotted in Lee.[6] The aiding fields correspond to decreasing impurity density from emitter to collector [$L>0$ in (5)] and the retarding fields correspond to increasing impurity density from emitter to collector [$L<0$ in (5)]. It is interesting to note that for uniform distribution or for an impurity distribution producing an aiding field, the cutoff region has a slope in excess of 6 db per octave, while for retarding fields the cutoff slope may be very close to the 6 db per octave RC type cutoff. The minimum phase associated with the cutoff characteristic is uniquely determined by the amplitude frequency characteristic of the current ratio given by (7).[7] Therefore, the common-base phase shift is clearly dependent on the distribution of impurities, and since the slope of the amplitude cutoff increases as the aiding field increases, the phase shift at the common-base cutoff frequency will also increase as the "aiding field" increases.

A question of basic interest is whether the phase shift of the transfer current ratio given by (6) is of the "minimum" or "nonminimum" type. The current ratio has infinitely many poles along the negative real axis. However, it has neither finite poles nor finite zeros in the right-half p plane. Accordingly, the logarithm of the current ratio has no finite poles in the right-half plane. The ratio does have a zero at infinity in the right-half p plane. The zero at infinity results in a pole in the logarithm at infinity which is of the order $p^{1/2}$. The fact that the logarithm of the transfer current ratio has no

[5] H. Kröemer, "On diffusion and drift transistor theory," *Arch. Elek. Übertr.*, vol. 8, pp. 223–228; May, 1954.
[6] C. A. Lee, "A high-frequency diffused base germanium transistor," *Bell Sys. Tech. J.*, vol. 35, pp. 23–34; January, 1956.
[7] H. W. Bode, *op. cit.*, ch. 14.

inite poles in the right-half plane, and that the pole it infinity is of order less than unity, identifies this current ratio as a minimum phase function; *i.e.*, its total phase is calculable from its magnitude vs frequency characteristic.[8]

The transistor with aiding field is so similar in concept to the drift of electrons across the grid-anode space of a vacuum tube (definitely a nonminimum phase situation) that some further discussion is warranted. A solution to the diffusion equation implies a response to a step input at any distance at any positive time. We would not expect the diffusion equation to be a good representation of the physical situation for times corresponding to velocities in excess of thermal velocities. This would be a time of

$$t < \frac{w}{v} \tag{8}$$

where

v = thermal velocity,

w = base layer width.

The thermal velocity is of the order of 10^7 cm/second so that for a base layer width of 10^{-3} cm, the first 10^{-10} seconds would be in doubt. Now the principal phase shift associated with such a base layer corresponds to a delay of the order of 10^{-7} to 10^{-8} seconds so that we must conclude that the diffusion equation is a good approximation for the highest frequency junction transistors known to have been built, and that the phase shift is primarily of the minimum type. If the transistor frequency cutoff is sufficiently high, the diffusion equation begins to be a poor approximation and the form of the transfer function would have to be reconsidered.

Current-Gain Equations

In developing the expressions for common-base and common-emitter current gains, it proved to be algebraically intractable to develop the common-emitter gain across the entire frequency range from dc to the common-base cutoff frequency in a single solution. The solution is therefore divided into two frequency ranges, the first covering the range below 0.1 f_α and the second covering the range between 0.1 f_α and f_α where the common-emitter current-gain magnitude approaches its 6 db per octave uniform slope. The two solutions are then found to merge into a single solution.

First consider a generalized common-base current-gain frequency characteristic such as is illustrated in Fig. 1. The frequency at which the amplitude of the current gain is down by 3 db from its low frequency value is designated as the common-base cutoff frequency, f_α.

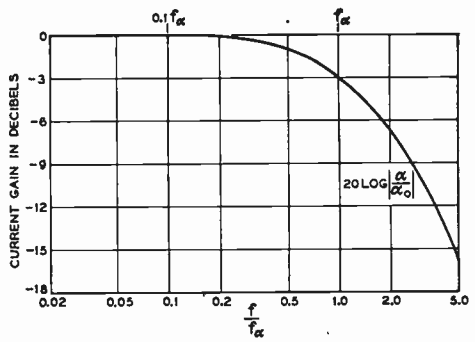


Fig. 1—Generalized common-base current-gain frequency characteristic.

Since it has been shown above that the phase of the common-base current gain is of the minimum type, the common-base phase characteristic of the generalized transistor can be determined from the amplitude characteristic of Fig. 1. Bode has shown that if we have an attenuation or gain characteristic which is nearly constant below some frequency $2f_h$, but which varies in any way at frequencies higher than $2f_h$, then the minimum phase associated with this real characteristic is very close to being linearly proportional to frequency at frequencies below f_h.[9] Since the real characteristic of Fig. 1 is well within the Bode requirements for $f_h = 0.1\ f_\alpha$, then the minimum phase, B_1, associated with this characteristic at frequencies below 0.1 f_α can be written as

$$B_1 = -\frac{1}{K}\frac{f}{f_\alpha} \tag{9}$$

where K is as yet an undetermined constant.

Since we know empirically and theoretically that $|\alpha| \doteq \alpha_0$ below 0.1 f_α, α can therefore be written

$$\alpha = \alpha_0 e^{-j(1/K)(f/f_\alpha)}, \qquad f < 0.1 f_\alpha. \tag{10}$$

If we now determine the common-emitter current-gain, β, corresponding to the common-base current gain given by (10) we get:

$$\beta = \frac{\alpha}{1-\alpha} = \frac{\alpha_0}{(1-\alpha_0)\left[1 + j\dfrac{f}{K(1-\alpha_0)f_\alpha}\right]}, \qquad f < 0.1 f_\alpha. \tag{11}$$[10]

Eq. (11) defines an RC type cutoff characteristic. No restriction was placed on the base layer impurity distribution of the hypothetical transistor whose common-base cutoff characteristic is given in Fig. 1. Eq. (11) therefore tells us that any reasonably high alpha junction transistor will have an RC type common-emitter cutoff extending up to at least 0.1 f_α. (The high alpha requirement is necessary to assure that the 0.1 f_α re-

[8] The derivation of the phase integral (H. W. Bode, *op. cit.*, pp. 305–307) requires that the integral of attenuation around a semicircle in the right-half p plane approach zero as the radius becomes infinite. If the attenuation function has no finite poles in the right-half plane, this condition is met even when there is a pole at infinity in the right-half plane providing that the pole is of order less than unity.

[9] H. W. Bode, *op. cit.*, ch. 14, p. 307.

[10] This relationship was presented without mathematical proof by D. E. Thomas at the Semiconductor Device Research Conference, Philadelphia, Pa.; June, 1955.

striction of the proof is at least above the common-emitter cutoff frequency). The constant K of (9) is given by

$$K = \frac{f_{ace}}{(1 - \alpha_0)f_\alpha} \tag{4}$$

where f_{ace} is the common-emitter cutoff frequency.

Although the shape of the common-emitter cutoff is independent of the common-base cutoff characteristic, this independence does not hold for the constant K. Because of its dependence upon common base phase, K is determined by the impurity distribution in the base layer of the transistor. For instance, K will be unity for a transistor having a retarding field in the base layer of the right magnitude to give an RC type common-base cutoff defined by

$$\alpha = \frac{\alpha_0}{1 + j\dfrac{f}{f_\alpha}}. \tag{12}$$

K will be 0.82 for a uniform base layer impurity distribution having a common-base cutoff characteristic defined by

$$\alpha = \alpha_0 \operatorname{sech} \sqrt{j\frac{f}{0.822f_\alpha}}. \tag{13}$$

For "built-in fields" corresponding to a decrease in impurity concentration from emitter to collector, K will be less than 0.82 and will decrease as the built-in field increases.

Eq. (11) has been experimentally verified by many engineers who have measured the common-emitter attenuation and phase of a variety of junction transistors on transmission and phase sets of different types. A typical common-emitter current gain and phase characteristic for a diffused base germanium transistor is shown in Fig. 2. The solid curves show magnitude and phase characteristics mathematically computed for an RC cutoff having a 3-db point at 6.8 mc. The individual points plotted are the actual transistor gains and phases measured on a 20-mc precision gain-phase transmission measuring set. The measured magnitude is in agreement with the theoretical magnitude within the limits of experimental error. The measured phase is seen to be 45° at the 3-db cutoff frequency as it should be in accordance with (11). The agreement between measured and theoretical phase is good up to twice the common-emitter cutoff frequency. Above this the measured phase exceeds the phase given by (11). This is the proper direction of the departure from the RC phase at higher frequencies as will be shown later.

The transistor of Fig. 2 has an α_0 of 0.9833 and a common-base cutoff frequency of 540 mc. The constant K for this transistor is therefore given by

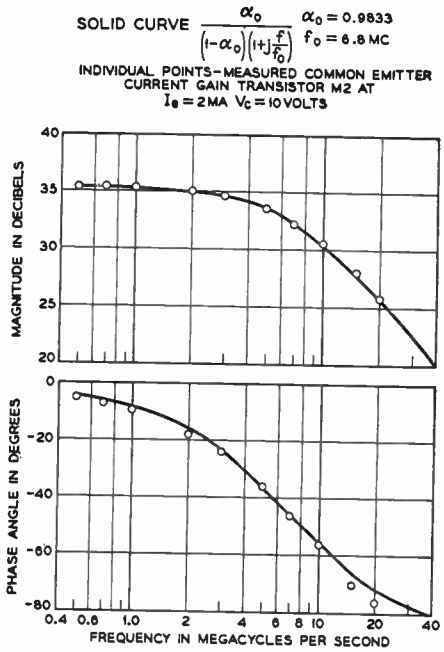


Fig. 2—Typical common-emitter current-gain magnitude and phase.

$$K = \frac{f_{ace}}{(1 - \alpha_0)f_\alpha} = \frac{6.8 \times 10^6}{(1 - 0.9883) \times 540 \times 10^6} = 0.75.$$

This transistor is a germanium diffused base transistor with a base layer impurity density which decreases between the emitter and collector. The K value of less than 0.82 is therefore in accordance with the dependence of K on base layer impurity distribution discussed above.

The first step in extending our current gain solutions above 0.1 f_α to the common-base cutoff frequency is the development of a generalized expression for the common-base current gain which is valid from dc to f_α. Referring to Fig. 3, the dotted curve shows the magnitude of the short-circuit current gain, referred to its low frequency value, of a transistor having a common-base current gain defined by the completely general expression

$$\alpha = |\alpha| e^{j\Psi} \tag{14}$$

where $|\alpha|$ and Ψ are as yet both undefined functions of frequency. The solid curve shows an RC cutoff defined by

$$\alpha_{RC} = \frac{\alpha_0}{1 + j\dfrac{f}{f_\alpha}} = |\alpha_{RC}| e^{-j \arctan f/f_\alpha}$$
$$= |\alpha_{RC}| e^{j\Psi_{RC}}. \tag{15}$$

In accordance with the observation that all transistors have a common-base short-circuit gain magnitude which is very close to an RC magnitude up to the 3-db loss frequency, f_α, the two curves have been drawn to

coincide at frequencies below f_α. The difference in gain magnitudes of the RC and the generalized transistor is given by

$$20 \log \left|\frac{\alpha}{\alpha_0}\right| - 20 \log \left|\frac{\alpha_{RC}}{\alpha_0}\right| = 20 \log \left|\frac{\alpha}{\alpha_{RC}}\right| \qquad (16)$$

which is also plotted on Fig. 3, and which, in accordance with our assumption is zero or close thereto below f_α.

The minimum phase associated with the magnitude variation of α will then be the sum of the minimum phase associated with the two magnitude characteristics,

$$20 \log \left|\frac{\alpha_{RC}}{\alpha_0}\right| \quad \text{and} \quad 20 \log \left|\frac{\alpha}{\alpha_{RC}}\right|,$$

which when summed give

$$20 \log \left|\frac{\alpha}{\alpha_0}\right|.$$

The minimum phase associated with $20 \log |\alpha_{RC}/\alpha_0|$ is given by Ψ_{RC} from (15). Since the magnitude of $20 \log |\alpha/\alpha_{RC}|$ is essentially 0 below f_α, and small immediately above f_α reaching appreciable magnitudes only at frequencies greater than $2f_\alpha$, then in accordance with the Bode theorem,[9] the minimum phase associated with this characteristic will be linear with frequency. Therefore, the phase Ψ of α can be written as

$$\Psi = \Psi_{RC} - C\frac{f}{f_\alpha}, \qquad f < f_\alpha. \qquad (17)$$

Since the magnitude of α follows the RC curve below f_α, α as a function of frequency can then be expressed as

$$\alpha = |\alpha_{RC}| e^{j(\Psi_{RC} - Cf/f_\alpha)} = \frac{\alpha_0}{1 + j\dfrac{f}{f_\alpha}} e^{j(-Cf/f_\alpha)}, \qquad f < f_\alpha. \qquad (18)$$

If the expression for α given in (18) is compared with that given in (10) at frequencies sufficiently low that $|\alpha| = \alpha_0$, we discover that the C of (18) is equal to $(1-K)/K$. We can therefore write (18) as

$$\alpha = \frac{\alpha_0}{1 + j\dfrac{f}{f_\alpha}} e^{j[(K-1)/K]f/f_\alpha} \qquad (1)$$

and from (17) we see that

$$\frac{K-1}{K} = -C = (\Psi - \Psi_{RC})_{[f=f_\alpha]} = \Psi_{[f=f_\alpha]} + \frac{\pi}{4} \qquad (3)$$

a result not surprising since we already knew the constant K to be a function of common-base phase shift.

Eq. (1) is a very close approximation to the common-base short-circuit current gain of any junction transistor

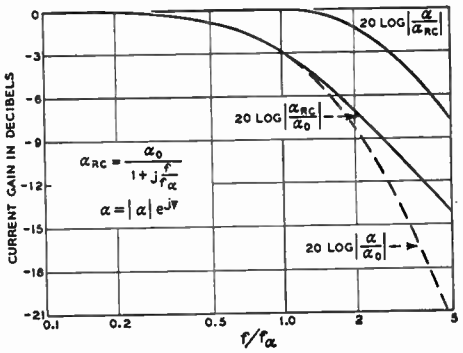


Fig. 3—RC and RC difference components of generalized common-base current gain.

falling within the limits of the analysis as defined in the introduction. As an example of its accuracy, consider a transistor having a hyperbolic secant (*i.e.*, a uniform base layer impurity distribution transistor) common-base cutoff frequency characteristic. Eq. (1) gives the magnitude of α for such a characteristic to within 0.01 db of true value at frequencies below f_α and to within 0.15 db at frequencies from f_α to $2f_\alpha$, and the phase of α to within 0.1° of the true phase at frequencies below f_α and to within 0.3° from f_α to $2f_\alpha$.

Now that we have a convenient and general expression for α extending from dc to f_α, we can examine the common-emitter cutoff at frequencies above the $0.1f_\alpha$ limitation of (11). To do this we will compare the magnitude and phase of the common-emitter current gain corresponding to the common-base gain of (1) to the common-emitter gain given by (11) with the $f < 0.1f_\alpha$ restriction removed. We will call this nonfrequency restricted common-emitter current gain, β_1. β_1 will therefore be given by

$$\beta_1 = \frac{\alpha_0}{(1-\alpha_0)\left[1 + j\dfrac{f}{K(1-\alpha_0)f_\alpha}\right]} \qquad (19)$$

which we will abbreviate as

$$\beta_1 = |\beta_1| e^{j\Phi_1}. \qquad (20)$$

And since it will be algebraically more convenient to compare reciprocals of common-emitter gain we will further define

$$\frac{1}{\beta_1} = \left|\frac{1}{\beta_1}\right| e^{j\theta_1} \qquad (21)$$

where

$$\theta_1 = -\Phi_1.$$

The common-emitter current gain corresponding to the common-base current gain given by (1) will be defined by

$$\beta = \frac{\alpha}{1-\alpha} = \left|\frac{\alpha}{1-\alpha}\right| e^{j\Phi} \qquad (22)$$

and the reciprocal of this common-emitter gain will be defined as

$$\frac{1}{\beta} = \left|\frac{1}{\beta}\right| e^{j\theta} = \frac{1}{\alpha} - 1 \tag{23}$$

where

$$\theta = -\Phi.$$

Substituting α of (1) in (23) we get

$$\frac{1}{\beta} = \frac{1}{\alpha_0}\left[\left(1 + j\frac{f}{f_\alpha}\right) e^{j[(1-K)/K]f/f_\alpha} - \alpha_0\right]. \tag{24}$$

If x is substituted for f/f_α and γ for

$$\left(\frac{1-K}{K}\right)\frac{f}{f_\alpha}$$

in (24) and the exponential term is written in complex form then,

$$\frac{\alpha_0}{\beta} = [(-\alpha_0 + \cos\gamma - x\sin\gamma) + j(\sin\gamma + x\cos\gamma)]. \tag{25}$$

If the square of the magnitude of α_0/β is next obtained in terms of the series expansion of $\sin\gamma$ and $\cos\gamma$ dropping terms containing powers of γ higher than the second we get

$$\left|\frac{\alpha_0}{\beta}\right|^2 = (1-\alpha_0)^2 + x^2 + 2\alpha_0 x\gamma + \alpha_0\gamma^2 \tag{26}$$

which is within 1/2 per cent of the true value for $K>0.8$ and $f<f_\alpha$. Again substituting x for f/f_α, $1/\beta_1$ can be written from (19) as

$$\frac{1}{\beta_1} = \frac{1}{\alpha_0}\left[(1-\alpha_0) + j\frac{x}{K}\right] \tag{27}$$

and

$$\left|\frac{\alpha_0}{\beta_1}\right|^2 = (1-\alpha_0)^2 + \frac{x^2}{K^2}. \tag{28}$$

Taking the ratio of $|\alpha_0/\beta|^2$ to $|\alpha_0/\beta_1|^2$ as given by (26) and (28), respectively, we get

$$\frac{\left|\frac{\alpha_0}{\beta}\right|^2}{\left|\frac{\alpha_0}{\beta_1}\right|^2} = \frac{|\beta_1|^2}{|\beta|^2} = 1 + \frac{(1-\alpha_0)\left(1-\frac{1}{K^2}\right)}{\frac{1}{K^2} + \left(\frac{1-\alpha_0}{x}\right)^2}. \tag{29}$$

The second term of (29) which represents the departure of $|\beta_1|^2/|\beta|^2$ from unity increases monotonically as $x=f/f_\alpha$ increases. Its maximum value for $f<f_\alpha$ will therefore be at f_α and will be given by $(1-\alpha_0)(K^2-1)$ which for $K>0.8$, and $\alpha_0>0.9$ will be less than 0.036. This means that $|\beta_1|$ and $|\beta|$ will be within 0.3 db of being the same for $K>0.8$, $\alpha_0>0.9$, and $f<f_\alpha$. Eq. (11) therefore gives the magnitude of common-emitter gain within 0.3 db at all frequencies below f_α.

Next compare the phases of the reciprocals of the common-emitter current gains given by (24) and (27) in the cutoff region, *i.e.*, $0.1f_\alpha<f<f_\alpha$. From (23) and (25) we get

$$\tan\theta = \frac{\sin\gamma + x\cos\gamma}{-\alpha_0 + \cos\gamma - x\sin\gamma}. \tag{30}$$

Since we are interested in the phase in the cutoff region where θ is close to $\pi/2$, we will consider the small angle $\pi/2-\theta$ whose tangent is the reciprocal of (30). Again expanding $\sin\gamma$ and $\cos\gamma$ and dropping terms containing powers of γ higher than the second we get

$$\tan\left(\frac{\pi}{2} - \theta\right) = \frac{(1-\alpha_0) - x\gamma - \gamma^2/2}{x + \gamma - x\gamma^2/2} \tag{31}$$

which is within 1 per cent of true value for $K>0.8$ and $f<f_\alpha$. Designating $(1-\alpha_0)$ by Δ, substituting x/K for $x+\gamma$, and taking the quotient of (31) retaining terms only up to and including the second power of x we get

$$\tan\left(\frac{\pi}{2} - \theta\right) \doteq \frac{\Delta K}{x} - K\gamma - \gamma^2 K/2x. \tag{32}$$

Anticipating that $\pi/2-\theta$ will be a small angle, we can write

$$\tan\left(\frac{\pi}{2} - \theta\right) \doteq \pi/2 - \theta \doteq \frac{\Delta K}{x} - K\gamma - \gamma^2 K/2x, \qquad (2\Delta < x < 1). \tag{33}$$

Using a similar approach, we get $\pi/2-\theta_1$ from (21) and (27) as

$$\pi/2 - \theta_1 \doteq \frac{\Delta K}{x} \qquad (2\Delta < x < 1) \tag{34}$$

and from (33) and (34)

$$\theta_1 - \theta \doteq -K\gamma - \gamma^2 K/2x$$

and replacing γ by $x(1-K)/K$,

$$\begin{aligned}\theta_1 - \theta &\doteq -(1-K)\,x\left[1 + 1/2\left(\frac{1-K}{K}\right)\right] \\ &\doteq -(1-K)\,x\left[1 + \frac{1-K}{K}\right]^{1/2} \\ &\doteq -\frac{(1-K)}{\sqrt{K}}\,x\end{aligned} \tag{35}$$

since $(1-K)/2K$ is an order of magnitude smaller than one.

Substituting $\Phi=-\theta$ and $\Phi_1=-\theta_1$ in (35) we get

$$\begin{aligned}\Phi &= \Phi_1 + \frac{K-1}{\sqrt{K}}\,x \\ &= \Phi_1 + \frac{K-1}{\sqrt{K}}\,\frac{f}{f_\alpha}.\end{aligned} \tag{36}$$

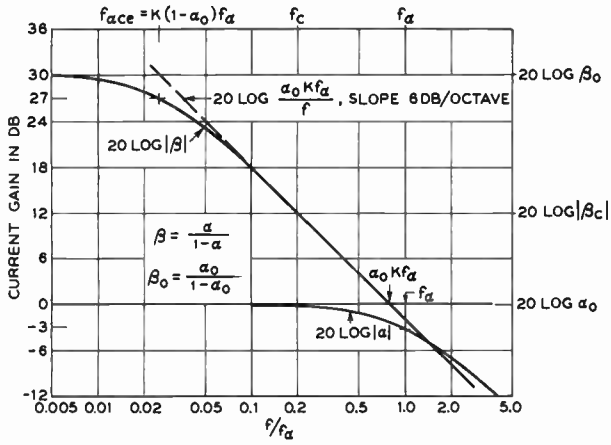


Fig. 4—Junction transistor current-gain magnitudes; 20 log $|\alpha|$, common-base short-circuit current gain; 20 log $|\beta|$, common-emitter short-circuit current gain.

The equality of $|\beta|$ and $|\beta_1|$ has already been established so that (19) for β_1 gives the magnitude of β. It also includes the Φ_1 portion of Φ as given by (36). Therefore, by adding an all-pass phase term to (19) which vanishes at low frequencies but which reflects the additional phase Φ of β over Φ_1 of β_1 in the common-emitter cutoff region, the following expression giving both the magnitude and phase of β is obtained.

$$\beta = \frac{\alpha_0 e^{j[(K-1)/\sqrt{K}]f/f_\alpha}}{(1-\alpha_0)\left[1 + j\dfrac{f}{K(1-\alpha_0)f_\alpha}\right]}, \quad f < f_\alpha. \tag{2}$$

Using (1) and (2) only three magnitude measurements are needed to determine the complete short-circuit current gain and phase characteristics, common emitter and common base from direct current to f_α. The values needed are α_0, the low frequency magnitude of α, f_α, the common-base current-gain cutoff frequency, and the magnitude of the common-emitter current gain at a single frequency f_c in the common-emitter 6 db per octave cutoff region. This is graphically illustrated in Fig. 4 where the magnitudes of the common-base and common-emitter short-circuit gains of a junction transistor having a low frequency common-emitter short-circuit current-gain of 30 db are shown as a function of frequency. The magnitude of the common-emitter current gain in the 6 db per octave cutoff region illustrated in Fig. 4 is given by (2) as

$$20 \log |\beta| = 20 \log \frac{\alpha_0 K f_\alpha}{f_c} \tag{37}$$

where f_c is any frequency meeting the requirement that

$$\left[\frac{f_c}{K(1-\alpha_0)f_\alpha}\right]^2 \gg 1.$$

The constant K can be determined from (37) and knowing f_α and α_0, the current gains can be computed from (1) and (2).

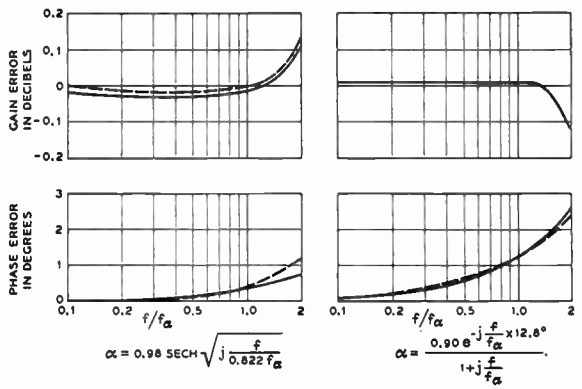


Fig. 5—Errors in the approximations of equations [(1) and (2)]; - - - common-base current-gain errors, —— common-emitter current-gain errors.

The frequency at which the common-emitter current gain is zero, ($|\beta| = 1$, 20 log $|\beta| = 0$, see Fig. 4) is of special interest. This will be defined as $f_{|\beta|=1}$ and is given from (37) as

$$f_{|\beta|=1} = \alpha_0 K f_\alpha. \tag{38}$$

Therefore, f_α, the common-base cutoff frequency, is given by the frequency of common-emitter zero gain divided by the product of K and α_0.

In order to illustrate the accuracy with which (1) and (2) give the common-base and common-emitter current gains, the common-base characteristic given by

$$\alpha = \alpha_0 \operatorname{sech} \sqrt{j\frac{f}{0.822 f_\alpha}}$$

was chosen and its short-circuit current-gains common base and common emitter were machine computed for $\alpha_0 = 0.98$ as a function of f/f_α up to $f/f_\alpha = 2$. The magnitude of the common-emitter current gain for $f/f_\alpha = 0.4$ was then taken from the machine computations as the single known gain in the common-emitter cutoff region, and with $\alpha_0 = 0.98$ the common-base and common-emitter current-gain magnitudes and phases were computed using the procedure just described. The departures of the values so determined from the true machine computed values are plotted as a function of f/f_α in Fig. 5(a). The maximum error in magnitude of current gain, either common base or common emitter, was only 0.02 db at frequencies below f_α and 1.4 db at frequencies up to $2f_\alpha$, and the maximum error in phase common base or common emitter was only 0.4° at frequencies below f_α and 1.1° at frequencies up to $2f_\alpha$. In order to test the approximations for the minimum value of α_0 considered in the development of (1) and (2), the current gains were computed from the three magnitude measurements for a transistor having an α given by

$$\alpha = \frac{0.90}{1 + j\dfrac{f}{f_\alpha}} e^{-j(f/f_\alpha)12.8°}. \tag{39}$$

The resultant errors in the approximations to the true common-emitter current-gain magnitude and the common-emitter and common-base phase are plotted in Fig. 5(b). Since the form of (39) is the same as our generalized expression for common-base current gain given by (1), the common-base magnitude determined from (1) will be exact. The estimated common-emitter gain is within 0.01 db of the true value at frequencies below f_α and within 1.2 db at frequencies up to $2f_\alpha$. The phases of the common-emitter and common-base current gains are within 1.3° of the true value at frequencies below f_α and within 2.6° of the true value at frequencies up to $2f_\alpha$.

As a more severe test of the reliability of (1) and (2), a transistor was assumed having an α_0 of 0.95 and a base layer impurity distribution which would produce a high built-in field and a correspondingly low value of K. The K value for the distribution assumed was 0.585 which is well below the 0.8 minimum K of the approximations.[11] However, even in this case, the estimated common-emitter and common-base gains were good to within 0.1 db below f_α and the estimated common-emitter phase was in error by only 5.6° at f_α. The estimated common-base phase was only out by 5.9° at f_α. For a higher α_0 of 0.98 where the approximation is somewhat better, the common-emitter and common-base gains were within 0.07 db of the true value at frequencies below f_α and the common-emitter-phase was within 3.8° of the true value at f_α. The common base phase was within 3.5° of the true value at f_α.

Because of the limitations of the analysis given in the Introduction, (1) and (2) are not applicable when there are appreciable cutoffs due to collector body resistance and base resistance or emitter resistance combined with collector junction capacitance. However, if as is usually so, these cutoffs have only a second order effect on the short-circuit current gains, their effect can be segregated in characterizing a given type of transistor.

One final comment should be made on the minimum vs nonminimum phase situation. Although (1) and (2) each include a phase shift linear with frequency and amplitude independent insofar as the equations are concerned, these phase shifts are not nonminimum transit time phase shifts but rather are due to a departure of the magnitudes of the true current gains at frequencies greater than f_α from the magnitudes given by these equations. However, if junction transistors were made which had appreciable nonminimum transit time common-base phase shift, we would expect this phase shift to be linear with frequency. In this case (1) would also accurately express the common-base current gains for these transistors. Consequently, (2) would give their common-emitter current gains providing the amplitude independent phase term of (1) falls within the limit set in the derivation of (1) and (2).

Acknowledgment

The authors wish to thank W. Kantrowitz for supplying the gain and phase data of Fig. 2, F. R. Stansel for the first experimental confirmation that the common-emitter phase at the common-emitter cutoff frequency was 45° over a wide range of operating conditions and transistors, and J. M. Klein for the many measurements of diffused base transistor magnitudes of common-base and common-emitter short-circuit current gains, the behavior of which encouraged the search for the generalized expressions for transistor short-circuit current gains derived in this paper.

[11] The data for this evaluation was supplied by C. A. Lee.

Correction

On page 739, in the second column, the superscript 1,2 should be deleted from the end of the second paragraph.

In (10) on page 750, the numerator a in the last part should be a_0.

Louis Weinberg, author of "Exact Ladder Network Design Using Low-Q Coils," which appeared on pages 739–750 of the April, 1958 issue of PROCEEDINGS, has requested that the following corrections be made to his paper.

Power Transistors*

M. A. CLARK†, SENIOR MEMBER, IRE

Summary—The objectives of power transistor design are expressed in general terms which make evident the reason for certain trends in the development. The design theory is discussed in terms of the physical phenomena which are believed to be important at high current densities and high voltages. The course of development over the past ten years is then described in terms of the important factors revealed by the design theory. A brief discussion of the present status of power transistor development is followed by speculation as to the probable future results of continued development.

Introduction

THIS paper deals with those transistors which have been designed and produced expressly for the purpose of delivering a specified output level, either power, current, or voltage, in addition to other objectives such as gain. This definition includes a range of transistors from those designed for high power at low audio frequencies to those designed to deliver, at the highest frequencies, the greatest power output of which the technology is capable, even though only milliwatts.

Now this is not so unreasonably broad a definition as it might appear at first. If power is a primary objective, the transistor must be designed to operate at high current densities, and this is also what is required of transistors which are to operate at high frequencies. Hence the high-frequency transistor intended to deliver a specified power output poses some of the same design problems as does the very much higher power audio-frequency transistor. In particular, the design must result in ample maximum current, maximum voltage, and heat transfer.

The specific problems associated with high frequencies are well covered elsewhere in this issue. Therefore, only those high-frequency design problems characteristic of power transistors are discussed. For the most part, only low frequencies are considered, for this is the area in which power transistors have found thus far the most extensive application.

During the past ten years, a wide variety of transistor structures has been evolved, and each structure has been exploited to some extent for the handling of substantial power. The history of this evolution, from the prejudiced viewpoint of a power transistor engineer, appears as a sequence of attempts to improve the power-handling capability of transistors. This history is presented from this point of view.

The design theory for power transistors consists of the original junction transistor theory of Shockley,[1] plus a collection of analyses attempting to avoid certain simplifying approximations, plus a large measure of empirical design. Attempts to improve upon the theory mostly take into consideration the two or three-dimensional nature of the structure and the effect of operation at high current densities. Such theories are both multidimensional and nonlinear, and for the most part do not consider simultaneously all of the departures from the simple theory that are important. Therefore, these theories serve as qualitative guides to the design and as such are very useful, although falling short of a precise, quantitative design theory.

A wide variety of methods for fabrication, of resulting structures, and of specific models, has resulted from the combined creative efforts of the semiconductor industry. No one method of fabrication has proven to be so superior in all respects as to displace all others. No single transistor structure is optimum for all applications. No single semiconductor material is best adapted to all uses and all environments. Engineers faced with the necessity of making decisions as to which fabrication method, structure, and material to use in constructing transistors are in the same kettle with the circuit engineer selecting a transistor to be designed into an electronic system which he hopes will evade obsolescence for at least a year or two. The fact that attention is concentrated here on power transistors does only a little to simplify these decisions.

The purpose of this article is to make as clear as possible the principles of power transistors, the factors influencing the above decisions, and the present state and future prospects for power transistors. Given some understanding of these generalities, the engineer may readily obtain detailed mathematical analyses, transistor specifications, and circuit designs from the large body of publications rapidly accumulating and from the excellent data sheets and applications notes available from manufacturers.

Applications

Most electronic systems have an output which requires substantial amounts of power. Commonly, some form of electromagnetic transducer is employed. In other cases the load may consist of an electrostatic transducer, or perhaps an antenna. Power transistors are prescribed for the output stage of such systems.

The power input is a second site for power transistors. In the power supply, transistors enter as inverters, converters, regulators, and controllers. It is natural that in transistorized systems the power-supply requirements can best be met with the use of transistors in the

* Original manuscript received by the IRE, March 13, 1958; revised manuscript received, April 11, 1958.

† Pacific Semiconductors, Inc., Culver City, Calif. Formerly at Bell Telephone Labs., Murray Hill, N. J.

[1] W. Shockley, "The theory of *p-n* junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

power supply, because the voltage and current levels are most appropriate for transistors.

A third natural application for power transistors is as switches, where they substitute for relays. Extensive use in this role may need to await further cost reduction, higher current and voltage ratings, and improved reliability. All of these requirements are within reach and, in many cases, already available to a satisfactory degree. The power transistor, of course, is capable of much faster operation than is a relay, and the power gain is comparable.

Many articles and a number of books have been written on transistor circuits. These contain circuits and circuit theory for power transistors. All that need be done here is to call the reader's attention to a few circuits which illustrate circuit features characteristic of the use of power transistors.

Four such elementary circuits are given in Fig. 1. The first is a simple power converter consisting of a transistor oscillator, with a transformer providing both the necessary feedback and the desired output voltage. Commonly, a rectifier and filter are used to provide a dc output. The transformer is built with a core material chosen to give square-wave operation of the transistor, and hence a very high efficiency, usually greater than 90 per cent for the complete circuit. The frequency of oscillation is made as high as possible without too severely lowering the efficiency, thereby simplifying the filter design. Push-pull circuitry is used if large power output is required.

Transistors make possible, in principle, a type of circuit not possible with tubes. This is the complementary symmetry circuit[2] shown in simple form in Fig. 1. By the use of transistors of opposite polarity type, the necessity for transformers is eliminated, while retaining the desirable features of a push-pull circuit with output transformer. In particular, even harmonics are cancelled and no direct current flows in the load. Furthermore, no input transformer or phase inverter is required, since a two-terminal input provides a push-pull drive of the transistors. A possible disadvantage is the need for a center-tapped power supply. This circuit has not been extensively used for the unfortunate reason that matched pairs of power transistors of opposite polarity have not been available in production quantities at a reasonable cost. Most manufacturers have concentrated on one type of power transistor and have not seen fit to attempt to produce both polarities, to say nothing of matched complementary pairs. Almost certainly, this situation is not permanent, and the accumulation of skill will make possible the alternative production of either type and the improved control needed to produce matched pairs of transistors. At that time, complementary symmetry circuits should prove valuable.

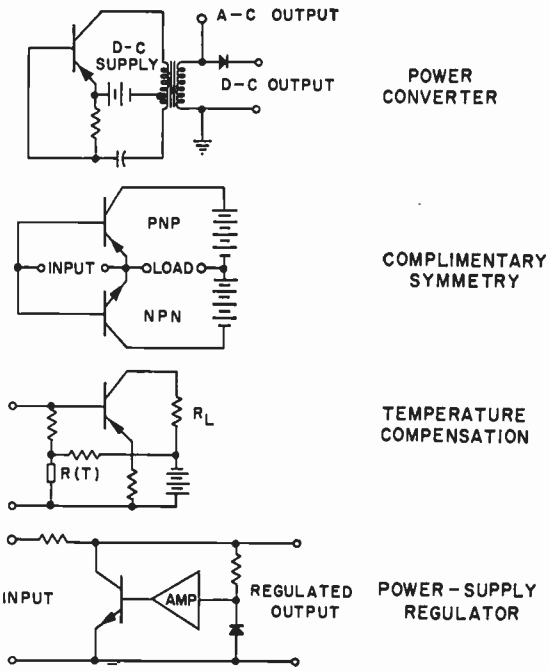


Fig. 1—Circuits characteristic of power transistor applications.

Just as the power requirement affects the design of the transistor, so it affects the design of the circuit. This is particularly true with regard to temperature variations because of the sensitivity of transistors to this variable. Intensive studies have been made of the stability of transistor circuits under conditions of internally generated heat and ambient temperature changes.[3] The principal transistor parameter subject to variation is the collector leakage current. Because this current increases exponentially with absolute temperature, and because the power dissipation may be due in significant part to the leakage, a run-away condition may occur, leading to circuit failure and possibly to the destruction of the transistor. This is much likelier to occur in some circuits than others. In general, common-base dc biasing is the most stable because, with this arrangement, the emitter current is controlled and run-away is prevented. However, the common-emitter circuit is more often used because of its advantages with regard to gain and impedance levels. This situation leads to circuit designs which are common emitter with regard to the signal but common base with regard to the input biasing. In addition, the use of temperature-sensitive elements, such as thermistors or semiconductor diodes, in the bias circuit is often recommended as a way to compensate for temperature variations. Such elements preferably are mounted adjacent to the transistor so as to follow closely the temperature of the transistor. A typical circuit arrangement is illustrated in Fig. 1.

Low-voltage power supplies, such as are needed for systems using transistors, are a natural place to use

[2] G. C. Sziklai, "Symmetrical properties of transistors and their applications," PROC. IRE, vol. 41, pp. 717–724; June, 1953.

[3] R. F. Shea, "Transistor operation: stabilization of operating points," PROC. IRE, vol. 40, pp. 1435–1437; November, 1952.

transistors.[4-6] In conjunction with low-power voltage-regulator diodes as voltage references, and a transistor dc amplifier if necessary, excellent regulation can be obtained with a good efficiency. The transistor regulator will function effectively with a very low voltage drop and is, therefore, nearly as efficient as is possible in such a system. Voltage regulators of both the series and shunt type are practical. Current regulators may also be designed which function very effectively over a limited but useful current range. A typical basic circuit for a voltage regulator is seen in Fig. 1.

Transistors make highly efficient rectifiers and as such may be used to provide regulation and control in a manner similar to thyratrons. The emitter and collector terminals may be used as rectifier terminals, and an ac bias provided on the base, such as to make the transistor conduct on one half cycle and cut off on the other half cycle. Because the voltage drop in the "on" state may be very low, lower than for a diode rectifier, the efficiency may be very high. In addition, by shifting the phase or modifying the waveform of the base bias current, the transistor may be caused to conduct only over a portion of the half cycle, so that the rectifier output may be controlled.

Finally, one circuit feature which must be mentioned in any such discussion as this is transient overload protection. Because power transistors operate at high currents and cannot withstand excessive voltage, the occurrence of destructive transients is probable unless precautions are taken against them. Ordinarily, such precautions take the form of careful design to avoid inductive surges, and the use of capacitors and diodes to absorb such surges as do occur.

Objectives of Power Transistor Design

Output Ability

Power output ability has two components, electrical and thermal. Electrical ability requires surviving and operating effectively at high voltages, and carrying high currents without failure or undue nonlinearity. Thermal ability is the capacity of the transistor to rid itself of the heat generated internally, and to operate at the elevated temperatures resulting from its own heat generation and from the ambient temperature. The specification of power ability, therefore, consists first of a statement of the maximum voltage, the maximum current, the thermal resistance, and the maximum temperature, all determined according to some agreeable conventions.

Power output ratings and deratings will be discussed later in connection with the packaging and mounting problems. At this point it is only necessary to call attention to the necessity of minimizing the temperature sensitivity of the transistor, by the choice of an appropriate semiconductor material and by careful design, and the necessity of minimizing the thermal resistance of the transistor and its associated heat sink.

The voltage rating of transistors has been no less arbitrary than the power rating. Collector junctions may exhibit a very sharp "breakdown" or a "soft" breakdown. By these terms it is meant that an oscilloscope presentation of the voltage-current characteristic reveals an abrupt increase in reverse current at a well-defined voltage or a gradual increase of current over a range of voltages. In the latter case, it is obvious that an arbitrary definition of breakdown is required. What is not so obvious is that the well-defined voltage characterizing the sharp breakdown may not be a useful rating, because for various reasons it may be necessary to restrict the voltage to lower values. Two obvious reasons for such derating are the need for a safety margin in the case of transients, and the necessity of setting a specification which can be met by a reasonable percentage of the transistors produced. The maximum voltage is commonly defined as the voltage at which a certain leakage current is caused to flow.

The current rating is no less of a definition problem. In general, three things may occur to limit the current which can be carried by the transistor. The first is the melting of an internal lead; this offers no very difficult rating problem. The second is a decrease in current gain too severe to be tolerable. For this problem, an arbitrary definition is necessary. Usually this takes the form of a specification that the current gain will not be less than a certain value at the maximum rated current. Third, the collector voltage required to have active transistor action may be so great at high currents that operation within the rated dissipation is not possible. In this case, at rated maximum collector current the specification may require that the collector voltage be no greater than a stated value, with a stated value of input current. This collector voltage is commonly referred to as a "saturation" voltage if it cannot be further reduced by increasing the input current. The ratio of this saturation voltage to the collector current at which it is measured is usually called the "collector saturation resistance." For some transistors, this usage is unfortunate because it implies, incorrectly, that the voltage drop is due to a resistance, when in fact nearly all of the voltage may be attributable to p-n junction phenomena. In other cases, the voltage may, in fact, be largely due to a parasitic resistance in the collector lead.

Gain

The power gain, current gain, and voltage gain may be significant individually or in combination. The power gain is, of course, the primary objective in most cases,

[4] F. H. Chase, B. H. Hamilton, and D. H. Smith, "Transistors and junction diodes in telephone power plants," *Bell Sys. Tech. J.*, vol. 33, pp. 827–858; July, 1954.

[5] S. Sherr, P. Levy, and T. Knapp, "Design procedures for semiconductor regulated power supplies," *Electronic Design*, vol. 5, pp. 22–25; April 15, 1957.

[6] R. D. Middlebrook, "Design of transistor regulated power supplies," Proc. IRE, vol. 45, pp. 1502–1509; November, 1957.

but current gain is often important for its own sake. For the purpose of discussion, the simplest useful equivalent circuit of a power transistor is probably that of Fig. 2. For low frequencies, the circuit elements of interest are the parasitic resistances r_b', r_c', and r_e', and the current generator aI_e. The emitter-junction resistance r_e is inversely proportional to the current, and for the high currents encountered in power transistors is often negligible. At high frequencies, the junction capacitances and the variation of a with frequency become important.

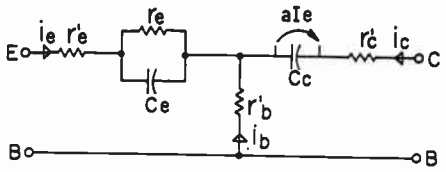


Fig. 2—Simple equivalent circuit for power transistors.

This equivalent circuit has seen wide use in describing the small-signal properties of transistors. It might be said that the object of the development of power transistors has been to make this circuit applicable to power transistors as well. In particular, it is usually assumed that the parameters of the circuit are constant with respect to the current and voltage variations caused by the signal being amplified. This, of course, is not true for large signals. The emitter resistance, r_e, is an unavoidable variable. The current amplification factor a is normally a very important variable also. Much of the power transistor development of the past few years has centered around attempts to make the value of a high and more constant with respect to variations of emitter current, and especially to prevent its value from decreasing too greatly at high currents.

The most widely used circuit configuration for power transistors is the common-emitter amplifier. The common-base circuit has the disadvantage of offering no current gain at low voltages, a serious matter for most power transistor applications. Although offering some advantage with respect to maximum collector voltage and frequency response, the common-base circuit has almost no power-transistor usage. The common-collector circuit has a higher input resistance than the common-emitter circuit, and a correspondingly lower power gain. Although often used in low-power stages for its high input resistance, the common-collector circuit is rarely used for power stages.

The current gain of a common-emitter power amplifier is given by

$$G_I = \beta = \frac{a}{1 - a}, \tag{1}$$

or by

$$G_I = B = \frac{A}{1 - A} \tag{2}$$

for the large-signal or dc gain. As has been mentioned, one objective of power transistor development has been to make these quantities more nearly equal.

The 3-db cutoff frequency of the common-emitter current gain is

$$f_{ce} = (1 - a)f_a \tag{3}$$

where f_a is the diffusion cutoff frequency of the current generator, aI_e.

The input resistance of a common-emitter power amplifier is approximately equal to the transistor base resistance, at least at high currents. At low currents, the input is nonlinear, due to the emitter resistance, or what is the same thing, the emitter-junction forward-voltage drop. The circuit designer needs to take account of this nonlinearity; the transistor designer has treated it as inevitable, and has concentrated only on the base-resistance component of the input resistance.

The power gain therefore is given approximately by

$$G_p = \left(\frac{a}{1 - a}\right)^2 \frac{R_L}{r_b'} = \beta^2 \frac{R_L}{r_b'}. \tag{4}$$

A definite limitation on the value of load resistance is imposed by the maximum rated voltage of the transistor. This voltage sets the maximum peak-to-peak output voltage, and for a given power output the load resistance cannot be greater than the value derived from the following relation for a Class-A amplifier:

$$P_0 = \frac{1}{8} \frac{V_m^2}{R_L}. \tag{5}$$

The frequency response of most power transistors is governed by the cutoff frequency of the current generator a rather than by capacitances, because the impedance levels are necessarily very low. This is not true for transistors designed for high-frequency use, nor for some of the newer types of power transistor having very good high-frequency response. The reader is referred to articles on high-frequency transistors for more detailed discussions of the design considerations pertaining to frequency response. In general, the design problems are the same, provided it is recognized that a power-output requirement sets a lower limit on the size of the transistor, because of the necessary heat transfer. For this reason, power output is an additional boundary condition on the high-frequency design, making the design more difficult.

Figures of Merit

The objectives of power transistor development can be understood better if the desirable characteristics are combined, insofar as possible, into figures of merit. This can be done as follows. Combining (4) and (5), and placing the performance parameters on one side of the new equation, gives

$$G_p P_0 = \left(\frac{a}{1-a}\right)^2 \frac{V_m^2}{8r_b'} . \quad (6)$$

It is apparent that the product of gain and power output is a measure of the performance of the transistor. In (6) this measure of performance is expressed in terms of certain characteristics of the transistor, which are fixed quantities for a given transistor and are independent of the circuit. However, these characteristics and the figure of merit will depend on the operating point. In particular, the load resistance no longer appears. Furthermore, the gain and power output may be exchanged, one for the other, by varying the load resistance, without changing the product of the gain and the power output. This assumes, of course, that the current gain does not vary.

Similarly, the high-frequency figure of merit may be obtained from (3) and (6), and is

$$G_p P_0 f^2 = a^2 f_a^2 \frac{V_m^2}{8r_b'} . \quad (7)$$

In this case the low-frequency current gain is unimportant and is replaced by the cutoff frequency, f_a. In this figure, three quantities are involved. As before, the gain and power output may be interchanged by choice of load resistance. In addition, the gain and bandwidth may be interchanged in the usual way, say by the use of feedback. So in this figure of merit, also, the performance parameters may be exchanged one for the other, and their product is a function only of the transistor characteristics. Such a product characterizes the transistor in a most convenient manner for the present purpose, which is to examine developmental objectives.

Summary of Objectives

The quantities of primary importance are the low-frequency current gain, the maximum rated voltage, the base resistance, and the cutoff frequency f_a. At high frequencies, the collector and emitter capacitances and other effects must be considered. The parasitic resistances r_e' and r_c' play an obvious role in the input and output circuit and are of great importance in some transistor structures.

What is desired of the power transistor development is high current gain with a high cutoff frequency, high collector-voltage rating, low parasitic resistances, and low capacitances. The variation of these properties with operating and environmental conditions is of vital importance.

Design Theory

The theory of junction transistors given by Shockley contained certain simplifying assumptions. It has proven to be quite difficult to improve upon those approximations in a self-consistent manner. The tendency has been to treat one approximation at a time rather than to undertake the chore of making the necessary computations involved in a "perfect" theory. For power transistors, the important approximations made by Shockley were: 1) that the transistor is one dimensional with no transverse electric fields and no surface effects, 2) that the electric field is zero everywhere except in the space-charge layers at the junctions, 3) that the emitted minority-carrier concentration is much less than the majority-carrier concentration, and 4) that the net space charge and electric field are zero except in the space-charge layers at the junctions.

Considering the equivalent circuit of Fig. 2, these assumptions imply that the parasitic resistances are zero, since there are supposed to be no electric fields in the regions which contribute these resistances. Of course, the parasitic resistances were never actually denied. The emitter and collector parasitic resistances usually are easily calculated because these regions are essentially one dimensional in most transistors. The base resistance is not so simple because it has a distributed nature. The base current originates in the base layer at recombination centers, at the emitter as reverse emission, and at recombination centers on the surface. Furthermore, a two or three-dimensional geometry is involved and the resistivity of the base layer is affected by conductivity modulation.

Conductivity modulation results whenever minority carriers are generated in a semiconductor in a concentration comparable to that of the majority carriers. Space charge neutrality requires that the number of majority carriers increase by the same amount and, therefore, each additional minority carrier adds a total of two free carriers to the material and thereby increases the conductivity. Conductivity modulation occurs in the base layer of a transistor at high emitter current densities. This not only makes the computation of base resistance difficult but also complicates the flow of emitted carriers to the collector and interferes with the emission itself.

It is outside the scope of this paper to present the mathematical analyses necessary to approach an accurate quantitative understanding of these effects. The attempt here is to summarize the pertinent conclusions of the low-level theory and to give some qualitative understanding of what is taking place at the high voltages and high current densities encountered in power transistors.

Simplified Low-Level, One-Dimensional Theory

The low-frequency current gain ought to increase as the base layer is made thinner, because both the emitter efficiency and the transport efficiency are higher for a thin base layer. The current gain should not be dependent on the current level nor on the voltage. The common-emitter current gain is determined approximately by[7]

[7] W. Shockley, M. Sparks, and G. K. Teal, "*P-N* junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; July 1, 1951.

$$\frac{1}{h_{fe}} = \frac{1}{2}\left(\frac{w_b}{L_b}\right)^2 + \frac{\rho_e w_b}{\rho_b L_e} \tag{8}$$

where

$h_{fe} = \beta = B = a/(1-a) = A/(1-A)$,
w_b = base-layer thickness,
L_b = minority-carrier diffusion length in the base layer,
L_e = minority-carrier diffusion length in the emitter,
ρ_b = resistivity of the base layer, and
ρ_e = resistivity of the emitter.

Although this simple theory makes the small-signal and large-signal current-gains equal, such is not the case in practice. Fig. 3 shows the difference usually observed.

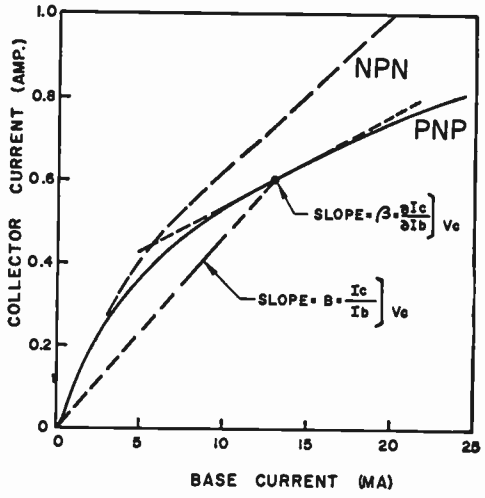


Fig. 3—Large-signal, common-emitter, current transfer characteristic.

Because of the nature of the diffusion process by which the minority carriers are transported, the cutoff frequency of the current generation, aI_e, should be dependent on the base-layer width as follows:[8]

$$f_a = \frac{1.2 D_b}{\pi w_b^2} \tag{9}$$

where D_b is the diffusion coefficient for minority-carriers in the base layer.

The common-emitter cutoff frequency, given by f_a/h_{fe}, is therefore not strongly dependent on the base-layer width; in fact, if the emitter efficiency is very high, this frequency is independent of the base-layer width. It is partly for this reason that the transistor designer prefers not to deal with this frequency. The circuit-theory arguments for and against the use of the common-emitter frequency in transistor specifications need not concern us here, since one may compute the cutoff of a from the cutoff of β and vice versa.

Both the current gain and the cutoff frequency theory must be modified somewhat from that given by Shockley if there is a nonuniform distribution of donors and acceptors in the base layer. The theory of the transistor may still be a low-level, one-dimensional theory, but electric fields exist outside the space-charge layer, and the minority carriers do not move by diffusion alone but also drift in the field. This condition occurs to some degree in all transistor structures with the exception of the alloy transistor. The diffused-base transistor has this effect to a marked degree. A general name for transistors having a significant, helpful electric field in the base layer is "drift transistor."

Simplified Two-Dimensional Theory

The base resistance can only be treated in a two or three-dimensional model of the transistor because it concerns the flow of current at right angles to the principal axis of the transistor. If the base current is generated uniformly throughout the base layer and the transistor is treated as two dimensional, and with a base layer, $a \times b$, of thickness w_b, having a base contact along side a of the rectangle, the base resistance is given by[9]

$$r_b' = \frac{1}{3}\,\frac{b}{a}\,\frac{\rho_b}{w_b}. \tag{10}$$

The voltage drop in the base layer is of special interest, since the internal base voltage must depend on distance from the base contact. The simple model used here results in a voltage which varies with distance x from the base contact as follows:

$$V_{bx} = I_b r_b' \frac{3}{2}\left(2 - \frac{x}{b}\right)\frac{x}{b}. \tag{11}$$

Because this base voltage governs the bias on the emitter junction, it is apparent that the emitter current density will vary over the face of the emitter and that the assumption of uniform base-current generation is not valid if the transverse voltage drop in the base layer is comparable to kT/q, about 0.026 volt. Hence, serious problems arise, even in the simplified theory, as soon as the transverse base current is considered. This will be reviewed again after other complications have been described.

Conclusions from the Simplified Theory

From the theory just sketched, certain conclusions may be drawn which remain valid in a more sophisticated analysis. For a high current-gain, the base layer should be thin and the minority-carrier diffusion coefficient, or mobility, in the base layer should be high. For a low base resistance, the base layer should be thick

[8] R. L. Pritchard, "Frequency variations of current-amplification factor for junction transistors," PROC. IRE, vol. 40, pp. 1476–1481; November, 1952.

[9] R. L. Pritchard and W. N. Coffey, "Small-signal parameters of grown-junction transistors at high frequencies," 1954 IRE CONVENTION RECORD, pt. 3, pp. 89–98.

and of high conductivity and a geometry should be used which brings the base contact close to all parts of the emitter. The conflict between the requirements for a thin base layer and a thick base layer are resolved in favor of a thin layer because of the dominant role played by the current gain and cutoff frequency in the circuit performance, as indicated by the figures of merit.

High Current Density Effects

At high current densities, the assumption is not valid that the minority-carrier concentration is much less than the majority-carrier concentration. Furthermore, the electric fields necessary to the flow of current in the emitter and collector region, and for the base current in the base layer, may affect the minority carriers. In the simple theory, the minority carriers are assumed to move only by diffusion.

One of the first observations of failure of the simple theory occurred when it was discovered that the current gain of early grown-junction transistors would increase at high emitter currents and might make the common-emitter circuit unstable. This was determined to be due to the fact that the field in the collector, moving majority carriers *away* from the collector junction, was increasing the flow of minority carriers *to* the junction from the collector region. That is to say, the collector saturation current, normally attributed to diffusion, was being increased by the electric field. At an elevated temperature, the saturation current is quite appreciable, and because the field enhancement of this current is roughly proportional to the total current, the effect is to increase the current gain of the transistor. The term "alpha crowding" has since been adopted to describe the appearance of the static characteristics if the current gain decreases at high currents, and this opposite effect in grown-junction transistors would be called "alpha spreading." The terms "alpha multiplication" and "collector multiplication" have been used also, but are best reserved for a high-voltage phenomenon to be mentioned later.

Alloy transistors have a very high conductivity collector region; therefore, the electric field in that region is truly negligible and high-current alpha spreading does not occur in such transistors. On the contrary, the current gain decreases at high currents, as shown by Figs. 3 and 4. Several explanations for this are possible. The following discussion treats one phenomenon at a time, although it is obvious that a precise theory must take all into account simultaneously.

Eq. (8), for the current gain given by the simple theory, suggests several possible sources of gain variation with current density. These depend upon the fact that the material properties, L and ρ, may be dependent upon the concentration of minority carriers. The concentration of those carriers may be estimated, using the assumptions that the flow through the base layer is due to diffusion and that the base layer is very thin. The

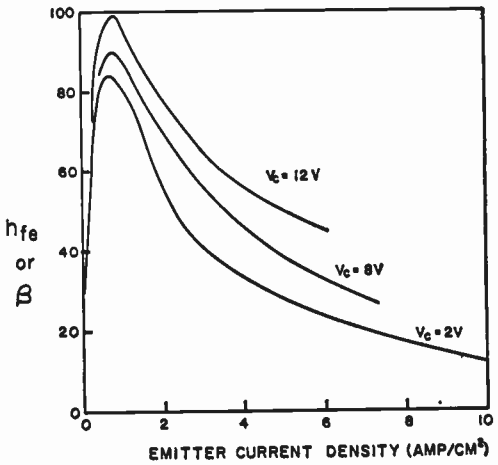


Fig. 4—Variation of small-signal, common-emitter current amplification factor with emitter current.

concentration of emitted minority carriers is then given by

$$n = \frac{w_b}{qD_n} J_e \tag{12}$$

where, of course, p may be substituted for n if the transistor is a p-n-p type rather than an n-p-n, and J_e is the emitter current density. The minority-carrier concentration will be equal to the majority-carrier concentration in the base layer at the following values of emitter current density:[10,11]

$$J_e = \frac{1}{2} \frac{kT}{q} \frac{1}{w_b \rho_b} \quad \text{for a } p\text{-}n\text{-}p \text{ transistor,}$$

$$J_e = 2 \frac{kT}{q} \frac{1}{w_b \rho_b} \quad \text{for an } n\text{-}p\text{-}n \text{ transistor.} \tag{13}$$

The constant *2* is the approximate value of the ratio of electron mobility to hole mobility. From these relations it would appear that the n-p-n transistor should be less susceptible to effects stemming from high current density. While there is considerable experimental evidence in agreement with this, as indicated in Fig. 3, one should not forget that the two transistor types are inherently different in other ways, due to the differences in fabrication, so that a controlled experiment to compare the two types is difficult to perform.

These equations suggest that a design to reduce this effect of high current density would aim for a thin base layer and a low-resistivity base layer. Note, however, that according to (8), lowering base-layer resistivity alone would reduce the emitter efficiency and thereby lower the current gain, a limitation which is encountered in diffused-base transistors and similar transistors.

[10] W. M. Webster, "On the variation of junction-transistor current amplification factor with emitter current," PROC. IRE, vol. 42, pp. 914–920; June, 1954.

[11] E. S. Rittner, "Extension of the theory of the junction transistor," *Phys. Rev.*, vol. 94, pp. 1161–1171; June 1, 1954.

The effect of the distributed base resistance in causing the internal base-layer voltage to vary with distance from the base contact has already been mentioned. Because of this effect, the emitter current will tend to be concentrated near the base contact. Therefore the emitter current density near the base contact is actually higher than indicated by (12), and the effect is more serious. On the other hand, one consequence of the high concentration of minority carriers is a decrease in the base-layer resistivity, thereby reducing the base resistance and the transverse biasing of the base layer.

Now, what effect will the high minority-carrier concentration in the base layer have on the current gain? We look for parameters in (8) which are affected by this concentration. One which is clearly subject to variation is the base-layer resistivity, whose conductivity modulation has been mentioned already. The lowered resistivity of the base layer will apparently result in a lower emitter efficiency. A completely satisfying quantitative experimental proof of this explanation for alpha crowding is difficult to accomplish because of the large number of poorly controlled variables and the complexity of the situation. Nevertheless, design improvements based on this theory have proven to be profitable.

To compensate for the decreased resistivity, (8) shows that several actions may be taken. One possibility is to decrease the resistivity of the emitter,[12,13] another is to make the base layer thinner; both of these design improvements have been used. The third possibility, increasing the diffusion length in the emitter, seems difficult to accomplish. Even the measurement of this quantity poses a formidable experimental problem.

The high minority-carrier concentration has another important effect on the transistor action. Because of the very strong tendency to maintain space-charge neutrality, the base-layer majority-carrier concentration increases near the emitter. These carriers then also have a concentration gradient and diffuse toward the collector, where they can only be reflected. Of necessity, there exists an electric field just sufficient to prevent the flow of the majority carriers. This field then *aids* the flow of minority carriers. The condition is one of ambipolar diffusion, just as sometimes occurs in a gas discharge.

The aiding field should increase the effective diffusion coefficient and thereby increase both the current gain and the diffusion cutoff frequency. The current gain and cutoff frequency, in fact, do increase as the emitter current is increased from zero [see Figs. 3 and 5(a)]. The aiding field may be a satisfactory explanation of this, at least for some types of transistors. Unfortunately, at a fairly low current, both the gain and the cutoff frequency reach a maximum; they then decrease

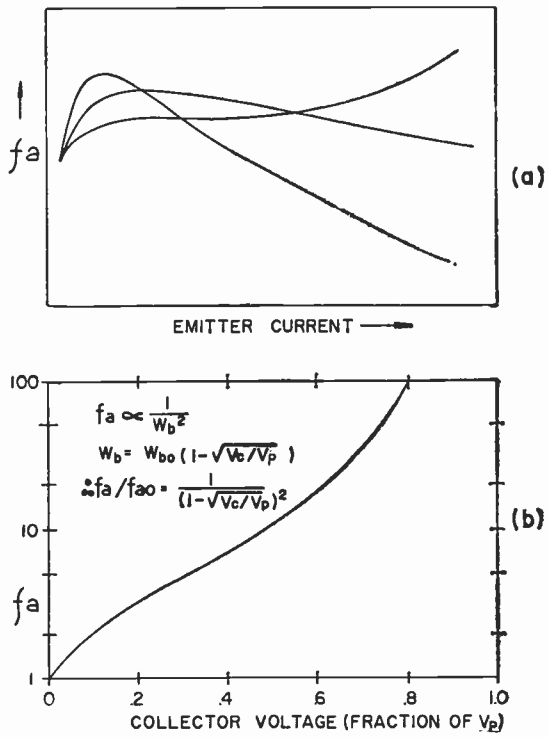


Fig. 5—(a) Variation of f_a with emitter current. (b) Theoretical variation of f_a with the collector voltage for alloy transistors.

at higher currents. Exceptions have been observed, as also shown in Fig. 5(a). If the rise in cutoff frequency is associated with the increase in minority-carrier concentration and the accompanying field, it is apparent that some other mechanism must be sought to explain the fall in cutoff frequency. Furthermore, the possibility must be considered that the fall in current gain is also due in part to this new mechanism even though the previous explanation, based on lowered emitter efficiency, has been shown to be sound.

The cutoff frequency, given by (9), is dependent only on the diffusion coefficient and the base-layer thickness. The possibility of an electric field in the base layer, modifying the diffusion, has been considered. A second source of variation in cutoff frequency must now be examined. This is the base-layer thickness. Variations in the effective base-layer thickness can be due to changes in the distribution of the emitter current, with a base layer of varying geometrical thickness. Such changes in emitter current distribution have already been mentioned. The geometrical thickness of the base layer may vary greatly in some transistor structures. The alloy transistor is perhaps the best example to consider.[14,15] In Fig. 6, a pictorial drawing is given which portrays events in an alloy transistor. The hole current is concentrated near the emitter edge. Electrons flow from the

[12] F. H. Stietljes and L. J. Tummers, "Behavior of the transistor at high current densities," *Philips Tech. Rev.*, vol. 18, pp. 61–68.

[13] L. D. Armstrong, C. L. Carson, and M. Bentivegna, "*P-N-P* transistors using high-emitter-efficiency alloy materials," *RCA Rev.*, vol. 27, pp. 37–45; March, 1956.

[14] A. R. Moore and J. L. Pankove, "The effect of junction shape and surface recombination on transistor current gain," PROC. IRE, vol. 42, pp. 907–913; June, 1954.

[15] K. F. Stripp and A. R. Moore, "The effects of junction shape and surface recombination on transistor current gain—Part II," PROC. IRE, vol. 43, pp. 856–866; July, 1955.

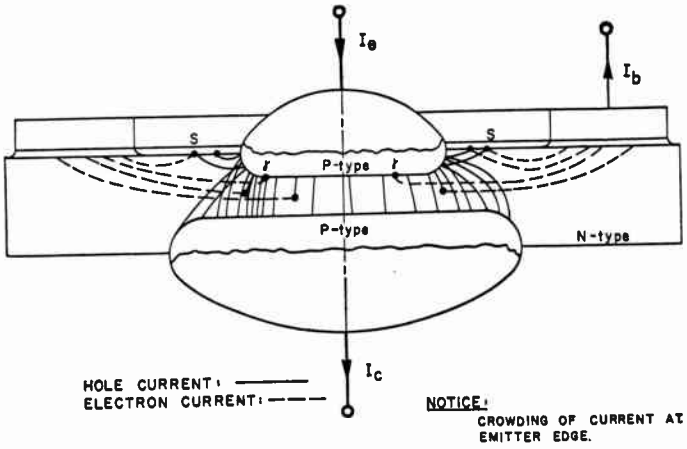


Fig. 6—Hole and electron flow in an alloy-junction transistor.

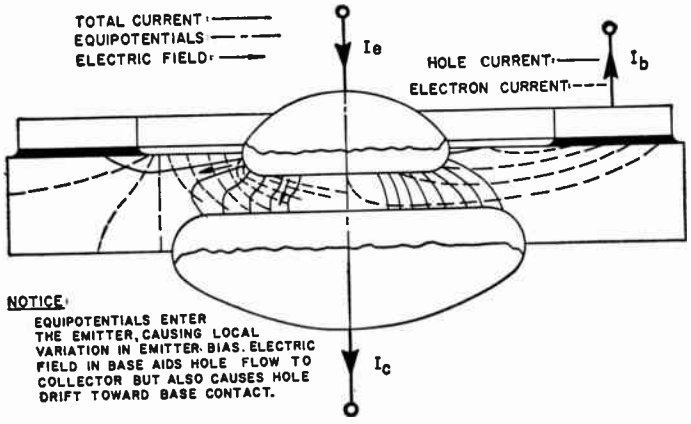


Fig. 7—Current and potential distribution in an alloy-junction transistor.

base contact to the emitter surface, where they enter the emitter (reverse emission), to the interior of the base layer where they combine with some of the holes (transport defect), and to the surface where they combine with holes (surface recombination).

It is immediately apparent from Fig. 6 that long diffusion paths are a real possibility and that the concentration of the emitter current at the edge of the emitter tends to produce such long paths. In addition, surface recombination enters the picture as a contributor to the loss of emitted carriers[10] and apparently the tendency to concentrate current at the emitter edge increases the likelihood of such losses.

In Fig. 7 an attempt has been made to depict the flow of hole and electron current on the right side of the drawing and the total current, and potential distribution, on the left side of the drawing. Here a new effect appears; there is an electric field tending to draw the holes away from the emitter-collector region and toward the base contact. This, of course, is the same field that is necessary for the transport of the base current.

For the simple rectangular model used previously in the calculation of base resistance, the field at the edge of the base layer would be

$$E = \frac{1}{a}\frac{\rho_b}{w_b} I_b. \tag{14}$$

This field will have a serious effect on the flow of minority carriers only when it produces a drift of carriers comparable to that which is due to the concentration gradient and accompanying diffusion. This will occur when

$$I_b = \frac{kT}{q}\frac{a}{\rho_b}. \tag{15}$$

This implies that, for an alloy power transistor with an emitter diameter of 0.120 inch and a base-layer resistivity of 3 ohm-cm, the field would be significant at a base current of only 8 ma. Apparently this might be a serious matter, because the field aggravates the tendency of the current to concentrate at the edge of the emitter, where it finds long paths, and would thereby increase the effective base-layer thickness. Examination of (8) and (9) shows that the base-layer thickness plays a dominant role in the transistor performance.

To reduce the field near the edge of the emitter, the designer uses the lowest possible base-layer resistivity and lengthens the emitter edge, a, to reduce the base current density. These are also the design features needed to make the base resistance low.

High Voltage Effects

A variety of phenomena limit the voltage that may be used on a transistor. The surface breakdown dominates in most germanium power transistors. This effect is the source of the practical limit, between 30 and 100 volts, which is usual for such transistors. Silicon transistors are capable of much higher surface breakdown voltages, presumably because of silicon's more stable surface. A satisfactory design theory and structural control does not exist for surface breakdown; it is dealt with in practice by following various recipes for surface preparation and then rating the transistor in accordance with the breakdown voltage that is observed.

The present discussion is confined to the design theory for the two bulk phenomena limiting the voltage. These are "punch-through" and "carrier multiplication."

Electrical punch-through is, in itself, a nondestructive effect which occurs when the collector space-charge layer extends entirely through the base-layer to the emitter junction, as illustrated in Fig. 8. This variation of the effective base-layer thickness has a strong effect on the transistor parameters.[16] The theoretical dependence of f_α on collector voltage for an alloy transistor is as given in Fig. 5(b). When complete penetration occurs, the barrier is lowered at the emitter and the emission is greatly increased, in a manner similar to

[16] J. M. Early, "Effects of space-charge layer widening in junction transistors," PROC. IRE, vol. 40, pp. 1401–1406; November, 1952.

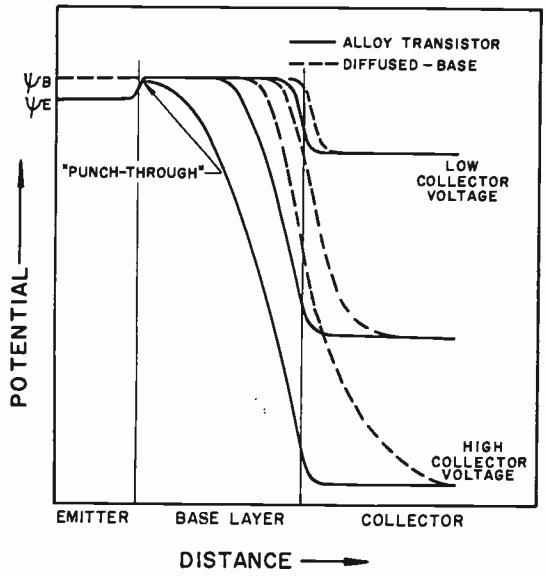


Fig. 8—Punch-through phenomena.

field emission at the cathode of a vacuum tube. The result is an effective short-circuit from emitter to collector, at voltages exceeding the punch-through voltage V_p; and the transistor is inoperative above this voltage. The effect of this voltage limit on an alloy transistor is to make the product $(V_p\rho_i f_a)$ a constant, whose value is dependent only on the semiconductor material. The value of the constant ranges from 1800 to 9600 for n and p-type silicon and germanium, with units of volt-ohm-cm-mc.

Other types of transistor generally have either a graded-collector junction, a high-resistivity collector, or both. Such transistors are not so troubled by punch-through because the collector space-charge layer is free to extend into the collector region, as indicated by the dashed lines in Fig. 8, and need not penetrate the base layer so far.

The second well-understood voltage limit is that which results from the multiplication of carriers in the collector space-charge layer.[17–19] The carriers normally collected by the potential drop in that region are accelerated to a high velocity by the high electric field. As a consequence, they acquire sufficient kinetic energy to collide with atoms in the crystal and dislodge electrons from such atoms, thereby creating electron-hole pairs. This process is seen to be similar to the ionizing collisions which occur in a gas discharge. The factor by which the current is multiplied is given by the empirical expression[17–19]

$$M = \frac{1}{1 - (V_c/V_B)^n} \tag{16}$$

[17] S. L. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, vol. 99, pp. 1234–1241; August 15, 1955.
[18] S. L. Miller and J. J. Ebers, "Alloyed junction avalanche transistors," *Bell Sys. Tech. J.*, vol. 34, pp. 883–902; September, 1955.
[19] S. L. Miller, "Ionization rates for holes and electrons in silicon," *Phys. Rev.*, vol. 105, pp. 1246–1249; February 15, 1957.

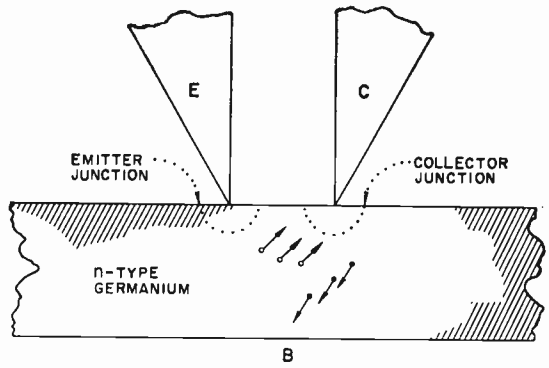


Fig. 9—Point-contact transistor. (Drawing taken from J. A. Morton, Proc IRE, November, 1952.)

where V_c is the collector voltage, V_B is the breakdown voltage, and n has values from 3 to 9. The breakdown voltage is that voltage at which the ionization creates an avalanche condition. Above this voltage, large collector current will flow with no emitter current; this is the p-n junction breakdown voltage. This voltage is roughly proportional to the resistivity of the high-resistivity side of the collector junction. It is apparent that the punch-through voltage and breakdown voltage make opposite demands on the base-layer resistivity of an alloy transistor. This makes impossible certain designs which would incorporate a high cutoff frequency with a high maximum voltage. Transistor structures which separate these effects by having a high collector resistivity are able to provide the desirable combination of a thin base layer and high maximum voltage.

The History of Power Transistor Development

This is primarily a history of methods of fabricating transistors of the conventional (n-p-n or p-n-p) sort. In addition, other semiconductor devices which might serve the functions of power transistors must be considered. These are treated only briefly since they have not as yet played a large role.

Point-Contact Transistors

By their very nature, these transistors are restricted to low power and low current. This is because the emitter and collector contacts consist of very fine wire points pressed against the semiconductor, as shown in Fig. 9. Such points would overheat if required to carry a very large current. In addition, the generation of heat occurs in a very small volume, with the result that an excessive temperature rise may be produced and the transistor fail to operate or be permanently damaged by overheating.

However, these alone are not sufficient reasons for discarding point-contact transistors as power transistors. If the frequency is very high, a power-output requirement of 100 mw may put the transistor into the power category and such a power output is within the ability of point-contact transistors.

The point-contact transistor, in fact, has been discarded insofar as further development or production for new applications is concerned. Obviously, this is not simply because it is a poor power transistor, but this is the only aspect of concern here. The point-contact transistor is deficient in the structural qualities which have been shown to be necessary for a power transistor. In particular, the base layer is determined by the point spacing and by a poorly understood process called "forming," by means of which the collector junction is produced. Neither these processes nor the "side-by-side" configuration of the emitter and collector lend themselves to the production of a structure with a thin, planar, base layer. Furthermore, the collector voltage is usually limited to low values, or the base resistance is too high, depending on the choice of material resistivity. Finally, the inadequate understanding of these transistors has caused attention to be turned into more scientific approaches which have in fact led to superior methods of constructing power transistors. It appears likely that the point-contact transistor has played its role to completion and need not receive further attention as a power transistor.

Grown-Junction Transistors

The development of transistors having large-area p-n junctions made possible the first power transistors in the narrow, obvious sense of the term. One of the first junction transistors was reported to be capable of a Class A power output of 2 watts. This transistor is made by the method of adding an acceptor impurity to the melt as an n-type crystal is being drawn from that melt, and quickly following with the addition of a large amount of donor element. The result is a high-resistivity n-type collector region, a low-resistivity p-type base layer, and a very low-resistivity n-type emitter.

The configuration of a grown-junction power transistor resembles Fig. 10. Examination of this sketch reveals certain of the difficulties which plague this transistor. For one thing, the base contact is not very effective if made as shown. This is the simplest way to make this contact but the result is a rather high base resistance. A second problem is that it is mechanically difficult to make the collector very short. Because the collector resistivity is high, this thick region will give a high collector parasitic resistance, r_c', as well as high thermal resistance.

Some work has been done to improve the grown-junction power transistor. It is possible to fabricate a base contact which extends along the base layer, and thereby reduce the base resistance. The techniques of producing grown junctions have advanced greatly. By the use of remelt and diffusion methods, grown-junction transistors may be made with a low collector resistivity, thereby reducing the collector parasitic resistance. This results in the same voltage problems which trouble the alloy transistor design, but the fabrication of a very thin base layer by such means may be easier than alloying.

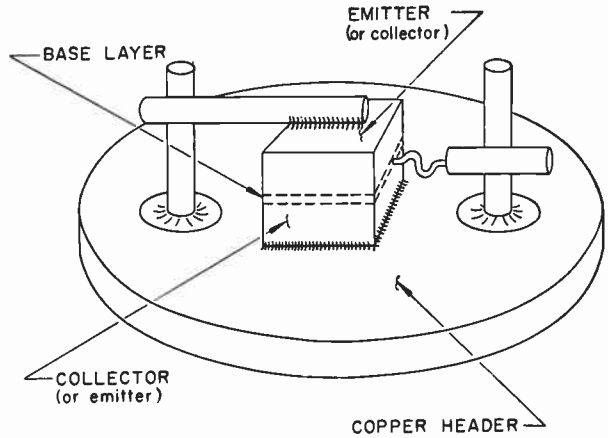


Fig. 10—Grown-junction power transistor.

Grown-junction power transistors appear to be rather complex to build if good performance is necessary. Although good performance can indeed be achieved, it is natural to seek fabrication methods which lend themselves more readily to the particular goals of power transistor design.

Alloy-Junction Power Transistors

Alloyed junctions offered immediate hope for the fabrication of high-power transistors. This transistor has inherently low parasitic resistances in the emitter and collector leads, and may readily be mounted with low thermal resistance. Furthermore, the base layer extends out where a good, extensive contact can be made. Alloying also offered the possibility of building elaborate structures which could not readily be made by growing junctions.

The basic configuration of a modern alloy power transistor is shown in Fig. 11. The collector button is commonly attached to the header, on the basis that the heat is generated at the collector junction. Actually, the emitter button is so very close to the collector that the thermal resistance would not be greatly increased by attaching the emitter to the header.

Shortly after the appearance of alloy transistors, the possibility of improved performance by means of a more elaborate structure was realized by Hall.[20] The structure he built is shown in Fig. 12. This has the desirable features of narrow emitter stripes and an extensive base contact in close proximity to the emitter. However, it does not have a thin base layer. A different arrangement of the electrodes was described by Fletcher[21–24] in an

[20] R. N. Hall, "Power rectifiers and transistors," PROC. IRE, vol. 40, pp. 1512–1518; November, 1952.
[21] N. H. Fletcher, "Some aspects of the design of power transistors," PROC. IRE, vol. 43, pp. 551–559; May, 1955.
[22] ——, "Self-bias cutoff effect in power transistors," PROC. IRE, vol. 43, p. 1669; November, 1955.
[23] ——, "A junction transistor of kilowatt pulses," PROC. IRE, vol. 45, p. 544; April, 1957.
[24] ——, "The high current limit for semiconductor devices," PROC. IRE, vol. 45, pp. 862–872; June, 1957.

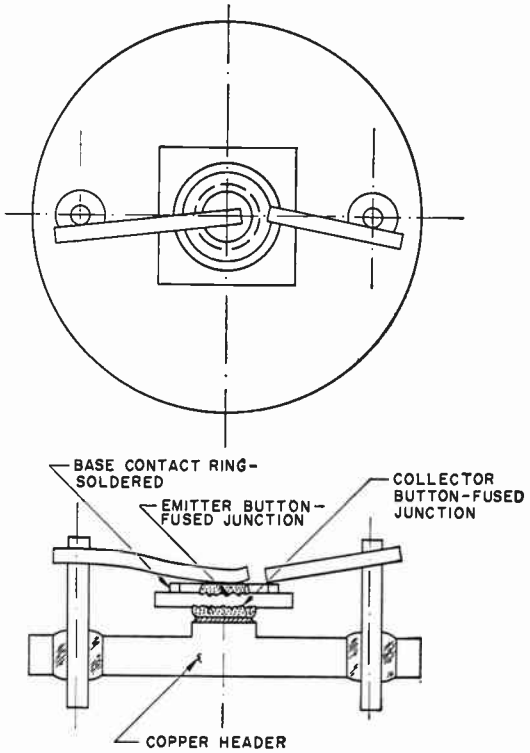


Fig. 11—Alloy-junction power transistor.

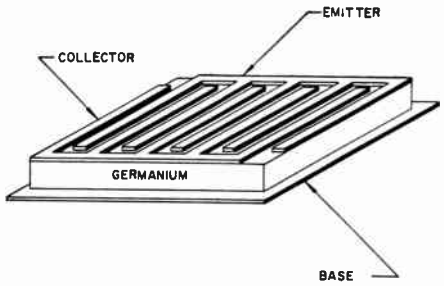


Fig. 12—Interleaved-stripe, alloy-junction power transistor. (Drawing taken from R. N. Hall, PROC. IRE, November 1952.)

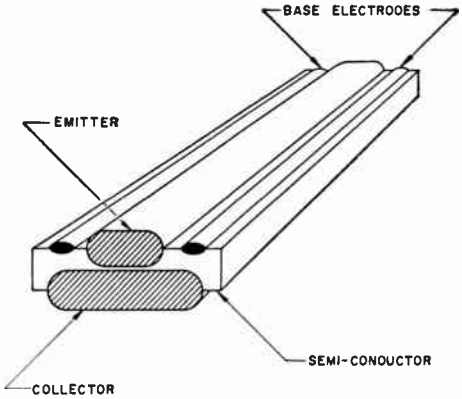


Fig. 13—Interleaved-stripe, alloy-junction power transistor. (Drawing taken from N. H. Fletcher, PROC. IRE, May 1955.)

outstanding paper on power transistor design.[21] The basic structure he proposed is shown in Fig. 13. In addition, he described transistors with several emitter stripes surrounded by base contacts.

An interesting possibility with alloying techniques is the construction of a tetrode power transistor.[25] A tetrode has an auxiliary base contact near the opposite edge of the emitter from the main base contact. The additional contact makes it possible to pass current through the base layer in such a manner as to counteract the natural electric field in that layer. Experience with such power transistors has shown that the base transverse bias has the effect of lowering the current gain at low currents, but not so much at high currents; therefore, the gain is less dependent on emitter current than for the triode transistor. In addition, greater circuit design flexibility becomes possible with the addition of the second base contact. The value of the tetrode features must, of course, be weighed against the increased complexity of the transistor.

In view of the almost certain advantages of complex alloy structures, particularly for power transistors, it seems necessary to offer some explanation for their absence from the market today. What is involved here is the mass of technical problems attendant to the production of a new and unfamiliar kind of device. In the face of these problems, the producer of transistors has been forced to seek simplicity. The alloy transistor with opposing circular buttons for emitter and collector and some sort of ring base contact offers such simplicity. The particular problem of the alloy power transistor manufacturer is that he must provide a jig which will constrain the molten alloy and force it to uniformly alloy over a large area. This means that the jig is a little complicated. To construct elaborate electrode configurations the jig would be still more complicated.

What has been done recently is to construct alloy power transistors with a "washer-shaped" emitter. The base contact is placed both inside and outside this washer, thereby creating a cylindrically symmetrical structure analogous to the linear stripe configuration of Fig. 13.

The electrical characteristics of an alloy, germanium power transistor of recent design are presented in graphical form in Fig. 14. The input, output, and transfer characteristics are given; from these the circuit performance may be determined graphically. It is apparent that a high power gain can be attained with such a transistor.

Diffused-Junction Transistors

The first diffused-junction transistors were developed for the purpose of improving the frequency performance of transistors. This required the fabrication of a very thin base layer, and diffusion has proven to be an excellent method for doing this. The necessity of a thin base layer in power transistors has been mentioned, and

[25] J. T. Maupin, "The tetrode power transistor," IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 1–5; January, 1957.

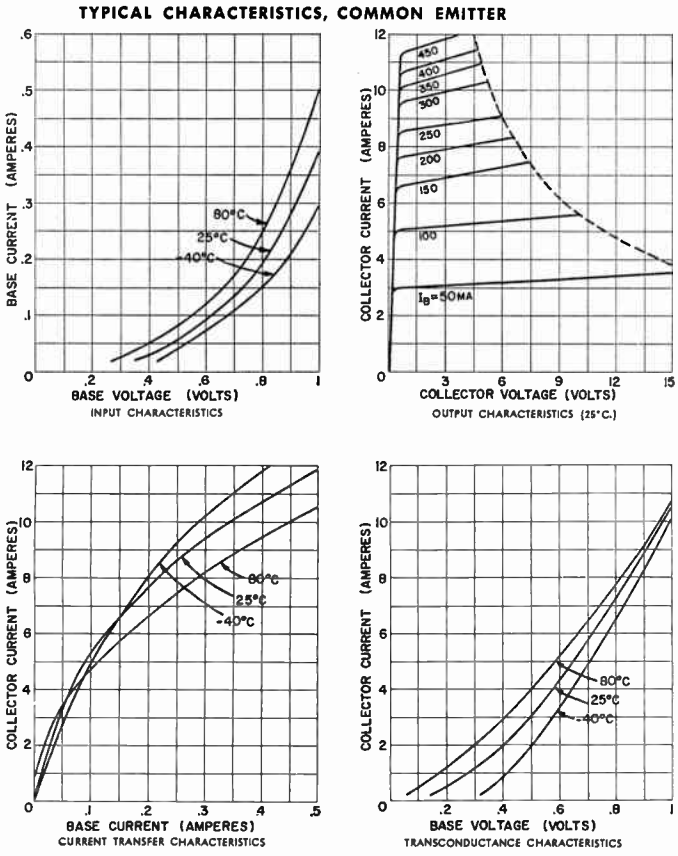


Fig. 14—Characteristics of alloy-junction, germanium power transistor. (Courtesy of Delco.)

it is natural therefore that diffusion be seriously considered as a method of fabricating power transistors.

In addition, diffusion provides a method for constructing large-area junctions with a high degree of uniformity. In particular, the planarity of such junctions, resulting from the simplicity of the process of diffusion into a homogeneous solid, is markedly better than can be attained over large areas with any other process. The rather mechanical processes of etching, growing, melting, or dissolving, used in other fabrication methods, are irregular and uncontrollable in comparison with solid-state diffusion, when evaluated in terms of large areas.

Although diffusion could be used to produce a variety of transistor structures, the diffused-base structure is the only one which has come into use. This structure has the advantage that the base width is determined by processes operating from one side of a wafer of material. Hence the base-layer thickness is unaffected by the wafer thickness. The diffused base has the additional advantage that the grading of impurities in the base layer creates an electric field in the base layer which aids the flow of carriers.

The emitter of the diffused-base transistor may be produced by diffusing or by alloying. The advantages of diffusion are such that alloyed emitters will probably not be seen on power transistors. Various methods are available for producing more or less elaborate electrode structures. This type of transistor is still quite new and much work is being done in the laboratories to establish methods of producing such transistors on a large scale.

Several diffused-base silicon power transistors have been described in detail in the literature. One such structure is shown in Fig. 15, next page. This is an *n-p-n* transistor with a simple electrode configuration. The impurity distributions achieved by diffusion are shown in Fig. 15. An example of a more detailed electrode structure for a diffused-base *p-n-i-p* transistor is shown in Fig. 16. This was designed as a high-frequency silicon power transistor, and includes a diffused collector region to reduce the collector parasitic resistance, as well as a narrow emitter stripe with base contacts close by, on each side.

Diffused-base transistors offer hope for both higher performance and lower cost. Most of the diffused-base power transistor development work is directed toward silicon transistors because of their excellent high-temperature performance and because the diffusion techniques are somewhat more advanced for silicon than for germanium. Higher frequency performance also seems unavoidable because of the inherent thin base layers.

Other Semiconductor Power Amplifiers

Field-effect transistors received a brief flurry of attention as power transistors but seem to have gone out of style. These transistors employ space-charge layer widening at "gate" electrodes to control the flow of current from a "source" electrode to a "drain" electrode. From the standpoint of their promise as power transistors, these devices too closely resemble vacuum tubes in their characteristics. In particular, a rather high voltage is required for their operation at high currents. For this reason the efficiency cannot be as high as for a conventional transistor, and the advantage of possible operation at low voltages is lost. The high input-impedance is no particular advantage for a power transistor, assuming a given power gain. Field-effect transistors have not proven to be easier to construct, and only a few have been built in the laboratory. On the whole it appears unlikely that these devices will play much of a role in the power transistor field.

The analog transistors suggested by Shockley have also not been exploited. These transistors use a very pure semiconductor and only one type of carrier would be essential to their operation. They closely resemble vacuum tubes. Here again their resemblance to vacuum tubes implies that they would have some of the same disadvantages of that device by comparison with conventional transistors. In particular, the need for a high voltage is likely. Analog transistors are sufficiently difficult to construct and have so little promise of real utility as power transistors that they have received very little attention, even in the laboratory.

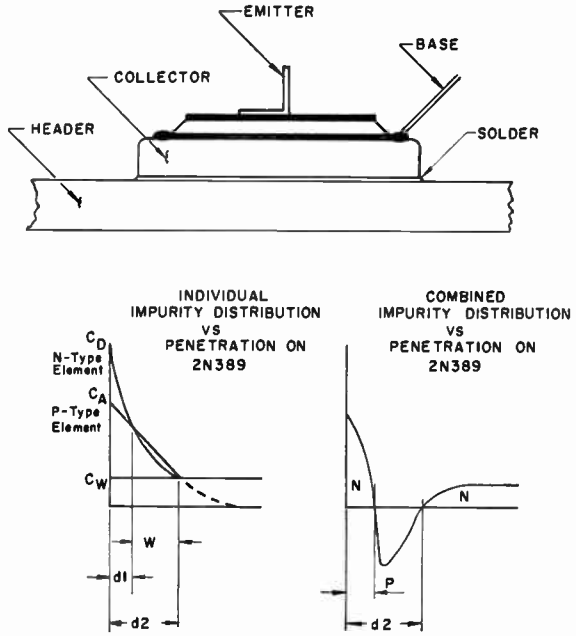


Fig. 15—Diffused-junction, silicon power transistor. (Drawing taken from E. A. Wolff, 1957 IRE WESCON CONVENTION RECORD, pt. 3.)

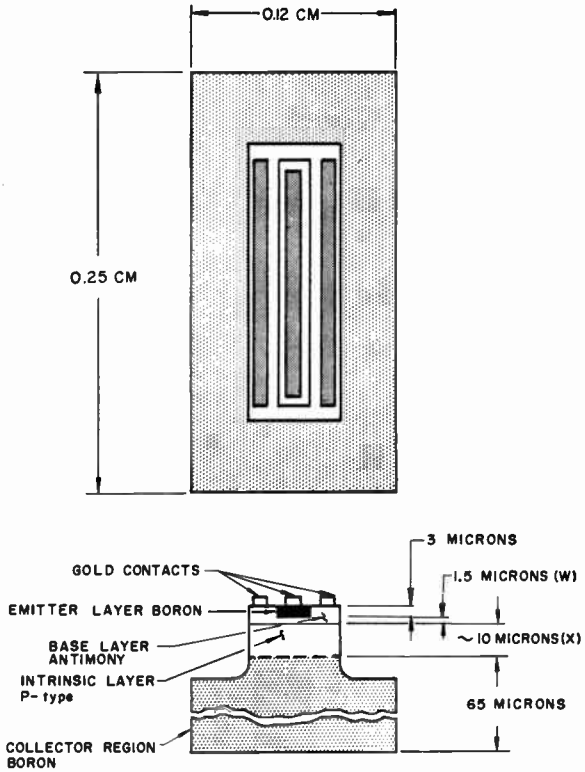


Fig. 16—Interleaved-stripe, diffused-junction, silicon power transistor. (Drawing taken from J. T. Nelson, *et al.*, 1957 IRE WESCON CONVENTION RECORD, pt. 3.)

Besides linear amplifiers, there are several types of bistable devices which hold some promise for those applications which could be satisfied by a controlled switching operation. This might include power converters and controlled rectifiers, as well as high-power switches.

The unijunction transistor or double-base diode consists of a single junction with ohmic contacts placed adjacent to opposite edges of the junction. Because the current path is of necessity rather long, the minimum voltage is appreciably higher than for a conventional transistor. On the other hand, this device is quite simple and may find important applications where its potential low cost and bistable characteristic are advantages. It is unlikely that it can substitute for conventional transistors at really high power levels or high current levels because of its relatively high minimum voltage.

The *p-n-p-n* diode is a more complex device with a bistable characteristic. The device is capable of a very low minimum voltage because the current flows through a very short path, just as it does in a junction transistor. It appears likely that such a device will find extensive use in bistable applications, possibly at high power levels, once the technology is competent to produce so complex a device at reasonable cost. The low versatility of a bistable device makes large-scale production and accompanying cost reduction difficult.

The avalanche (or "delayed-collector-cutoff") transistor is also a candidate for development as a bistable device. Some laboratory models have been made, and many ordinary transistors exhibit the phenomenon to a marked degree. The comments made in regards to the *p-n-p-n* diode apply also to the avalanche transistor. Design details differ, of course. It seems likely that the *p-n-p-n* diode is capable of a lower minimum voltage. On the other hand, the avalanche transistor might be easier to produce. In either case, the result is a highly specialized device.

This by no means exhausts the list of amplifying semiconductor devices. Others are discussed elsewhere in this issue. What is striking about an examination of the list of possible devices is that, as a power amplifier, the junction transistor first described by Shockley still seems to offer the greatest promise for further development. Already this device has found wide use, and reasonable designs offer reason to expect much improved performance in the near future.

PACKAGING AND HEAT TRANSFER

Typical Package Types

The outer limits of the package of a power transistor are not always easy to define. Because heat transfer is of such vital importance, the surroundings of the transistor must be considered in detail. One of the functions of the package is the transfer of heat to the environment. This function is performed by the header on which the active element is mounted, by the shell which encloses and protects that element, by the socket, if used, by the chassis or heat sink fin to which the header is attached, and by the surrounding air and other apparatus which affects the motion of the air. Finally, the heat must be transferred from the circuit enclosure to the room air.

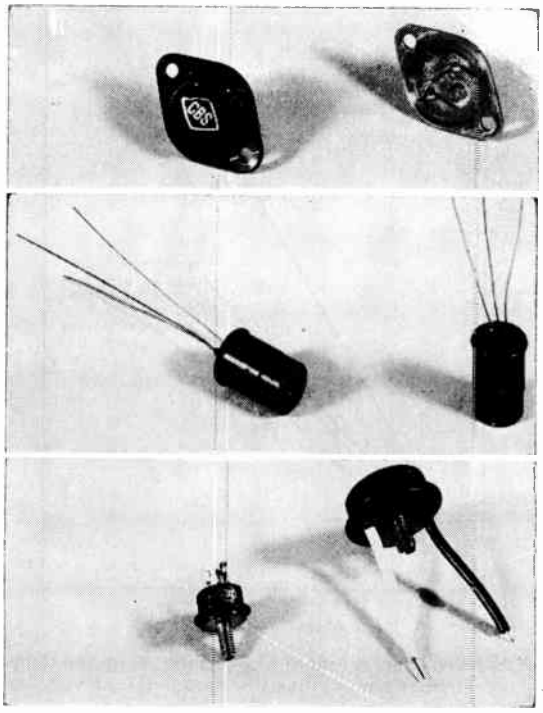


Fig. 17—Representative power-transistor structures. (Courtesy of CBS.)

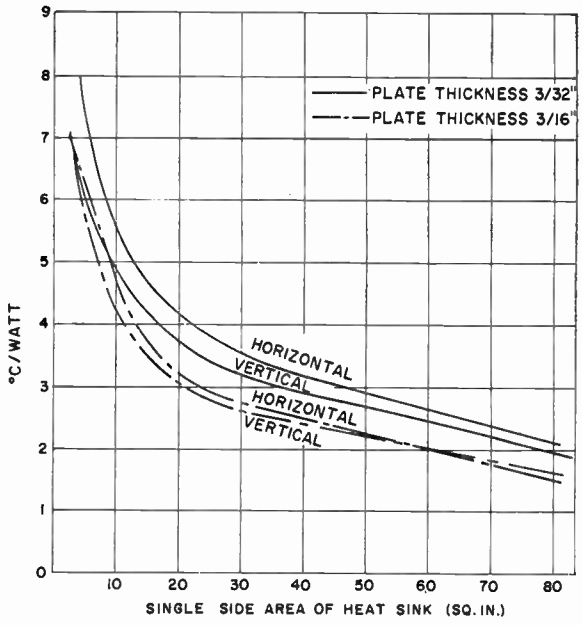


Fig. 18—Heat-transfer characteristic of square aluminum heat sink. (Courtesy of Delco.)

The manufacturers of power transistors have set various limits upon the extent of their responsibility to provide a complete package. One extreme is to provide a transistor having one flat surface, with blind tapped holes, a stud and nut, clear holes, or a spring clip by means of which the transistor may be attached to some sort of heat sink. The heat sink is then to be provided by the circuit engineer or his mechanical engineer associate, who is laying out the system assembly. The other extreme is sometimes attempted, by attaching fins of some sort directly to the transistor, so that only lead connections are required of the user. The latter is only a partial solution, for as pointed out above, the heat transfer is not complete at this stage.

It is to be expected that this situation of various degrees of manufacturer packaging will continue. Consider the two extremes of power transistors: an ultra-high-frequency, 10-mw transistor and a 1-kw, water-cooled transistor. The former apparently requires very little attention to external cooling because almost any container will provide adequate heat transfer to the environment. Admittedly, the internal design may need to take careful account of the high power density. The water-cooled transistor, on the other hand, must be designed with hose connections at least, and possibly with an elaborate circulating system. Between these extremes lie the power transistors as they are now known, with the transistor manufacturer and the transistor user sharing responsibility for insuring that the heat transfer is adequate.

Some transistor packages, without any of the associated heat sinks, are illustrated in Fig. 17. These are representative of the packages being used but do not include all of the structures being marketed by the numerous companies now producing power transistors. In general, the packages have provision for attachment to a heat sink, but do not incorporate the heat sink as a part of the package. A notable exception, not illustrated, has a compact fin structure on the package in addition to provision for attachment to a larger heat sink.

Power transistor manufacturers are able to provide assistance in the design of heat transfer systems and to some extent to provide the items of hardware needed to construct such systems. Although books on heat transfer are plentiful, they rarely give the kind of information needed for the design of small systems applicable to individual electronic components. Several good discussions of heat-transfer for transistors have appeared in the literature.[26,27] A rule of thumb which the author has found to be quite accurate and very useful is that one square inch of metal surface will dissipate 8 mw if its temperature is 1°C above the ambient air temperature. This value is applicable for the temperatures ordinarily encountered with transistors and the dissipation is very nearly proportional to the temperature rise. Such factors as the color of the surface and the proximity of objects interfering with air circulation obviously modify the effectiveness of the heat transfer to the air, but usually not by a large factor. The spatial orientation also has an effect as may be seen in Fig. 18. This graph shows the temperature rise at the center of a flat plate in free air with a power transistor mounted at the center of the plate.

[26] H. T. Mooers, "Design procedures for power transistors," *Electronic Design*; July, September, and October, 1955.
[27] B. Reich, "Thermal considerations in power transistor applications," *Elec. Mfg.*, p. 162; May, 1957.

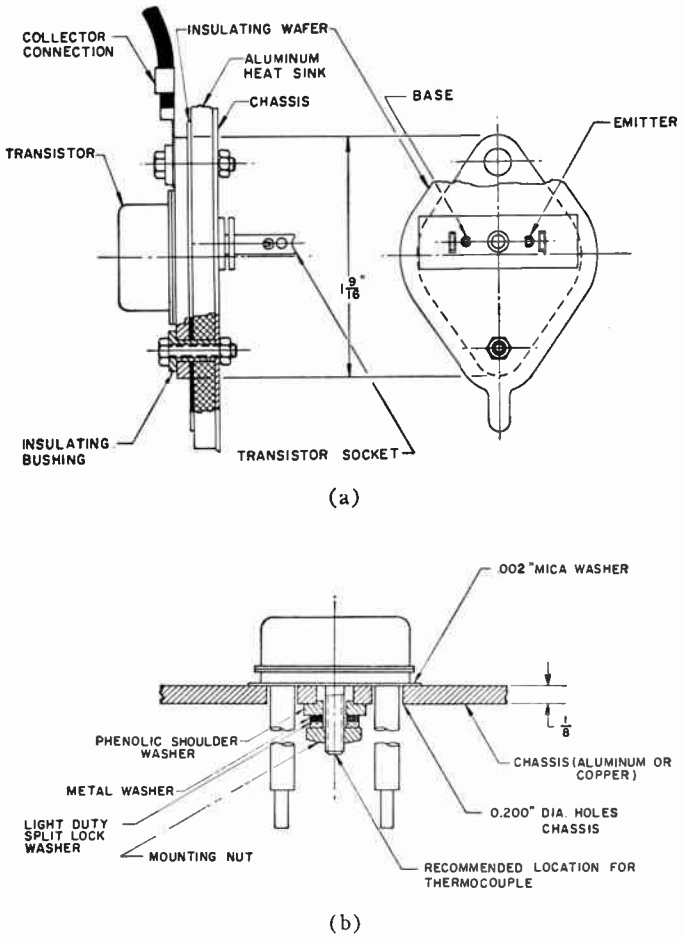


Fig. 19—Insulated mountings for two types of power transistors. [(a) Courtesy of Bendix, (b) Courtesy of Delco.]

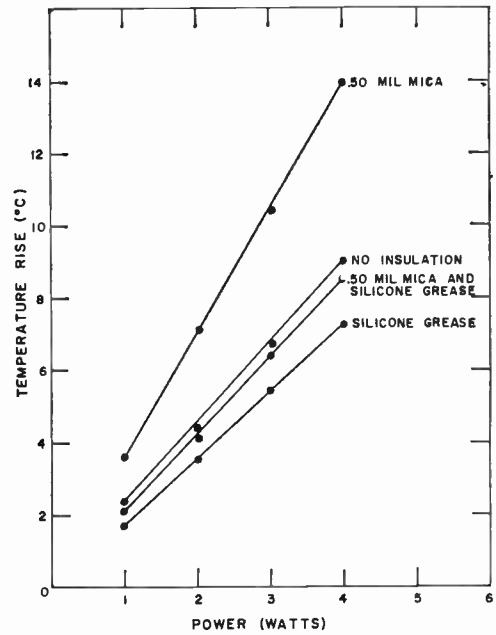


Fig. 20—Observed effects of mounting techniques on the temperature rise of a particular power transistor mounted on a very large heat sink.

More often than not, the transistor must be insulated from the chassis. In general, the farther the electrical insulation is separated from the transistor the less will be its effect upon the heat transfer. Practically, this means that it is preferable to attach the transistor directly to the heat-dissipating fin, and then insulate the fin from the chassis, rather than to use insulation directly under the transistor. However, the latter is sometimes simpler and is widely used. Mounting details for two common types of package are shown in Fig. 19.

Mica is commonly used for insulation because it may be obtained in very thin sheets and will stand the mechanical abuse involved in being clamped between the transistor and the heat sink, with a high probability of sharp edges and burs which would puncture many other materials. The insulator necessarily adds some thermal resistance, but the amount added may not be serious, values of 0.2 and 0.5°C are quoted. Fig. 20 plots the temperature rise vs the power dissipated for one specific transistor mounted on a nearly ideal heat sink. The graphical presentation contributes little except to show the linearity of the temperature rise and to show several points on each line, which serve to substantiate the data. The internal thermal resistance was apparently about 1.8°C per watt, judging from the lowest thermal resistance which was obtained. However, mounting without grease to exclude air gave a resistance of 2.2°C per watt and the use of a mica insulator and grease gave only 2.1°C per watt. The use of the mica insulator without grease was very disadvantageous, giving a resistance of 3.5°C per watt. Although these measurements were repeated successfully with transistors of the same type, different results are to be expected for other transistors, depending on details of the method of mounting. It is essential that the mating surfaces be flat and free of burs, and that the mica sheet be of uniform thickness. The use of a grease, just viscous enough to remain in place, makes the mounting conditions less critical. A thicker mica sheet than indicated in Fig. 20 is used ordinarily for ease of handling and less likelihood of shorting.

Power Rating

The power rating of a power transistor is a function of a number of factors. A complete statement of the rating requires a description of those factors. The difficulty in making a simple statement of the rating is that the allowable power dissipation is a strong function of the environment in which the transistor is to be used. Competitive pressures have resulted in power dissipation claims based on ideal environmental conditions not realizable in practice. Fortunately, experience and increased knowledge on the part of both maker and user have led to a more satisfactory situation in which the emphasis is on accurate and sufficient data.

The minimum information necessary to determine the allowable dissipation consists of: 1) the maximum permissible junction temperature, 2) the internal ther-

mal resistance, 3) the thermal resistance introduced by the mounting method, 4) the thermal resistance between the mounting surface of the heat sink and the environment, and 5) the temperature of the environment.

The transistor manufacturer ought to provide the first three data: the user must provide the last two. In many cases the transistor manufacturer also gives some examples of heat sinks, and may even furnish a heat sink, and he will also give some estimates of the allowable dissipation with a typical heat sink in a specified environment. This kind of information is helpful to the potential user in guiding him to a transistor which offers some possibility of satisfying his need.

The designer responsible for the choice of a power transistor to match a particular need must estimate what the environmental condition will be. This consists of the ambient temperature and the thermal resistance between the transistor and that ambient. Then, by adding all of the thermal resistances, he can calculate the power which can be dissipated without having the transistor junction exceed its rated temperature. The determination of the effectiveness of the heat sink in a particular environment cannot be done with great accuracy by means of heat-transfer theory because the situation is far too complicated. On the other hand, the effectiveness of the heat sink can very quickly be determined by measurement.

To determine the heat sink thermal resistance, all that is required is: 1) a thermometer to measure the ambient temperature, 2) some method of determining the temperature of the transistor mounting surface or of the transistor junction, and 3) a method of connecting the transistor so as to dissipate a measured power. Of these, only item 2) may be missing in the electronic laboratory.

Two methods of determining the temperature of the junction are in common use. One way is to use a thermocouple and its associated meter or potentiometer to measure the temperature of the base of the transistor. A recommended location for this is to be noted on one of the transistors of Fig. 19. An alternative method is to make use of the temperature sensitivity of some transistor parameter, so as to use the transistor junctions themselves as an internal thermometer. To do the latter, it is usual to first calibrate the transistor as a thermometer, using an oven or oil bath and an ordinary thermometer. A complication in this method is that it will usually not be possible to measure the temperature-sensitive parameter while at the same time dissipating the requisite power. This is certainly true if the collector-junction reverse leakage current is used as the temperature-indicating parameter. In such a method, it is necessary to switch rapidly from the circuit providing the power dissipation to the measurement circuit, and to accomplish the measurement before the temperature has fallen. An oscilloscope is essential for this observation. An alternative is to measure a parameter such as

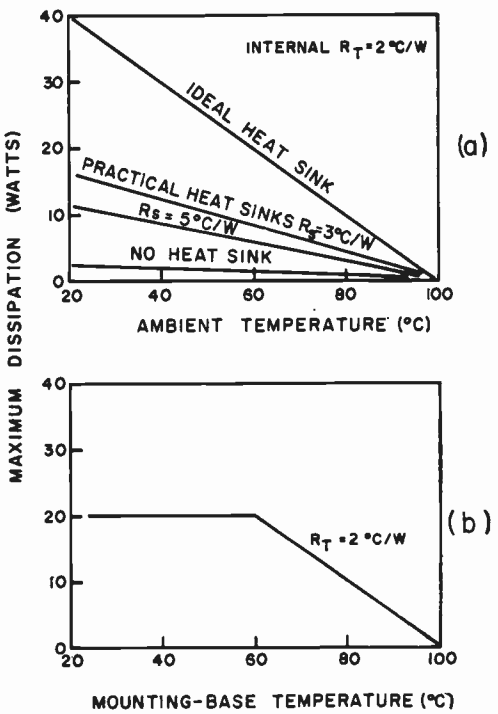


Fig. 21—Derating curves for hypothetical power transistors. (a) Based on ambient temperatures. (b) Based on mounting-base temperature.

the emitter resistance r_e while dissipating power at a constant emitter current. This latter method requires confidence in the emitter-resistance theory and measurement, or a calibration procedure; and the temperature sensitivity is not as great as that of the leakage current. On the other hand, the emitter resistance is more stable and reproducible than is the leakage current.

Power derating curves are of three types, those based on ambient temperature, those based on mounting-base temperature, and those which take into account the circuit stability. Circuit stability considerations, although of vital importance, will not be discussed here.[8]

Fig. 21 shows derating curves stated in two ways. The upper family has as abscissa the ambient temperature. Now it is obvious that such derating is dependent on the the heat sink as well as on the transistor characteristics. Only the line labeled "Ideal Heat Sink" is characteristic of the transistor. This line has a slope given by the internal thermal resistance of the transistor and reaches zero power at an ambient temperature equal to the maximum allowable junction temperature. The advantage of this type of derating graph becomes clear when there is plotted on the same graph some derating lines for practical heat sinks, including no heat sink at all. From such a graph the reader may quickly get an estimate of the allowable power dissipation in the environment he is able to provide.

An alternative power derating graph has as abscissa the mounting-base temperature. This graph has the advantage that it is characteristic of the transistor and not dependent on the environment. On the other hand,

the reader is left to seek elsewhere for data that will give him an idea as to the practical power-dissipation ability of the transistor, because such a graph applies directly only to an ideal heat sink.

Present Performance and Future Prospects

However desirable it might be, the compilation of transistor characteristics will not be attempted here. There are now about 100 transistors which are described as power transistors, and more than a dozen manufacturers of such transistors. The compilation of the characteristics of these transistors is tremendously complicated by the fact that each manufacturer has his own methods of specification. Standardization is evident only in packaging, and there only to a degree. Nevertheless, the data sheets now available are very informative.

Graphs for use in discussing the present performance and future prospects of power transistors are given in Figs. 22 and 23. The area of collector voltage and collector current now covered by such transistors is approximately that of the shaded area of Fig. 22. Note, however, that the highest voltages and highest currents are not simultaneously useful, with the possible exception of high-efficiency, square-wave applications. Also, let the user beware of using the device to the limit of its ability, for in almost all applications transients are present which may damage the transistor if no safety margin is provided. The steady-state dissipation of 100 watts is probably the upper limit with present transistors, and only with extreme cooling measures, such as are rarely practical.

No known reason exists, in principle, why the voltage and current limits cannot be extended to the limits of Fig. 22, namely 1000 volts and 100 amperes—or higher. The limitation which does exist is nonetheless real, and is a technological one. To achieve such voltage and current ratings requires a degree of perfection of materials, processing, and surface condition which does not now seem to be practical. In addition, improved fabrication methods will be needed to produce the electrode configurations and junction configurations which are necessary.

A second method of portraying power transistor performance is that of Fig. 23. This is the power-output, frequency plane. The shaded area is believed to be about what is now available in power transistors. The degree of availability is the debatable point, of course. Certainly, laboratory transistors are being constructed which approach the "Foreseeable Future" rainbow, but these are not generally available. In the case of transistors which may be purchased, the high cost of some types makes them effectively unavailable for many applications. The growing use of such transistors will reduce their cost by virtue of increased production. At the same time performance will continue to advance.

The dotted curve labeled "Foreseeable Future" in Fig. 23 is the result of several developments whose impact has not yet really been felt. Notable among these is the increasing ability to produce silicon transistors. This material offers hope of greatly increased power ratings in the next few years. At high frequencies, the diffused-base transistor has much to offer, with either germanium or silicon as material.

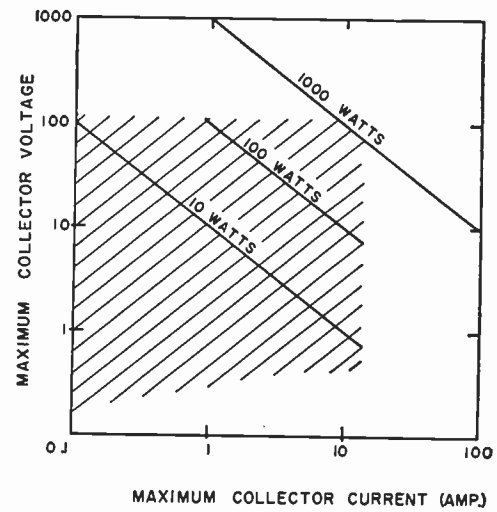


Fig. 22—Voltage and current capability of present power transistors.

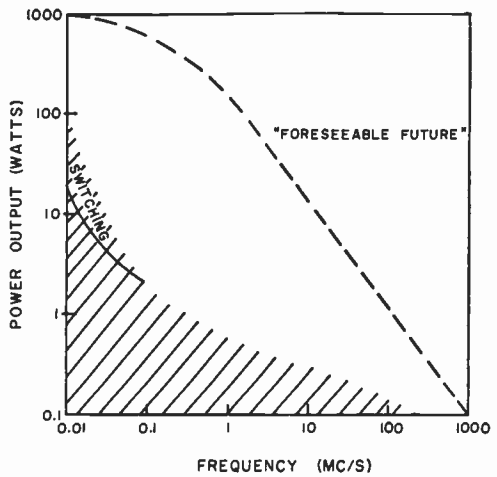


Fig. 23—Power output and frequency capability of present power transistors.

Beyond the foreseeable future the picture for conventional power transistors is hazardous to estimate, but it seems useful to look briefly at the three areas in which progress is to be made.

Design

By following the design trends over the past decade, and by the more scientific procedure of looking at the design theory, one may draw certain general conclusions about the designs to be seen tomorrow. For reasons that have been discussed, thinner base layers are in order for power transistors, regardless of whether the frequency objective requires this or not. This is not to say that the optimum design for a given application will always re-

quire the thinnest attainable base layer. What is claimed is only that, in general, thin base layers are advantageous and are likely to be seen increasingly in the future.

A second trend in power transistor design is the increased recognition of the desirability of a narrow emitter, inasmuch as the portions of the emitter remote from the base contact are probably inactive. Although these portions of the emitter may not directly degrade low-frequency performance, it is advantageous to use the area for additional base contacts and thereby improve the performance. Further improvement in emitter efficiency by the use of higher doping concentrations is also to be expected.

Summarizing, the design of power transistors with high performance is likely to include a thin base layer and an intricate electrode structure, having narrow emitter stripes and closely spaced base contact stripes. By no means does this imply that simpler structure will vanish. Traveling-wave tubes have as yet no pretensions toward replacing the simpler tubes in your television receiver; by the same token the simple, economical, power transistors sold today are likely to be in use in essentially their present form for a long time to come.

Techniques

The thin base layers and intricate electrode structures demanded by advanced designs will require substantial advances in techniques. In addition, improved reliability, lower cost, and better performance of the present designs are much in demand, and require both improvement in the myriad details of processing and some technological advances worthy of being called "breakthroughs."

For example, the maximum collector voltage of germanium power transistors has not yet gone above 100 volts, although that high a voltage has been attainable in the laboratory for a long time, and higher voltage germanium diodes are available. The difficulty in manufacturing transistors with higher breakdown voltage has been presumed to be a surface problem. Research in this area generally has avoided attacking this problem directly, perhaps because of more pressing problems, or perhaps because the research man cannot get his hands on transistors at an intermediate stage of processing without running the risk of being caught up in the problems of the processing, and being converted into a process engineer. In addition, the success in achieving high voltages with silicon has held out hope of evading the problem altogether.

Improved reliability is most likely to come from refined processing, along the lines established by the manufacturers of premium tubes. In general, this means conservative ratings, excellent process control, the use of high quality materials, and a very high standard of cleanliness in the processing and assembly. In addition to these necessary steps, a better knowledge of surface chemistry will lead to instructions for the fabrication of reliable surfaces.

Materials

The present materials being exploited are germanium and silicon. Silicon has the important advantage of a wider energy gap, and therefore better high-temperature performance. The lower free-carrier mobilities in silicon impair the high-frequency performance of a given structure. In power transistors, this theoretical difference will be obscured by differences in fabrication techniques for some time to come.

There are three general areas in which work on improved materials is showing promise for better power transistors. One of these is the development of silicon-germanium alloys. Such alloys may make possible a compromise design of the material such as to combine, to a degree, the best attributes of silicon with those of germanium. Power transistors offer a natural place to employ such a material, if it can be produced as high-quality single crystals, and methods of fabrication developed. On the other hand, one would not expect a very great improvement over devices made of the elements, and the fabrication problems may be quite difficult.

A second area of material development is that of the compound semiconductors composed of elements from Column III and Column V of the periodic table. Good crystals of such materials are becoming available. Much work needs to be done before such materials are of the perfection required to make power transistors. It is likely that their transistor usage will start, if it does, with small transistors which are simpler to fabricate and are less demanding of material perfection.

A third material of particular interest for power transistors is silicon carbide. This semiconductor has a wide energy gap and may be useful for making transistors which function when at a red heat. An exceedingly refractory material, very difficult to machine, silicon carbide holds more immediate promise as a material for rectifiers and diodes rather than transistors. Its eventual use for very high-power transistors or for transistors to operate in very high ambient temperatures is a real possibility, but a long developmental period must be expected.

Summarizing, although better materials than germanium and silicon for the design of power transistors are known and being actively worked on, it is highly probable that the art of producing good single crystals of these materials and of manufacturing useful transistors from these materials will require several more years of research and development. The ability to produce good power transistors of germanium and silicon makes such research and development less urgent and therefore less well supported. On the other hand, the techniques developed for the elemental semiconductors provide a valuable guide which already is visibly accelerating the work on the more complex materials.

ACKNOWLEDGMENT

The author wishes to express his appreciation to the many people at Bell Telephone Laboratories who contributed to his general experience and to his power transistor work there. In particular, thanks to W. J. Pietenpol and R. M. Ryder for their most patient encouragement. The author's coworkers at Bell Telephone Laboratories and Pacific Semiconductors who, in one way or another, contributed to the background of this article are too numerous to list with fairness to all. Several companies have been kind enough to assist in the compilation of this article by furnishing data, literature, photographs, and their own comments on the development of power transistors. To list some such companies and not to list all would risk an injustice. The reader who pursues work with power transistors will soon acquire his own complete list of suppliers and will recognize that each has contributed in its own way to the work summarized here. Finally, it is obvious that this article is in no way creditable to the author, but represents the labors of countless scientists and engineers in the semiconductor industry.

Measurement of Transistor Thermal Resistance*

BERNARD REICH†, MEMBER, IRE

Summary—**A method of measuring the thermal resistance of transistors using the collector cutoff current as an indicator is described. The author uses a series of two steps to determine the over-all junction-to-ambient thermal resistance. This type of arrangement is most accurate and less time consuming than the one-step over-all junction-to-ambient measurement.**

INTRODUCTION

ONE of the problems still facing the transistor designer and the applications engineer is a realistic appraisal of the power-handling capability of transistors. At one time in the not too distant past, the solution of the power-handling capability problem seemed peculiar to the germanium power transistor. Over the last year, the necessity has been seen for specifying the thermal resistance of germanium small- and large-signal devices in general. At this time, silicon transistors of all types are coming into existence, and the problem of thermal resistance measurement must be faced. The author has written on the subject, but the article was directed mainly to power transistors.[1] At this time, an extension of this method to other types of transistors is presented, based on further investigation. Specific information is given here on the measurement of power transistors and other types fabricated of both silicon and germanium.

GENERAL COMMENTS ON THERMAL RESISTANCE MEASUREMENTS

There are two broad methods by which thermal resistance measurements can be made on a fabricated transistor by nondestructive techniques. The first method is concerned with the direct thermal resistance measurement between the operating junction and the air environment. When a measurement of this kind is made, the thermal resistance between the transistor and the ambient can be calculated by the relationship

$$\theta_T = \frac{T_j - T_a}{P_T} \tag{1}$$

where

T_j = operating junction temperature,
T_a = air ambient,
P_T = total power dissipated, and
θ_T = total thermal resistance between the junction and the ambient.

There are many distinct disadvantages in making a measurement of this sort. A precise transistor calibration is required whether it be by the use of collector cutoff current, I_{co}, or by the input voltage, V_{BE}, varying with temperature. A long thermal time constant between the case and the ambient requires a relatively long period of measurement. Finally, in a measurement of this sort, a rise in case temperature follows the rise in junction temperature leading to a measurement of small differences between the junction and the case, even with the use of a moderate size heat dissipator.

The second general method, developed by the author, eliminates long time measurements by bypassing the thermal time constant between the case and the air ambient. More specific details are presented later in this paper. Essentially, therefore, a two-step process is indicated; first, a measurement of the thermal resistance between the junction and the case of the device which is

* Original manuscript received by the IRE, February 28, 1958.
† U. S. Army Signal Eng. Labs., Ft. Monmouth, N. J.
[1] B. Reich, "Transistor thermal resistance measurement," *Electronic Design*, vol. 4, pp. 20–21; December 1, 1956.

truly the "transistor" thermal resistance. The second step of the process, which is done on a design basis only, is the measurement of the thermal resistance between the case of the transistor and the ambient. A "bogie" value for this latter resistance can be attained by measuring a few transistors of a particular design. Again, these general remarks are applicable to any basic method contemplated. These statements lead to equations useful in calculating θ_T, the over-all thermal resistance. θ_I, the resistance between the junction and the case, may be calculated from

$$\theta_I = \frac{T_j - T_c}{P_T},$$

and θ_c, the case thermal resistance, may be calculated from

$$\theta_c = \frac{T_c - T_a}{P_T}$$

where T_c is the case temperature and the other symbols are as previously defined.

Junction-to-Case Transistor Measurements

At present, two temperature sensitive parameters of the transistor are being utilized to measure transistor thermal resistance; I_{co}, the collector cutoff current, and V_{BE}, the input voltage of the transistor. The author is aware of the latter method, but interested persons are directed to other papers on the measurement of thermal resistance by the V_{BE} method.[2]

For the past two years, the author has utilized the I_{co} technique with success in measuring thermal resistance between the junction and the case of transistors. This technique requires two characteristics of a transistor.

1) It must have a reasonable I_{co}, and an I_{co} which increases monotonically with temperature.
2) At a *fixed* collector voltage, the I_{co} at a particular temperature must be stable over a measurement period of approximately *two minutes*.

It is believed that these are not stringent requirements for a transistor with any usable characteristics. With the above transistor requisites in mind, techniques of measurement are described covering two general types of transistors, power devices, and others.

Power transistors are defined for the purposes of discussion as devices having one of the three elements tied to the case, and confined to a JETEC 30 or larger package. It is realized that this definition is arbitrary, but it has been the author's experience that devices of this sort are easily mounted or clamped to a water-cooled dissipator. The design of the heat dissipator should be governed by its ability to remove heat effectively and rapidly from the case of the transistor. Experience has indicated that a dissipator $3\frac{1}{2} \times 3\frac{1}{2} \times \frac{1}{2}$ inches, fabricated from copper and shown in Fig. 1, is quite adequate.

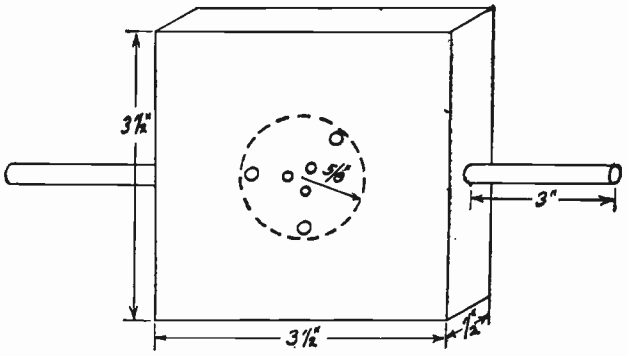


Fig. 1—Heat dissipator.

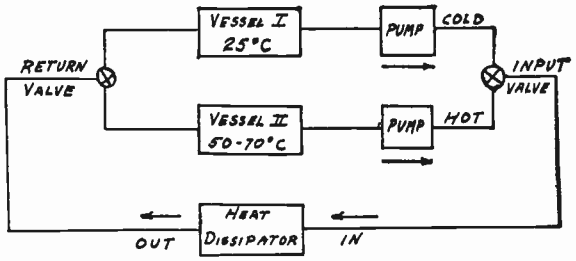


Fig. 2—Thermal circuit.

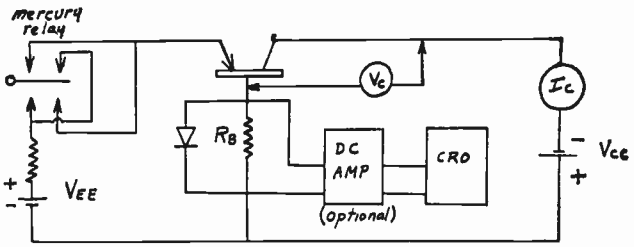


Fig. 3—Thermal resistance measuring circuit.

The coolant suggested for the dissipator is contained in two containers of water, one at room temperature, the other at 50–70°C for germanium devices and 70°C for silicon devices. Arrangements should be made for allowing either temperature water to be pumped through the dissipator. The thermal circuit suggested for the measurement of thermal resistance is seen in Fig. 2. This thermal arrangement allows for versatile control of the mounting base temperature of the transistor.

The electrical portion of the thermal resistance measurement system is illustrated in Fig. 3. This circuit consists of a constant voltage supply in the collector. The magnitude of the voltage used is dependent on the voltage rating of the transistor and the amount of current the unit will conveniently handle. These two variables determine the amount of power the transistor will be subjected to during the course of measurement. A measuring resistor is included in the base circuit of the measurement set. The value of base resistance is again dependent on the measured magnitude of I_{co} at higher operat-

[2] See, *e.g.*, H. G. Lin and R. E. Crosby, Jr., "A determination of thermal resistance of silicon junction devices," 1957 IRE National Convention Record, pt. 3, pp. 22–25.

ing junction temperatures. In order to avoid circuit measurement complications, resistances of 1,000 ohms or less are quite suitable. The mercury relay used is excited by 60-cycle ac, and is connected in such a manner as to have emitter current flowing except during the time the relay reed is moving from one contact position to another. The dc amplifier is included in the circuit to allow for amplification of low levels of I_{co} as would be encountered with silicon power transistors. One of the requirements placed on the amplifier is that it faithfully reproduce the waveform across the measuring resistor R_B. The silicon diode is included to divert forward base current from the measuring resistor and resultant voltage from the dc amplifier.

With this preliminary information in mind, the technique for the measurement of thermal resistance is now set forth.

1) An estimate of the value of I_{co} of a typical transistor at a temperature above the "hot" mounting base temperature should be made. This estimate does not have to be accurate since it is *not* used in the final calculation of thermal resistance.

2) The transistor mounted on the dissipator is now subjected to the higher temperature coolant. An amount of power, y watts, is dissipated in the transistor to bring the value of I_{co} to the approximate value as determined in step 2) above.

3) The device is subjected to the lower temperature coolant. Another value of power, x watts, is dissipated in the transistor, bringing the value of I_{co} monitored on the scope to the same point as in step 2).

4) The value of thermal resistance θ_I, previously defined, can then be calculated by

$$\theta_I = \frac{(T_2 - T_1)^\circ \text{ C}}{(x - y) \text{ watts}} .$$

In the previous discussion, the measurement techniques have been directed to power transistors mounted on water-cooled dissipators. This is not possible with smaller transistors, for example, devices mounted in the JETEC 30 type package and *not* having any transistor electrodes mounted to the case. The author has devised a method of transistor cooling for these smaller devices which lends itself nicely to the identical electronic equipment used with the power devices with minor modifications. For the smaller devices, two temperature baths are used. These baths are 1-liter beakers of transformer oil stirred at a rate at which the heat developed about the transistor mounted in the bath can be removed effectively.

The bath temperatures used are the same as the heat dissipator temperatures formerly referred to in the section on power transistor measurements. For these low power devices, a shielded cable is run from the electronic portion of the measuring equipment to the bath. The transistor, mounted in a socket, is completely immersed in the transformer oil and the thermal resistance measurement process is repeated as outlined in the previous section. It becomes imperative, when silicon devices are measured, that the dc amplifier be used to amplify the reverse voltage drop across the base measuring resistor. It also is necessary that the reverse currents of the specimen measured be much greater than any reverse current normally developed in the diode which clamps the base measuring resistor. The matter of using the dc amplifier with germanium devices is recommended only if the reverse currents are not measurable directly across the base measuring resistor.

This technique can also be extended to high-frequency small-signal and large-signal amplifiers and oscillators. If, however, the process of just using a shielded cable and mounted transistor to connect the electronic circuit and the oil bath is continued, circuit difficulties may arise. It has been found that when units are connected in this fashion that, because of the long leads and generally sloppy circuit conditions, instability may arise during the course of the measurement. When this occurs, certain peculiarities in the measurement which are not immediately evident may arise. In particular, when such a situation of instability occurs, "negative" thermal resistances show up. Physically, in terms of the semiconductor device, this is not possible, and it is suggested that a collector-to-base external capacitance be included. In all cases, the collector-to-base capacitance may not be adequate if instability still occurs and the emitter-to-base capacitance should be added to damp any oscillations. The physical location of the capacitor should be directly at the socket or the receptacle for the transistor.

Case-to-Ambient Measurements

It is necessary now for a complete junction-to-ambient measurement to include the portion of case thermal resistance. When the junction-to-case measurement was made, as described previously, many units of one type were measured to insure product acceptability in terms of this parameter. In determining the case-to-ambient resistance, only a few transistors must be measured since, in general, this parameter of the over-all thermal resistance should not vary appreciably from case to case. The following procedures are recommended.

1) The transistor should be mounted in an enclosure where air currents are minimized.

2) It then is suggested that a measurable amount of power be introduced in the transistor. The case temperature and air temperature should be recorded after thermal equilibrium has been reached.

3) The transistor should be mounted in many positions to determine the worst possible mounting condition. After determination has been made of the thermally poorest mounting condition, two or three power level measurements should be made so that the investi-

gator is assured that the thermal resistance of the case has been well defined. If a few measurements are made on a number of packages, it will be evident that a number can be readily assigned to the case-to-ambient thermal resistance.

Conclusions

A method of thermal resistance measurement of the junction to the ambient is suggested in two discrete steps. It is felt that this is by far the most accurate and simplest way of arriving at an over-all factor, θ_T. This method of measurement is applicable to all silicon and germanium types which meet the basic requirements set forth in this paper. It will be evident from storage and life test data whether the individual investigator has performed the measurement satisfactorily.

Acknowledgment

The author acknowledges the efforts of T. Redgate of these Laboratories for his valuable assistance in the measurements performed during the course of this study.

Measurement of Internal Temperature Rise of Transistors*

J. T. NELSON† and J. E. IWERSEN†

Summary—**A method for measuring the internal temperature rise of a transistor, making use of the variation of alpha with temperature, is described. It consists of a comparison of static characteristics taken at constant temperature by means of a low-average-power pulse technique and characteristics taken under continuous power dissipation. The advantage of the method lies in the fact that measurement is made at temperature equilibrium in the hottest portion of the transistor, and is made with current and voltage distributions essentially identical with those encountered in normal operation.**

Introduction

INTERNAL temperature rise of a transistor, during operation, is one of the many parameters whose determination is necessary for complete characterization of the device. In this paper, we describe a method of measuring this rise which depends essentially on the variation of alpha with temperature. Minority carrier density increases with increasing temperature, tending to fill and deactivate the recombination traps which determine the minority-carrier lifetime in the base layer. Higher lifetime leads to higher alpha so that an increase in alpha with increasing temperature is to be expected in transistors which have a sufficiently high trap density to give an observable effect.

In brief, the temperature rise measurement is accomplished by determining the common emitter static collector characteristics at known base layer temperatures by means of a pulse technique and comparing them with the same characteristics taken at temperature equilibrium with known header temperature and steady dc bias currents for which the power dissipation is known. The intersection of the pulse characteristics and the continuous-power characteristics for equal base current occurs when alpha, hence base region temperature is the same under both methods of biasing. At the intersection, the base-layer temperature, header temperature, and dc bias power are known; hence the base layer temperature rise above the header vs dissipated power can be determined.

This method has at least two advantages over other methods. It measures the temperature in the portion of the base layer where the main current flows, and the measurement is made while the power is being dissipated. An examination of the structure of the many diffused base transistors will show the importance of these advantages. The necessity for making the base contacts often leads to a total collector junction area much larger than the emitter area. The thin emitter and base layer also result in the collector junction's lying close to the upper surface of the transistor wafer so that a large part of the thermal rise occurs in the bulk of the wafer when the collector is mounted on the header. Any method for measuring temperature based on collector junction reverse current[1] will lead to a reading averaged over the total junction area and/or perimeter and will consequently be lower than the temperature in the hot portion of the junction. A measurement made sequentially to the application of power, as the transistor is cooling, may also be misleading since the time constant

* Original manuscript received by the IRE, February 21, 1958.
† Bell Telephone Labs., Inc., Murray Hill, N. J.

[1] J. Tellerman, "Measuring transistor temperature rise," *Electronics*, vol. 27, pp. 185–187; April, 1954.

for thermal redistribution in the bulk of the wafer may be as low as tens of microseconds. The technique of measuring the voltage drop across the forward-biased emitter[2] (which depends on temperature) falls into this category, although it could probably be made simultaneous by pulse calibration similar to that described here. The effectiveness of this procedure could be limited at high currents where the (base current)×(base resistance) drop might mask the emitter drop. Other methods, such as the use of thermocouples or melting waxes, measure the temperature at some exterior point of the unit and are often cumbersome.

The next section of this paper gives the details of the alpha-variation method as applied to the 5 w, 10 mc silicon transistor previously reported on.[3]

Measurement

Static common emitter collector characteristics at constant temperature were obtained by using a pulse technique on the transistor maintained at various ambient temperatures. A dc collector voltage was supplied at the desired value, a current pulse was sent into the base and the collector-current pulse was measured. To give a steady-state indication, the pulse had to be long compared to the common emitter rise time of the transistor, and short enough that the temperature rise during the pulse would be insignificant. The observed common emitter rise time was 0.12 microseconds, which is in agreement with theory.[4] A 0.8 microsecond pulse, which satisfied the rise time requirement, was used.

The 0.8 microsecond pulse at a repetition rate of 120 per second resulted in an average power dissipation of only 10^{-4} w per watt of pulse power so that the average temperature rise in the transistor was negligible. The transient rise in the base layer during the pulse was estimated to be less than 0.2°C per watt of pulse power on the basis of one-dimensional heat diffusion away from the collector junction. Sidewise flow away from the junction was neglected so the estimate was conservative. This rise, however, was small within the power range of operation of the transistor (5–10 w). Verification of the fact that the transient had no effect on the readings could be determined readily by noting that the collector current remained constant during the pulse. If the temperature rise had been large enough to affect the current gain, a variation of collector current with time would have been observed.

The common emitter static characteristics obtained by this method are shown in Fig. 1 with temperature as an independent parameter. Also shown are two curves made by point by point plotting using continuous base and collector currents, while maintaining the header as near as possible to room temperature. Thermal equilibrium was attained in less than two minutes. Power dissipation limits these dc measurements to a V_cI_c product of approximately 10 w.

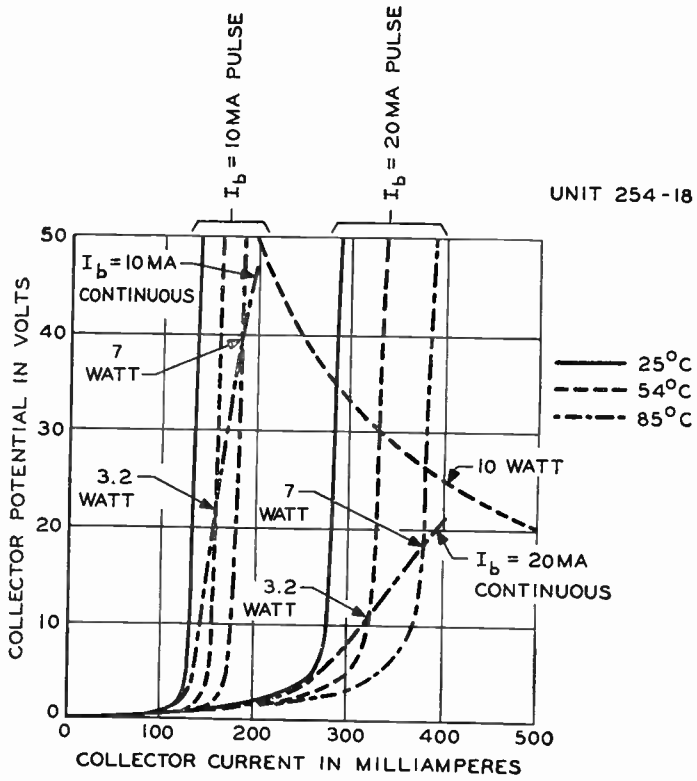


Fig. 1

The large variation of collector current with collector voltage present in the continuous current curves is absent in the pulse curves. Since both represent steady-state curves the difference must be attributed to temperature rise in the transistor when an appreciable amount of power is dissipated in the unit.

The dc curves intersect the 85°C pulse curves when the power dissipated in the transistor is 7 w. (Total power is taken as V_cI_c, since the contribution due to the base current is negligible.) For this power level the measured header temperature was 33°C, so that the rise is 52°C above the heat sink for 7 w dissipation. In the same manner, the intersections with the 54°C pulse curves indicate a 25°C rise (above the 29°C header temperature) for a dissipation of 3.2 watts. Both power levels lead to a rise of about 7.5°C per w.

Conclusion

By making use of the dependence of alpha on temperature the temperature rise in the base layer of a transistor can be measured. This measurement, at least in the case of many diffused base structures, is believed to be more effective than previously used methods in measuring the temperature of the hot, current-carrying region of the device.

[2] S. Bara and P. L. Schmidt, private communication.
[3] J. T. Nelson, J. E. Iwersen, and F. Keywell, "A five-watt ten-megacycle transistor," this issue, p. 1209.
[4] J. L. Moll, "Large-signal transient response of junction transistors," Proc. IRE, vol. 42, pp. 1773–1784; December, 1954.

A Five-Watt Ten-Megacycle Transistor*

J. T. NELSON†, J. E. IWERSEN†, MEMBER, IRE, AND F. KEYWELL†

Summary—A 5-watt, 10-mc, silicon power transistor has been developed. The device, made by solid-state diffusion, uses the intrinsic-barrier structure. Although primarily designed as a 5-watt, 10-mc oscillator, some laboratory samples have delivered as much as one watt of power when used as oscillators at 100 mc. As an amplifier at 10 mc, a unilateral gain in excess of 20 db is obtained at the 5-watt output level. The design, static characteristics, characterization in terms of an equivalent circuit, and performance data are given.

I. Introduction

SINCE the invention of the junction transistor, the general tendency in development has been to increase the frequency response while maintaining or improving the power handling capacity. This paper treats a transistor which falls into this mainstream of development. The design, structure, and characterization of a transistor capable of delivering 5 watts at 10 mc are described. The general approach to the problem has not been limited to making the simplest transistor satisfying this requirement, but has been directed toward increasing (power gain)×(bandwidth)[2] and collector efficiency at this power level.

In line with this plan three ingredients are used to advantage: silicon, solid-state diffusion, and the intrinsic-barrier structure. Silicon is desirable for use in a power unit since its thermal conductivity is higher than that of germanium, but a more important factor is that silicon's intrinsic level is lower for a given temperature. This permits operation at a higher temperature for some maximum-reverse-current specification, making possible the use of a smaller structure for a given power dissipation. The use of solid-state diffusion provides the required means of accurately making the thin base and barrier regions necessary for high-frequency operation. The advantages of the intrinsic-barrier structure have been treated elsewhere.[1] Its use results in a low collector capacitance and a high collector breakdown voltage in conjunction with low-resistivity collector and base regions. Since the space-charge region extends principally into the barrier, modulation of the base resistance, alpha, and the alpha-cutoff frequency by the collector voltage is not troublesome.

A *p-n-i-p* structure was chosen in preference to an *n-p-i-n* for two reasons. At the start of this development, *p-n-p* technology in silicon had advanced further

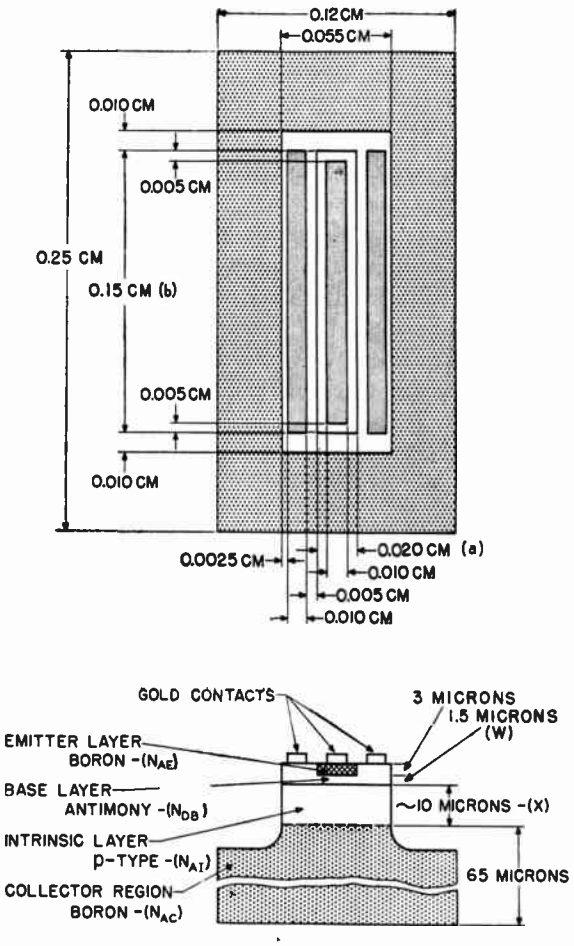


Fig. 1—Structure of the 5-watt, 10-mc silicon transistor.

than *n-p-n*, and avalanche multiplication is less severe.[2] In a *p-n-i-p*, space-charge considerations at high current dictate the use of a π-type material in the so-called intrinsic region rather than ν-type. Also, its use results in a better low-voltage alpha and alpha-cutoff frequency because the collector junction lies on the base side of the barrier. The only remaining detrimental aspects, with this choice of material, are high capacitance and a collector-body series resistance at low voltages.

As shown in Fig. 1, a linear geometry was chosen for the emitter and base contacts in preference to circular structures, which, although much in vogue in alloyed-junction transistors, were shown by Looney[3] and others to be less desirable from the standpoint of base resistance

* Original manuscript received by the IRE, February 21, 1958; revised manuscript received, April 2, 1958. This work was supported by a joint services contract, "Engineering Services on Transistors," Signal Corps Contract No. DA36-039-SC-64618, Signal Corps Project No. 323A, Dept. of the Army Project No. 3-19-03-031.

† Bell Telephone Labs., Inc., Murray Hill, N. J.

[1] J. M. Early, "*P-N-I-P* and *n-p-i-n* junction transistor triodes," *Bell Sys. Tech. J.*, vol. 33, pp. 517–533; May, 1954.

[2] S. L. Miller, "Ionization rates for holes and electrons in silicon," *Phys. Rev.*, vol. 105, pp. 1246–1249; February, 1957.

[3] At the beginning of this project, a quantitative design theory, which has been modified somewhat in the light of new data, was given by D. H. Looney in private communication.

and fringing (the driving of the emitter current toward the edges of the emitter by a transverse voltage in the base caused by the base current).

The next section indicates the direction taken in selecting the structural parameters for optimum performance.

II. Design Considerations

In this paper the treatment of the design is kept to a semiquantitative level,[3] using uniformly doped regions for the sake of simplicity. The approach taken will be to relate performance to structural parameters. A decision as to the desired electrical behavior will then indicate the direction in which the structure should be developed. It becomes apparent immediately that the optimum design cannot be fabricated so that a compromise must be made between fabrication limits and the structure which will give ideal electrical performance. It is necessary to select certain structure parameters near the fabrication limit and examine the resulting transistor. As fabrication techniques advance, improvements can be carried out.

Our electrical performance objectives will be: 1) 5 watts output, and 2) maximum unilateral gain and collector efficiency which are compatible with the power requirement. While power output and efficiency are self-explanatory, unilateral gain[4] was chosen as a figure of merit because it is a function of the transistor only (not the circuit in which it is used) and gives a quantity of great intrinsic value.

The power requirement fixes the product of maximum collector current and collector voltage ($V_{\max}I_{\max}$). The maximum collector voltage will be limited by the collector-junction breakdown voltage, which is approximately proportional to the thickness of the intrinsic barrier ($V_{\max}\sim X$) (see Fig. 1 for symbol definitions). To minimize an apparent series collector resistance caused by space-charge effects in the barrier region, the mobile-charge density must be limited in proportion to the fixed-charge density (N_{AI}). Since the mobile charge is swept through the barrier at maximum constant velocity the limit applies to current density, hence $I_{\max}\sim abN_{AI}$, if the full emitter area (ab) is responsible for current flow. The power requirement expressed in terms of structure parameters will then be

$$P(5 \text{ watts}) \sim N_{AI}abX.$$

In the frequency range of interest, a sufficiently accurate expression for the unilateral gain (U) of the device is

$$U = \frac{\alpha_0\omega_t}{4r_b'C_i\omega^2}$$

where α_0 is the low-frequency common-base current gain; ω_t is a characteristic frequency[5] roughly equal to ω_α, the common-base cutoff; r_b' is the base resistance; C_i is the (inner) collector capacitance[5] (roughly the capacitance directly under the emitter); and ω is the angular frequency of operation. It is assumed that, in order to minimize fringing, α_0 must be made close to one, so that α_0 is not to be sacrificed for an (apparent) improvement in some other parameter.

In terms of structure parameters,

$$r_b' \sim \frac{a}{\mu_n N_{DB} Wb} \quad (\mu_n \text{ is the mobility of electrons in the base}).$$

(This term contains only the resistance under the emitter since, in a diffused structure like that in Fig. 1, the part between emitter and base contacts is much smaller.)

$$\omega_t \sim \frac{D_p}{W^2} \quad (D_p \text{ is the diffusion constant for holes in the base})$$

and

$$C_i \sim \frac{ab}{X}.$$

Thus

$$U \sim \frac{D_p\mu_n N_{DB} X}{Wa^2} = \frac{N_{DB}'X}{Wa^2},$$

and is to be maximized. (N_{DB}' is a function almost proportional to N_{DB} since D_p and μ_n are slowly varying functions of N_{DB} only.)

Next we consider efficiency. Inefficiency in quasi-class-A operation is primarily due, in the case of this transistor, to the necessity of maintaining a certain finite voltage ($V_{\min}$) at the collector terminal to maintain low collector capacity by keeping the barrier swept out. (Series resistance in either the emitter or collector circuits also leads to a minimum-voltage requirement but, since such resistances only degrade performance, they should be kept as small as possible.) Then,

$$\text{efficiency (class A)} = 1/2[(V_{\max} - V_{\min})/(V_{\max} + V_{\min})]$$
$$\approx \frac{1}{2}\left(1 - 2\frac{V_{\min}}{V_{\max}}\right).$$

The term $V_{\min}/V_{\max}$ is therefore to be minimized.

$$V_{\min} \text{ (the sweep-out voltage)} \sim N_{AI}X^2$$

so

$$V_{\min}/V_{\max} \sim N_{AI}X.$$

The given power-output, gain, and efficiency objectives may be achieved by the following steps.

1) W should be made as small as possible.

[4] The term unilateral gain is used in the specific meaning assigned by S. J. Mason, "Power gain in feedback amplifier," IRE Trans. on Circuit Theory, vol. CT-1, pp. 20–25; June, 1954.

[5] The use of these parameters instead of the commonly used ω_α and C_c will be explained in a later section.

2) N_{DB} should be made as large as is consistent with good emitter efficiency (which means that the emitter doping concentration should be as large as possible).
3) Since b appears only in the expression for power output, and can thus be used freely to adjust this term to the desired value, the other factors of the product are released from this condition.
4) In view of step 3), a and N_{AI} should be made as small as possible.

The foregoing shows that the ideal transistor structure leans toward thin base layers, long narrow emitters, lightly doped intrinsic regions, and heavily doped emitters. All these quantities are limited by the fabrication technology. Once the best obtainable values have been determined, the one remaining parameter, (X), can be used to exchange efficiency and gain, subject to its own fabrication limits. It is surprising, at first glance, that so little compromising is involved in this design, but this is because the present technology stops short of the point where certain parametric interactions (for instance, those involved in reach-through) become important.

The principal phenomenon omitted from the above analysis, and which must be taken into account, is fringing. It has the effect of decreasing the width of the emitter in all design considerations except collector capacity. Therefore, any central unused portion of the emitter should be eliminated if possible, and if this is done, the analysis given above again becomes valid within its limits.

III. Structural Parameters

The transistors reported on in this paper have been made with parameters given in Table I. They represent the best approach, within reasonable fabricability, to the above analysis.

TABLE I

$N_{DB}=10^{17}$ cm^{-3} (average)
$W=1.5\times10^{-4}$ cm
$b=0.15$ cm
$a=0.020$ cm
$N_{AI}=5\times10^{14}$ cm^{-3}
$X\sim10^{-3}$ cm
$N_{AE}=10^{20}$ cm^{-3} (average)
$N_{AC}=10^{20}$ cm^{-3} (average)

Fringing has not been eliminated in this structure, principally because the emitter must be wide enough to accommodate a gold strip and lead large enough to carry the high current involved. The value of X is approximate because 1) its boundary with the diffused collector cannot be precise, and 2) it is very difficult to measure accurately.

The electrical characteristics and performance of this structure are discussed in the following sections.

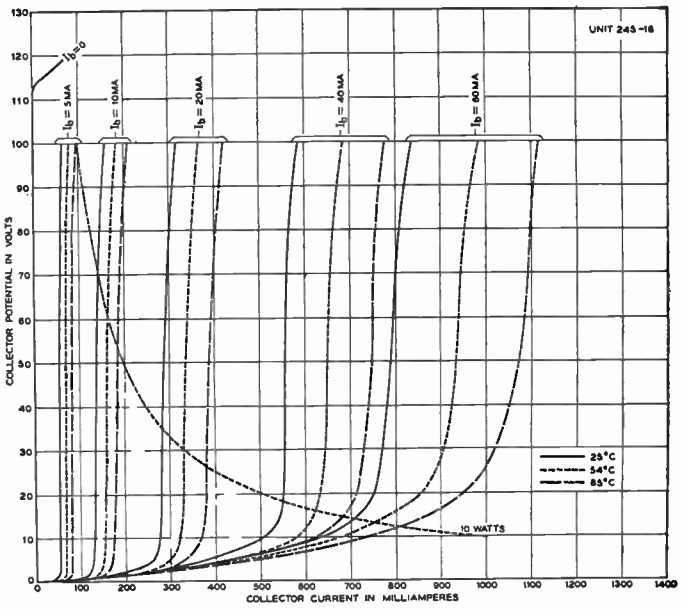


Fig. 2—Common-emitter static output characteristics.

IV. Characterization

Static Characteristics

Static, common-emitter output characteristics are normally determined by simple dc measurements or *VI* sweeping at a low frequency. If such a procedure is used for this transistor, however, the characteristics are influenced by the temperature rise within the transistor due to the dissipation of the bias power.[6] At higher temperatures minority carrier lifetime in the base layer is higher so that a higher alpha is expected. This temperature effect will mask the falloff of current gain at higher currents. Therefore, in order to get a true picture of the behavior of the transistor, it is necessary to hold the temperature of the transistor constant. To do this, we have used a pulse technique. A dc collector voltage was supplied at the desired value, a current pulse was sent into the base, and the collector-current pulse was measured. To give a steady-state indication, the pulse had to be long compared to the common-emitter rise time of the transistor, and short enough that the temperature rise during the pulse would be insignificant.

The common-emitter static characteristics obtained by this method are shown in Fig. 2 with temperature as an independent parameter. A slight increase of collector current with collector voltage below 80 volts is present in the pulse curves, but its magnitude can be explained by the normal spreading of the space-charge region into the base layer. Between 80 and 100 volts there is a slight knee in the curves which is suggestive of avalanche multiplication.

It should be noted that the maximum operating current for which the transistor was designed was 400 ma.

[6] This phenomenon can be used to measure internal temperature rise. See J. T. Nelson and J. E. Iwersen, "Measurement of internal temperature rise of transistors," this issue, p. 1207.

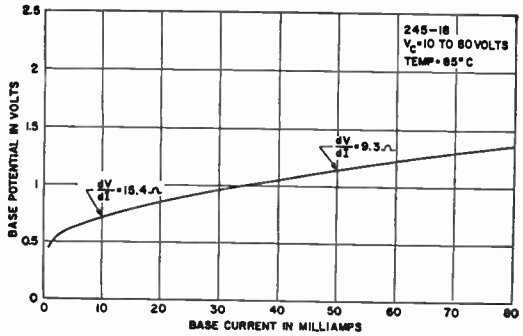


Fig. 3—Common-emitter static input characteristics.

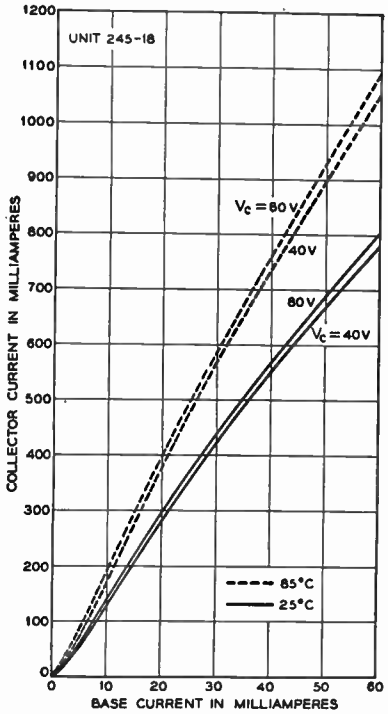


Fig. 4—Common-emitter current transfer characteristics.

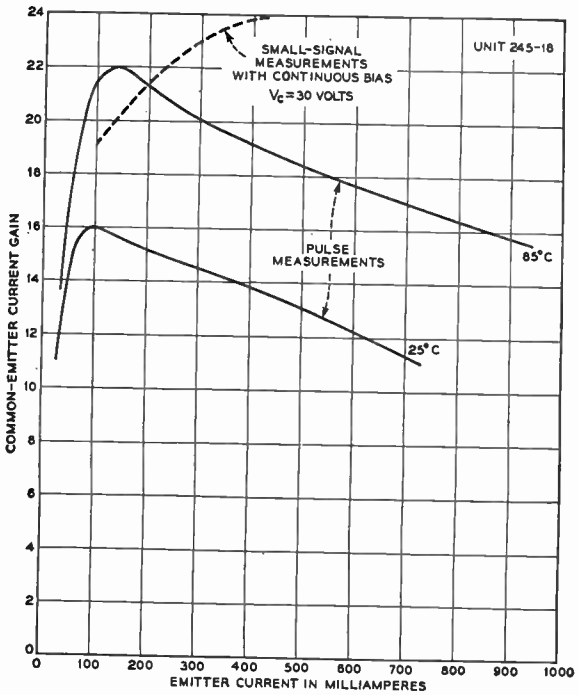


Fig. 5—Common-emitter current gain as a function of emitter current.

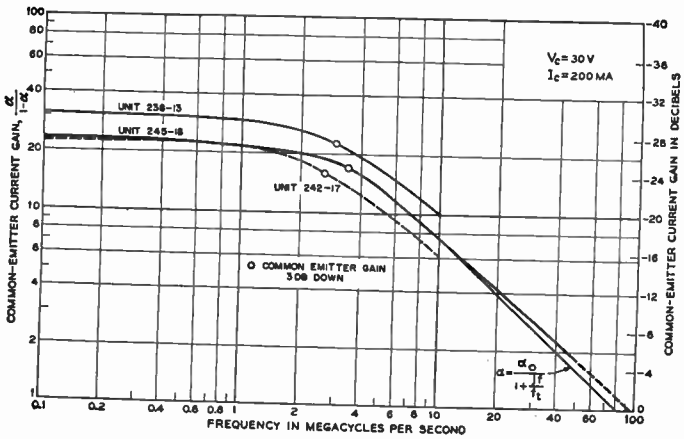


Fig. 6—Common-emitter current gain vs frequency.

In this range the collector saturation voltage indicates the presence of a series collector body resistance of 10 ohms. Above the operating range the resistance increases. Possible causes for this are concentration of emission into a narrow band at the edge of the emitter by the transverse base current and/or a (mobile) space-charge limitation in the collector barrier.

A typical static input characteristic, made by pulse methods at a temperature near that at which the device will presumably operate, is shown in Fig. 3. Above $I_b = 5$ ma, the slope of the curve is almost exactly the base resistance. The effects of fringing and conductivity modulation, both of which operate to decrease r_b' with increasing current, can be seen.

From the collector static curves the current-transfer characteristics have been formed (Fig. 4). The slope of the transfer curves is the common-emitter current gain shown in Fig. 5 as a function of emitter current. The small-signal measurements on the transistor are included for contrast. In these measurements the bias power was on continuously and was proportional to I_e. During such measurements, the falloff of current gain with emitter current is masked by the increase in gain with temperature rise due to increased bias power.

Variation of Current Gain with Frequency

Common-emitter current-gain (beta) measurements have been made to 50 mc. They are shown in Fig. 6 for transistor 245-18 (and for older units to 10 mc). An extrapolation of the curve indicates a unity current gain (for 245-18) at 95 mc (f_{oe}). It should be noted that the frequency at which beta equals one, obtained in this way, is somewhat lower than the common-base cutoff frequency f_α. Moreover, to describe the true behavior of alpha (by means of a simple formula) at frequencies small compared to both f_{oe} and f_α, a parameter f_t (less than f_{oe} and f_α) must be used in

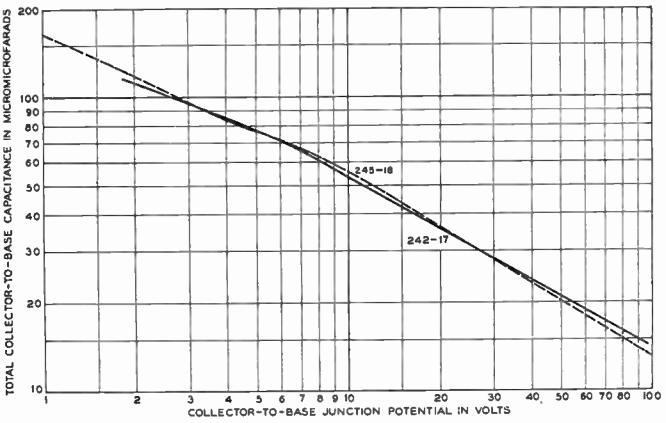


Fig. 7—Collector capacitance as a function of voltage.

$$\alpha = \frac{\alpha_0}{1 + jf/f_t} = \frac{\alpha_0}{1 + j\omega/\omega_t}$$

or the frequency-dependent current generator in the equivalent circuit at 10 mc. This discrepancy occurs because the above formula is only a simplified approximation to a more accurate and complicated function.

Collector Capacitance

Collector capacitance as a function of collector voltage is shown in Fig. 7. There is no leveling off of the capacitance as the space charge is swept into the collector region, as would be expected with a step junction between the collector and intrinsic layers. Rather, the presence of the diffused collector results in lower slope in the variation of capacitance with voltage than would be present without the collector (n-i junction only).

It should be pointed out that this is the total capacitance between the collector and base. It will be shown later that only about one third of this capacitance limits the amplifier performance of the transistor when used where neutralization is possible.

Proposed Equivalent Circuit and Measurements at 10 MC

Since the transistor is intended primarily for use at or near 10 mc we desire an accurate characterization at this frequency. Fig. 8 shows the proposed equivalent circuit which is seen to be the usual high-frequency T plus an element, C_o, called the outer capacity to distinguish it from C_i, the inner capacity.[7] C_o arises from the fact that a large part of the collector junction area is closer to the base contacts than to the emitter (see Fig. 1). $C_o + C_i = C_t$, the total collector capacity, which is the quantity measured by the usual means. We have plotted in a series collector-body resistance (r_c') in the equivalent circuit since the static characteristics of many units show that one exists and is of the order of tens of ohms. However, this measurement is made at a few volts and, if r_c' is due to unswept π-type material, as seems reasonable, it should decrease with voltage and, in a properly constructed transistor, should be negligible at the operating point.

[7] This equivalent circuit was chosen in preference to others because of its close correspondence to the actual physical structure of the transistor.

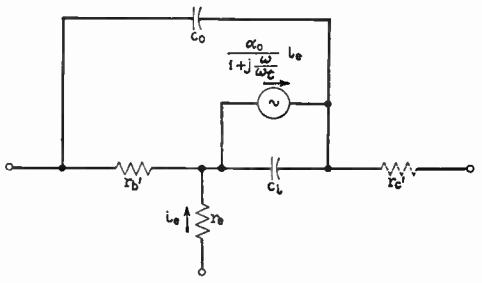


Fig. 8—Equivalent circuit of the transistor.

In order to verify the equivalent circuit, the elements of the circuit were measured in as simple and direct a manner as possible and four-pole parameters were calculated using these values in the proposed equivalent circuit. Some four-pole parameters were then measured directly at 10 mc. It was possible to get three independent parameters by measuring the input and output open-circuit impedances (z_{11} and z_{22}) and the input and output short-circuit admittances (y_{11} and y_{22}). The relationship $z_{11}y_{11} = z_{22}y_{22}$ allowed a check.

Measurements were performed on unit 242-17. $kT/qI_e = 0.13$ ohm at room temperature ($I_e = 200$ ma) but, making allowance for the operating temperature and some small contact and series resistance, r_e was assumed equal to 0.16 ohm.

The small-signal current-gain measurement at low frequency leads directly to α_0. A second measurement at a frequency well beyond the common-emitter cutoff frequency will yield ω_t by:

$$\text{common emitter gain} = \left|\frac{\alpha}{1-\alpha}\right| = \frac{\alpha_0}{[(1-\alpha_0)^2 + (\omega/\omega_t)^2]^{1/2}}.$$

$\omega_t = 391 \times 10^6$ sec^{-1} and $1 - \alpha_0 = 0.0426$ were taken in this way from Fig. 6.

The total collector capacitance ($C_o + C_i$) can be simply measured on a low-frequency capacitance bridge. C_t was taken from Fig. 7 at the bias point of 30 volts where it equals 28 $\mu\mu$f. The ratio C_t/C_i was estimated equal to 4 from the ratio of the total collector area to the emitter area. (The division was made in this way since nearly all the base resistance is close to the edge of the emitter.) Fig. 1 gives this ratio as 3.3 but the value 4.0 was measured on the actual transistor. The only parameter fitted to the four-pole measurements was r_b'; however, this procedure is fairly direct since the conductance portion of y_{11} is closely identical to $1/r_b'$. All these numbers, together with $r_b' = 35.5$ ohms, yield the values in the "calculated" column in Table II.

Measurements of these four parameters were made on a Boonton R-X meter, using appropriate transmission-line transformers, and the observations are given in the

TABLE II

Parameter	Calculated	Observed
z_{11}	$10.7+j3.8$ ohms	$10+j1.5$ ohms
y_{11}	$0.028+j0.0020$ ohm^{-1}	$0.028+j0.0020$ ohm^{-1}
$z_{11}y_{11}$	$0.292+j0.127$	$(0.28+j0.062)$
z_{22}	$86-j35$ ohms	$65-j25$ ohms
y_{22}	$0.0024+j0.0025$ ohm^{-1}	$0.0034+j0.0028$ ohm^{-1}
$z_{22}y_{22}$	$0.294+j0.131$	$(0.29+j0.10)$

"observed" column of Table II. The difference between the products of the measured values ($z_{11}y_{11}$ and $z_{22}y_{22}$) shows that there are inconsistencies in the measured values comparable to the differences between the measured and calculated values. There is such good agreement between the calculated and measured values, however, that we intend to adopt the equivalent circuit of Fig. 8 as representative of the transistor, at least near 10 mc.

The real part of the short-circuit input admittance at frequencies of interest was shown above to be close to $1/r_b'$ while the reactive term is only about 10 per cent of the real part. Because of this the absolute value of the expression is very close to $1/r_b'$. A measurement of the reciprocal of the absolute value is the input small-signal voltage-to-current ratio. It can be made in the same jig and at the same time that the small-signal current gain measurement is made. Thus, the transistor can be fully characterized without any difficult high-frequency bridge measurements; this method is used in the next section.

Unilateral Power Gain

It is desirable to know what power gain can be achieved by a transistor; one must be careful, however, either to specify the circuit used or find a gain expression which is independent of the circuit. We have chosen the latter course and have measured the unilateral gain. Agreement between the calculated and measured values would also serve as a further check on the usefulness of the equivalent circuit.

The principal source of feedback is the outer capacitance (C_o). The contribution by the emitter resistance is negligible.

There are several standard methods of removing the feedback caused by C_o. We have chosen to use an inductance from the output to input to form a parallel resonant circuit with C_o since this method neutralizes the effect of C_o in the input and output impedances as well as removing the feedback.

With the feedback removed the input impedance is very nearly r_b'. The capacitive part of the output impedance was resonated by use of parallel resonant circuit across the output terminals so that the output impedance appeared as pure resistance:

$$R_o = \frac{(1-\alpha_0)^2 + (\omega/\omega_t)^2}{\alpha_0(\omega/\omega_t)\omega C_i}.$$

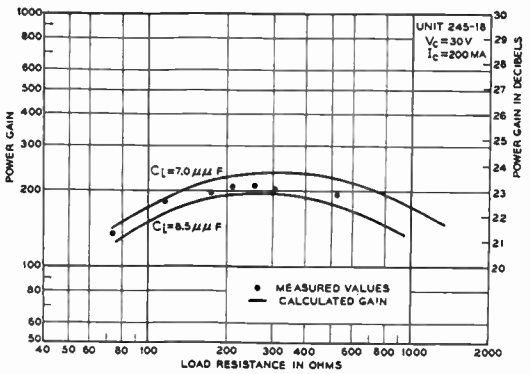


Fig. 9—Power gain as a function of load resistance.

The unilateral gain is

$$U = \frac{\alpha_0\omega_t}{4r_b'C_i\omega^2}.$$

The measured equivalent-circuit parameters of transistor 245-18 are:

$$\frac{\alpha_0}{1-\alpha_0} = 23, \qquad r_b' = 19 \text{ ohms},$$

$$f_t = 83 \text{ mc}, \quad C_i + C_o = 28\ \mu\mu f.$$

Unfortunately, the ratio of emitter area to total mesa area was not measured before the transistor was encapsulated, so C_i is not known exactly. Experience has shown that C_i is from 25 to 30 per cent of the total capacitance, which would give it a value of from 7.0 to 8.5 $\mu\mu$f. The unilateral gain and output resistance for these values at 10 mc are:

$$C_i = 7.0\ \mu\mu f, \quad R_o = 323\ \Omega, \quad U = 238;$$
$$C_i = 8.5\ \mu\mu f, \quad R_o = 266\ \Omega, \quad U = 196.$$

The measuring circuit was adjusted by tuning the output parallel resonant circuit for maximum output and tuning the feedback circuit for minimum signal at the input with a signal fed into the output. The output power was determined from the voltage across the load resistance and the input power is taken as the product of input current and voltage.

The power gain was measured for several values of load resistance centering about the calculated values so that both R_o and U could be determined. The points in Fig. 9 are the measured values, and the solid curves are the calculated gain for the two values of C_i stated above. The measured power gain falls between the calculated values although the maximum power gain occurs with a load resistance slightly less than the calculated R_o.

V. Performance

Amplifier

Gain measurements in the previous section were small-signal measurements. Since the transistor is not particularly linear for large signals, it was not deemed

feasible to make large-signal calculations. However, since the device will be used as a power amplifier, large-signal gain was determined as closely as possible with the existing equipment. The small-signal circuit of the last section was used except for the output resonant circuit which was changed so that the loaded Q would be about 10 to reduce harmonic output. The measurement of input power described above is no longer very good since the input impedance is a rather strong function of current (see Section IV).

With a collector bias of 62 volts and 200 ma and a load resistance of 250 ohms, 5 watts output could be obtained from unit 245-18 with an input power in the vicinity of 26 mw indicating a power gain of approximately 190 (22.8 db).

The bias power with these conditions was 12.4 watts giving a collector efficiency of 40 per cent. The most important source of power loss which causes deviation from the ideal 50 per cent is the necessity of maintaining the collector capacitance low to prevent distortion by keeping a certain voltage across it at all times. In this case 12 volts was necessary. An unknown portion of the 10-ohm r_c' present at ~4 volts (Fig. 2) may still be operative at ~10 volts, contributing to inefficiency.

Oscillator

Tests of the high-frequency power capability of the transistor were made at 10, 30, and 100 mc. For the first two frequencies the circuit shown in Fig. 10 was used. It can most easily be considered as a common-emitter circuit with a π-feedback network, ($C_3L_2C_4$), between collector and base. The collector is grounded to a heat sink. At 10 mc, the power output was determined by measuring the voltage across the load resistor R_3 with an oscilloscope. By adjusting the components in operation for maximum output, over 5 watts at 10 mc was obtained with an efficiency in excess of 40 per cent. In the case of 30 mc, a series of light bulbs was substituted for the load, and power was measured by brightness and color comparison with nearly identical bulbs driven by known dc power. Three watts at about 30 per cent efficiency resulted. At 100 mc, the transmission-line oscillator shown in Fig. 11 was used. Typically, 1 watt was obtained at an efficiency of about 15 per cent. The best performance to date has been 1.2 watts out at 110 mc with 17 per cent efficiency.

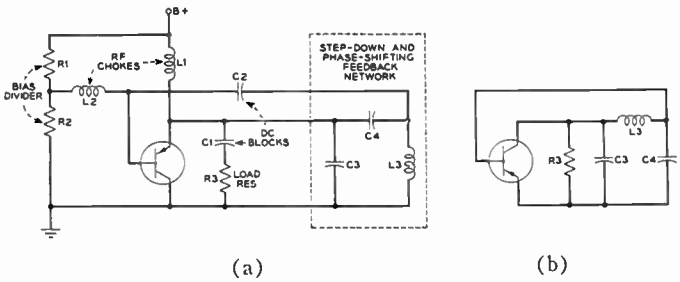


Fig. 10—10-mc and 30-mc oscillator circuit. (a) Oscillator circuit. (b) Simplified oscillator circuit (rf components only).

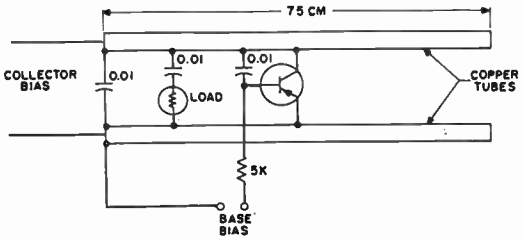


Fig. 11—100-mc oscillator circuit.

VI. Acknowledgement

The authors wish to express their appreciation to M. A. Clark under whose guidance this project was started and most of the work was done. Thanks are also due to I. M. Ross and J. M. Early for subsequent supervision. Substantial contributions were also made by J. F. Aschner, R. G. Barton, R. H. Braun, P. A. Byrnes, S. J. Carrubba, G. T. Loman, D. H. Looney, J. H. McConville, J. A. O'Sullivan, J. W. Peterson, and E. J. Stansbury. We also wish to acknowledge our indebtedness to a host of predecessors and contemporaries without whose ideas this transistor could not have been developed.

The Blocking Capability of Alloyed Silicon Power Transistors*

R. EMEIS† AND A. HERLET†

Summary—An extensive series of silicon *n-p-n* alloyed transistors has been made. One set of these transistors has base thicknesses of 40–45 μ; the other, 55–60 μ. In each group the base resistivity is varied from 2 ohm-cm to 7000 ohm-cm. The "blocking capability" of these transistors is plotted against base resistivity. The results are compared with theory. Agreement is good. Details of the current-voltage characteristic are discussed. Three base resistivity regions are distinguished: 1) pure breakdown; 2) approximately simultaneous occurrence of breakdown and punch-through; and 3) pure punch-through.

I. Introduction

POWER transistors which have been made by the alloying process usually have an *n-p-n* (or *p-n-p*) layer arrangement in which the emitter and collector regions are highly doped. On the other hand, the base region is relatively lightly doped. Since, in addition, one produces practically abrupt *p-n* transitions by means of the alloying process, the blocking capability of such an element is determined by the properties of the base region barring exceptional conditions. It is determined in particular by the resistivity of the base region.

First of all, the blocking capability is limited in exactly the same way as in rectifiers; that is, at the breakdown voltage U_b the field strength in the *p-n* transition (on the collector side) reaches a critical value. By collision ionization, an avalanche build-up of current carriers results, and the leakage current increases abruptly.[1,2] The maximum blocking voltage U_b determined by this effect increases with the resistivity ρ of the base region (Curve 1 in Fig. 1). The relation between U_b and ρ has been discussed theoretically[3] and investigated experimentally many times. The $U_b(\rho_n)$ relation has been measured for *n*-type silicon by Wilson.[4] Herlet and Patalong,[5] and Miller,[6] have determined the $U_b(\rho_p)$ relation. The results of both of the latter studies are in agreement over a wide range with the relation

$$U_b = a\rho^{3/4}, \tag{1}$$

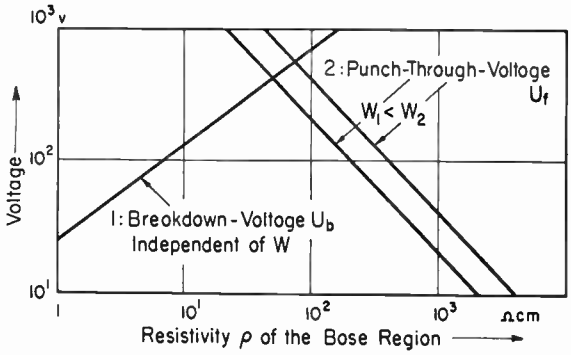


Fig. 1—The limit of the blocking ability of transistors (schematic).

however, with regard to the factor a there is nearly a factor of 2 variation.

Supplementing this limitation due to breakdown there appears in the transistor yet another limit to the blocking capability. If, namely, the space-charge region on the collector side grows into the base region with increasing voltage and finally breaks through to the emitter side of the *p-n* transition, a steep increase in the leakage current appears.[7,8] The limiting voltage

$$U_f = \frac{1}{2}\frac{W^2}{\epsilon\epsilon_0\mu_p\rho} \tag{2}$$

given by the "punch-through" effect decreases with increasing ρ values of the base region and is, moreover, dependent upon the base thickness W (Curve 2 in Fig. 1).

II. Experimental Results

The situation presented schematically in Fig. 1 was checked quantitatively in a large series of experiments. To this end two series of *n-p-n* transistors were made; one set had base thicknesses between 40 and 45 μ and the other, thickness between 55 and 60 μ. Within each series the resistivity of the starting silicon was varied, and thereby, the base region resistivity was varied from 2 ohm-cm to about 7000 ohm-cm. For every resistance value a group of about 10 transistors was alloyed. Fig. 2 shows the results. Every beamed cross of this figure represents a group of transistors. The vertical beam length gives the spread of the reverse voltage within each group (for example, U_b—335, 330,

* Original manuscript received by the IRE, April 2, 1958.
† Siemens-Schuckertwerke AG, Pretzfeld über Forchheim, Oberfranken, Germany.

[1] G. L. Pearson and B. Sawyer, "Silicon *p-n* junction alloy diodes," PROC. IRE, vol. 40, pp. 1348–1351; November, 1952.
[2] K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, vol. 94, pp. 877–884; May 15, 1954.
[3] S. L. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, vol. 99, pp. 1234–1241; August 15, 1955.
[4] The results of D. K. Wilson are cited in McKay, *op. cit.*, Fig. 3.
[5] A. Herlet and H. Patalong, "Die Sperrfähigkeit von legierten Si-Flächengleich-richtern," ("The blocking capability of alloyed silicon rectifying cells") *Z. Naturforsch*, vol. 10a, pp. 584–586; July, 1955.
[6] S. L. Miller, "Ionisation rates for holes and electrons in silicon," *Phys. Rev.*, vol. 105, pp. 1246–1249; February 15, 1957.
[7] W. Shockley and R. C. Prim, "Space-charge limited emission in semiconductors," *Phys. Rev.*, vol. 90, pp. 753–758; June 1, 1953.
[8] G. C. Dacy, "Space-charge limited hole current in germanium," *Phys. Rev.*, vol. 90, pp. 759–763; June 1, 1953.

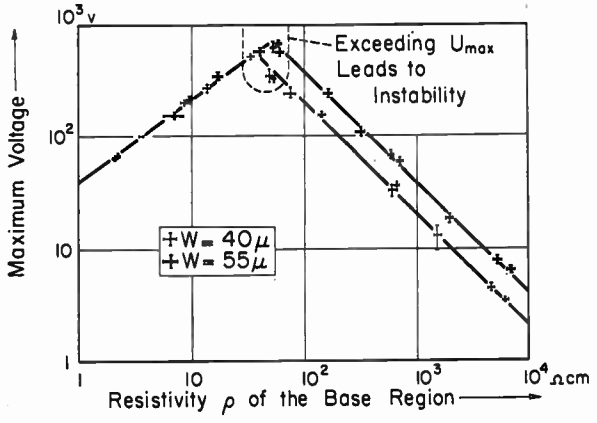


Fig. 2—Maximum blocking voltage of *n-p-n* silicon alloyed transistors (20°C). ✤ extent of scatter of the experimental results.

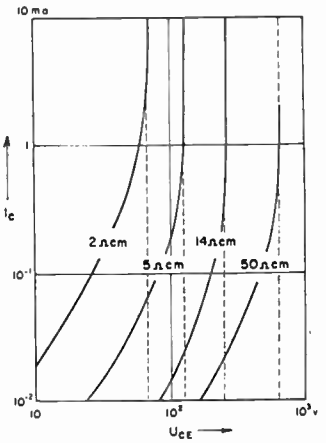


Fig. 3—Characteristic curves of the transistor in the breakdown region. (Direct current measurements.)

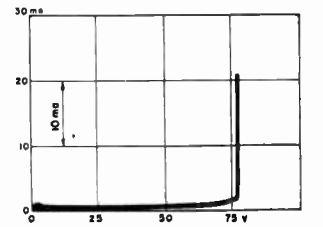


Fig. 4—Characteristic curves of the transistor in the breakdown region. (Taken from the cathode-ray oscilloscope.)

320, 340, 360 volts for 17 ohm-cm group). Naturally, defective transistors have not been considered for the evaluation. The horizontal beam length reveals the uncertainty of the determination of values of ρ. The details of the evaluation are in Section III.

The results given in Fig. 2 agree quantitatively with expectation. The separation of $U_f(\rho)$ values for the two base thicknesses is easily recognized. The upper curve lies a factor of 1.9 over the lower curve. According to (2), this factor corresponds to the difference of the base thicknesses. Both curves have a slope of 45 degrees which implies that the hole mobility μ_p is independent of ρ. On the basis of base thicknesses of 40 μ and 55 μ, a value of 450 cm²/volt-second results. This result is also very reasonable for the resistivity range above 50 ohm-cm.[9]

The course of $U_b(\rho)$ in Fig. 2 is closely approximated by the straight line

$$\frac{U_b}{\text{volts}} = 40\left(\frac{\rho_p}{\text{ohm-cm}}\right)^{3/4}. \qquad (3)$$

This relation is in agreement with the results of Miller within the limits of error.[10]

III. Details of the Evaluation

To determine the maximum blocking voltage in Fig. 2, the steep rise of the collector current in the grounded emitter connection was used. A more exact explanation must be given to supplement this general statement. The current-voltage characteristics of the transistors were namely very different, depending upon which domain of resistivity the base region contains. We are able to distinguish essentially three domains:

1) the domain of pure breakdown ($\rho \ll 50$ ohm-cm);
2) the transition domain with approximately simultaneous occurrence of breakdown and punch-through ($\rho \approx 50$ ohm-cm; and
3) the domain with pure punch-through ($\rho \gg 50$ ohm-cm).

It is relatively easy to establish the maximum blocking voltage in the first domain below 50 ohm-cm. The influence of the surface is exceedingly strong in this domain so that one has very high reverse currents for voltages below U_b. Because of this, the onset of the steep rise of the current in the very sensitive log log plot of Fig. 3 is strongly blurred, but is sufficient to recognize with certainty. For the evaluations in this domain, we have allowed only those characteristic curves which show on a linear plot (Fig. 4) a distinctly recognizable break under 2 ma. The current measurements and the cathode-ray oscilloscope observations were carried out with the short-circuited base-emitter circuit. In this ρ dimain, the full blocking voltage was not attained with the base open. The theoretical explanation was given by Miller.[3]

The outer characteristic curve in Fig. 3 (50 ohm-cm) already belongs to the transition domain of resistivity in which punch-through and breakdown set in at about the same voltage. In this domain the characteristic curves normally do not have the character shown in Figs. 3 and 4. In most cases the course of the voltage is downward after reaching a maximum, as shown in Fig.

[9] The method can give only an approximate value of μ_p. It does not permit a precision determination. The uncertainties in the determination of the base thicknesses are too great for that. The experimental results are, for example, also given within the limits of error by the theoretical curves with the values $W_1 = 57.5$ μ, $W_2 = 42.5$ μ, and $\mu_p = 500$ cm²/volt-second.

[10] We could not definitely clear up the difference between these values and the values of Herlet and Patalong, *op. cit.*, where, perhaps, the reason for the lower breakdown voltage is that the small rectifiers measured by Herlet and Patalong did not have a planar junction. (See H. L. Armstrong, E. D. Metz, and I. Weimann, "Design theory and experiments for abrupt hemispherical *p-n* junction diodes," IRE Trans. on Electron Devices, vol. ED-3, pp. 86–92; April, 1956.)

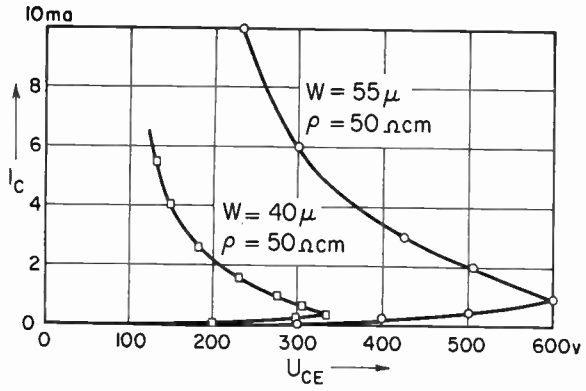


Fig. 5—Characteristic curves of the transistor in the transition region between breakdown and punch-through. (Direct current measurements.)

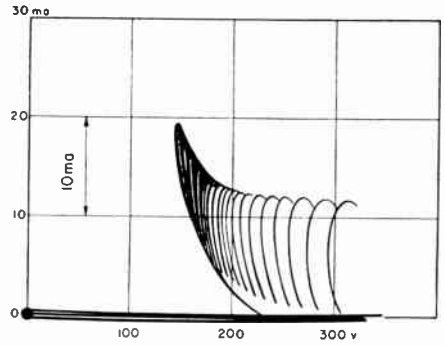


Fig. 6—Uncontrolled oscillations of transistors in the transition region between breakdown and punch-through ($\rho = 50$ ohm-cm, $W = 40$ μ, taken from the cathode-ray oscilloscope).

5 for two transistors with about the same ρ value but different base thicknesses. As observations with the cathode-ray oscilloscope show, these transistors are inclined to unstable oscillations in the region of the falling voltage-current characteristic.[11] An example is given in Fig. 6 for the same transistor for which the static characteristic is given in Fig. 5 ($W = 40$ μ). The kind of oscillations and the onset depend strongly upon the external impedance in the emitter-collector circuit.

In spite of limiting the collector current by a high value of external resistance, a number of the transistors were destroyed upon attaining the instability point. We carried out this measurement with a shorted base-emitter circuit, and took the onset of instability (with or without destruction) as the maximum blocking voltage in Fig. 2. While the voltage at the instability point remains within a narrow region of scatter within a given group of transistors, the collector current at the break point varies in part from 30 μa to 10 ma. This appears to the authors as a hint that thermal effects play no essential role in the instability depicted.

[11] Similar oscillatory phenomena in this domain have been reported by H. Statz and R. A. Pucel, "The spacistor, a new class of high-frequency semiconductor devices," PROC. IRE, vol. 45, pp. 317–324; March, 1957.

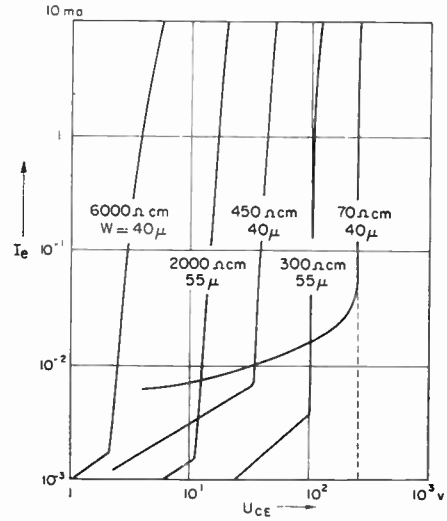


Fig. 7—Characteristic curves of the transistor in the punch-through region. (Direct current measurements.)

We explain the characteristic of Fig. 5 as follows. At the instability point almost all carriers are swept out of the base region because of punch-through. Therefore the transverse resistance of the base is high and prevents the external short circuit from being effective. The situation resembles the case of open base circuit. Thus the voltage decreases with increasing current towards the maximum voltage for open base circuit.

The domain in which instability appears in the case of 55-μ base thickness is between 40 and 70 ohm-cm (520–750 volts); in the case of 40-μ base thickness it is 30–55 ohm-cm (300–550 volts).

Fig. 7 shows the course of the characteristic curves in the third domain of pure punch-through. The steep region is affected very little by surface phenomena, and measurements for base open or shorted to the emitter are practically identical. In contrast to the characteristic curves of Fig. 3, the onset of the steep region can be seen as a very distinct break. However, the steep increase of the current is observed to be practically perpendicular only for the relatively low ρ values. In transistors with high values of base resistance, the current rise from the beginning onward is slanted and follows a nearly constant droop in the log log presentation. The droop is greater the higher the base resistance.

We wish first to refer to the characteristic curves for values of ρ greater than 1000 ohm-cm. First of all, it is in no way clear what value one should put on the maximum voltage for the punch-through effect. The voltage established by the bend in the characteristic from a flat to a steep rise in current is in no way appropriate for this. Namely, one can raise or lower arbitrarily this flat region of the characteristic by surface treatment of the transistor, and therefore, remove this voltage break at will.

Possibility of a physically definite determination is

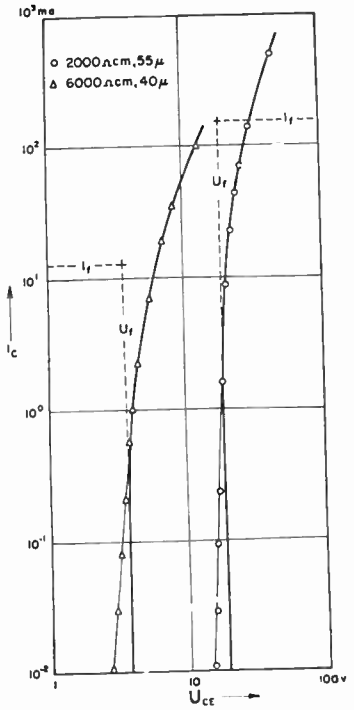


Fig. 8—Determination of U_f in the case of a high resistance base region by comparison with the theory of Shockley and Prim. —— theoretical curve; Δ ○ experimental points.

available if one pursues the course of the characteristic into high collector currents. Here the voltage-current relation is in good agreement with the relation derived theoretically by Shockley and Prim[12] (Fig. 8). At lower currents the characteristic goes over into a tilted decline with a constant slope instead of a perpendicular decrease. The sharp bending of the characteristic from a flat to a steep course in Fig. 7 occurs, therefore, long before the attainment of the punch-through voltage.[13]

We have inserted in Fig. 2 as maximum blocking voltages those values of U_f which result from comparison of measured characteristics with theoretical curves.

In the case of transistors with a base resistivity from about 300 ohm-cm down, the current rise was steep enough to take the bend-over voltage as the limit voltage with good precision. For base resistivity values of

[12] The agreement with Shockley and Prim, *op. cit.*, is better than with the modified characteristic relation of Dacy, *op. cit.* This is understandable since the field strength in the base for the ρ domain in question remains under the value 8000 volts/cm for all loads. Then the dependence of mobility upon field strength need not be taken into account, according to E. J. Ryder, "Mobility of holes and electrons in high electric fields," *Phys. Rev.*, vol. 90, pp. 766–769; June 1, 1953.

[13] A reason for this premature tilted rise can be, for example, that the space-charge zone is indeed not an exactly defined region, but that the boundary is smeared out over a set of several Debye lengths $x_0 = \epsilon kT/4\pi e^2 N_{A-}$. In this way a tail of the space-charge zone on the collector side prematurely punches through to the space-charge zone of the emitter side. Since the Debye length increases with increasing ρ values of the base region and therefore becomes greater relative to the base, it may be understood that the current rise becomes flatter with increasing ρ values of the base. Further, very small-faced narrow points of the base can be contributing to premature increase of current.

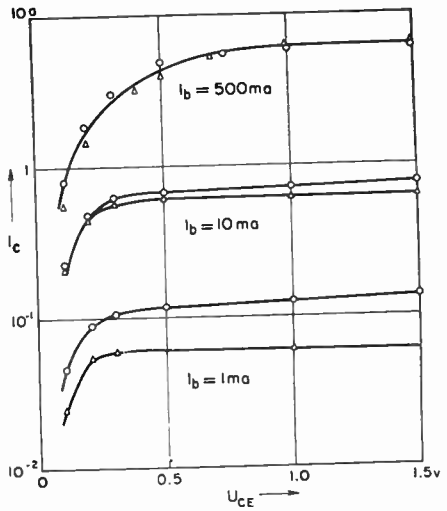


Fig. 9—Family of characteristic curves for the "on condition" in the grounded emitter connection: Δ 9 ohm-cm, 40 μ; ○ 1500 ohm-cm, 40 μ.

300–1000 ohm-cm, it was impossible to determine physically a definite voltage. The evaluation according to Shockley and Prim[7] gave no reasonable results in this instance; on the other hand, the tilting of the characteristic was too large to select the bend-over voltage as the limit voltage. We made a compromise in this case in that we gave the voltage domain for $I_c = 10^{-1} - 10$ ma.

With this we set forth in detail how the voltage values given in Fig. 2 were determined. Now, the determination of the ρ values is described briefly. For the preparation of a transistor group we have cut at times 14 silicon slices from a crystal, and have fashioned the middle 10 into transistors. The first two and last two slices have been alloyed with aluminum at the same time, thereby creating ohmic resistances from which ρ can be calculated from the dimensions and resistance value. The amount of spread in Fig. 2 is the difference of the ρ values between the first two and the last two slices. Values measured in this way agree well with ρ values on the single crystal using other methods; however, the spread of the ρ determination can be sharply reduced with this method.

IV. Concluding Remarks

Just as the forward conduction properties of alloyed rectifiers with high-level injection are independent of the resistivity of the high resistance midregion, so also the family of characteristics of transistors in the "on condition" are, in the main, independent of the base region resistivity.[5,14,15]

This is indeed so in a far reaching sense (Fig. 9). Fine

[14] R. N. Hall and W. C. Dunlap, "*P-n* junctions prepared by impurity diffusion," *Phys. Rev.*, vol. 80, pp. 467–468; November 1, 1950.

[15] R. N. Hall, "Power rectifiers and transistors," Proc. IRE, vol. 40, pp. 1512–1518; November, 1952.

differences naturally appear. So in the case of the transistor with a high base resistance, the rise of I_c with collector voltage[16] is clearly discernible. Further, this transistor has a better current amplification at small collector currents than a transistor with low base resistance. This is plainly due to the difficulty of reducing surface recombination rate in low resistivity starting material.

In the first approximation nevertheless, one can strive for a blocking voltage by specifying the base resistance without regard to the properties of the "on condition."

[16] J. M. Early, "Design theory of junction transistors," *Bell Sys. Tech. J.*, vol. 32, pp. 1271–1312; November, 1953.

V. Acknowledgment

The authors wish to express their thanks to Dr. E. Spenke, who suggested this examination, and to our coworkers H. Benda, Dr. A. Hoffmann, Dr. W. Keller, and Dr. K. Reuschel for supplying the starting silicon with all the desired ρ values. We especially thank K. Recknagel for patiently and carefully making the transistors examined in this work. To Dr. Angello we are greatly indebted for the translation of this paper.

The Effective Emitter Area of Power Transistors*

R. EMEIS†, A. HERLET†, AND E. SPENKE†

Summary—**The theoretical derivation in this paper is concerned first with the determination of an effective emitter area for a transistor with cylindrical geometry. From this it follows that out of the total emitter area only the border region in a width of one diffusion length is effective. This result is applied to a geometry of concentric emitter and base rings, each being considerably broader than two diffusion lengths.**

For comparing theory and experiment in such a geometry, different emitter rim lengths are realized easily when the different base rings are connected successively with one another and the emitter rings are also treated in the same manner. Theory and experiments agree. As a second theoretical contribution, the Appendix contains a rigorous solution for a special three-dimensional transistor model.

I. Introduction

THE physical conditions in a unidimensional transistor model are described by the statement of the junction voltages U_e and U_c. They determine the concentrations of the minority carriers on the emitter and collector boundaries of the base region (see Fig. 1):

$$n(0) = n_p \cdot e^{(e/kT)U_e} \quad (1)$$

$$n(W) = n_p \cdot e^{(e/kT)U_c}. \quad (2)$$

The concentration $n(z)$ remains between these two values on a nearly linear curve,[1] for the case in which the base is sufficiently narrow. The diffusion current of the electrons corresponds to this concentration gradient. When the carrier concentrations are increased, the diffusion current will be supported to a greater extent by a field current, and in the limiting case of strong injection will be directly doubled. Under the operative conditions of a current carrying (not blocking) switch transistor (condition "on"), as shown in Fig. 1,[2] we have

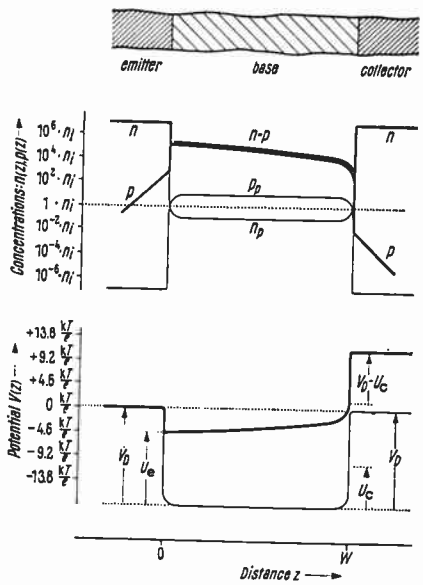


Fig. 1—Concentration and potential relations in an *n-p-n* transistor in the case of strong injection.

* Original manuscript received by the IRE, April 3, 1958.

† Siemens-Schuckertwerke AG, Pretzfeld über Forchheim, Oberfranken, Germany.

[1] In the semilogarithmic plot of Fig. 1 this will *not* produce a straight line.

[2] In order to prevent, from the start, any troublesome complications (which in other respects will produce no substantial consequences), strong injection $p = n \gg p_p$ is assumed in the whole base region, also including the region before the collector. This is not uncommon in case of a switch transistor in a current carrying condition, as it may appear to a communication expert, since the switch transistor is controlled in the saturation range of the characteristic only in its "off" position. In case of an "on" condition the saturation range of the characteristic is preferably avoided, because no further increase of the current can be obtained, and only voltage losses can be expected. But only in the saturation range of the characteristic will the voltage at the collector change to a blocking voltage.

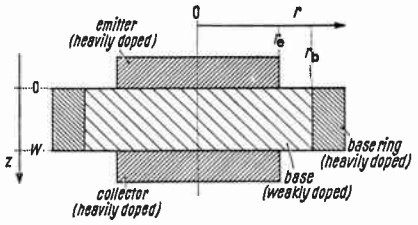


Fig. 2—Cylindrical transistor model.

$$n(0) \gg n(W),$$

so that the concentration gradient of the electrons, and thus the emitter current, will be decisively determined by the value of $n(0)$.

In the unidimensional theory U_e and, thus, $n(0)$ are considered to be constant across the whole emitter. This is permissible also in case of weak injection. But in case of strong injection, the lateral voltage drop which is related to the base current will influence the potential distribution to such an extent that an assumption of a value of U_e which is independent of position is not permissible anymore. On the contrary, the effects of the voltage drop in the lateral direction are of decisive importance for the properties of the transistor.

It is known that the voltage drop in the lateral direction continuously reduces the voltage drop of the junction with increasing distance from the base contact.[3-5] In the case of a circular emitter shape (Fig. 2) U_e and, thus, $n(0)$ will decrease from the edges of the circular surface towards the center. Therefore, the share of the respective single surface elements of the transistor with respect to the total current will be larger in the border region of the emitter surface and will become less towards the center.

To a major extent, the situation in the border region of the emitter which is specifically loaded the greatest is decisive for the properties of a transistor. The current density in this region is, for instance, determining for the thermal load limit of the transistor and has an important influence on the decrease in the (α) value with increased emitter current. It therefore appears meaningful to relate the transistor description to the situation at the emitter border region; *i.e.*, the drop of the current density towards the middle of the emitter surface is taken into account by the definition of an *effective* emitter surface

$$A_{eff} = \frac{I_e}{i_e(r_e)} \int_0^{r_e} \frac{i_e(r)}{i_e(r_e)} \cdot 2\pi r \cdot dr. \qquad (3)$$

This process leads to a particularly clear and practical description.

This question has been treated in the following theoretical section for the case of a circular emitter shape and in particular for the case of strong injection. The calculation is carried out for a simplified model having an ideal emitter. The layer sequence is: $n-s_p-n$, wherein s_p indicates a weak p doping. An experimental section follows, wherein measurements made on silicon transistors with several base and emitter rings are reported.

In the Appendix it is shown that in the case of a special geometry with an "ideal" emitter, collector, and base ring, and with strong injection, it is possible to carry out a rigorous three-dimensional theory of the transistor.

II. Theory

In Section II-A the conditions in the base region directly in front of the emitter junction are considered. Based on the determinations made in Section I, a calculation of the effective emitter area is made possible in Section II-B.

A. Concentrations, Potential, and Current Distribution Directly Adjacent to the Emitter Junction

Our considerations are based upon a cylindrical transistor model, as shown in Fig. 2. As mentioned above, we are limiting ourselves to the case of strong injection:

First, we will fulfill the Poisson equation in the neutral section of the base region by the following:

$$n(r,z) \approx p(r,z) \quad \left(\gg p_p \approx n_A^- > n_p = \frac{n_i^2}{p_p} \right), \qquad (4)$$

wherein p_p and n_p indicate the equilibrium densities of the holes and the electrons in the zero current condition, due to the doping n_A^- of the base region. n_i is the intrinsic density of the respective semiconductor.

Second, the recombination surplus $\mathfrak{R}$ is based upon the linear relation $(n-n_p)/\tau$ which is valid first of all for weak injection, and subsequently with another τ value also for strong injection.[6,7]

Third, the value of $-n_p$ in the numerator is disregarded, since we intend to use the relation only for strong injection, and because the value of $-n_p$ according to (4) will be negligible as compared with the value of n we obtain:

$$\mathfrak{R} = \frac{n}{\tau}. \qquad (5)$$

But, because of this condition, in future equations the transition to the thermal equilibrium will not be permissible, *i.e.*, to the zero current condition of the transistor, and finally, the transition to "all voltages $U \to 0$."

[3] J. M. Early, "Design theory of junction transistors," *Bell Sys. Tech. J.*, vol. 32, pp. 1271–1312; November, 1953.
[4] N. H. Fletcher, "Some aspects of the design of power transistors," Proc. IRE, vol. 43, pp. 551–559; May, 1955.
[5] R. Emeis and A. Herlet, "The effective emitter area of power transistors," *Z. Naturforsch.*, vol. 12a, pp. 1016–1018; December, 1957.
[6] With reference to weak injection compare W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.
[7] With reference to strong injection compare R. N. Hall, "Power rectifiers and transistors," Proc. IRE, vol. 40, pp. 1512–1518; November, 1952.

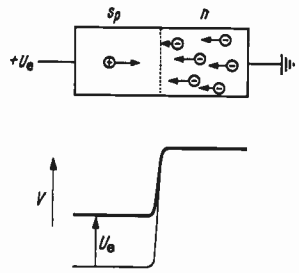


Fig. 3—An s_p-n junction. In order that the electrons can flow from an n region into an s_p region, the potential in the s_p region must be *increased*.

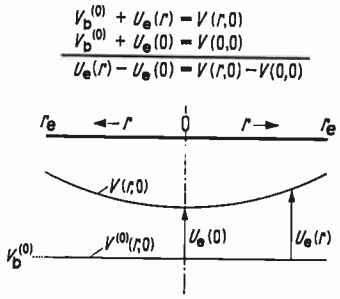


Fig. 4—The radial distribution of the potential increase in the s_p-doped base in front of the emitter junction.

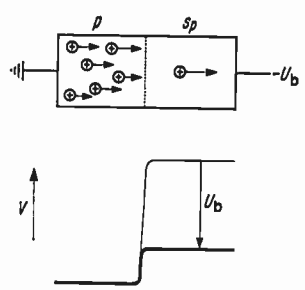


Fig. 5—A p-s_p junction. In order that holes can flow from a p region into an s_p region, the potential in the s_p region must be *reduced*.

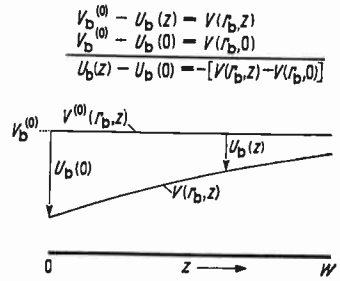


Fig. 6—The axial distribution of the potential reduction in the s_p-doped base in front of the base junction.

We first consider only a narrow band of the base region in front of the emitter junction $z=0$. The electron concentration n is, in this case, increased relative to the equilibrium value n_p by a coefficient $\exp(eU_e/kT)$.[8] Similar to (4),

$$n(r, 0) = n_p \cdot e^{(e/kT)U_e(r)} \tag{6}$$

must be valid for all r values between 0 and r_e. Also,

$$n(r, 0) = n(0, 0) \cdot e^{e/kT[U_e(r)-U_e(0)]}. \tag{7}$$

Due to the high conductivity of the highly doped emitter, potential differences of a local nature on the emitter side of the junction cannot be formed. The radial variation $U_e(r) - U_e(0)$ of the voltage U_e can be formed only by an equal radial variation of the potential V on the base side of the junction (compare also Figs. 3 and 4):

$$U_e(r) - U_e(0) = V(r, 0) - V(0, 0), \tag{8}$$

and from (7) it follows that

$$n(r, 0) = n(0, 0) \cdot e^{+e/kT[V(r,0)-V(0,0)]}. \tag{9}$$

The electrons will therefore arrange themselves in the radial direction in accordance with the Boltzmann law. By radial differentiation it follows that:

$$\frac{\partial}{\partial r} n(r, 0) = \frac{e}{kT} n(0, 0) e^{e/kT[V(r,0)-V(0,0)]} \frac{\partial V(r,0)}{\partial r}$$

$$= -\frac{e}{kT} n(r, 0) E_r(r, 0), \tag{10}$$

where $E_r(r, 0)$ is the radial component of the field strength E immediately before the emitter junction $z=0$. Therefore the radial component

$$+e \cdot \mu_n \frac{kT}{e} \cdot \frac{\partial}{\partial r} n(r, 0)$$

of the electronic *diffusion* current density will be entirely compensated by the radial component $+e\mu_n n(r, 0) \cdot E_r(r, 0)$ of the electronic *field* current density. Or it follows necessarily from the Boltzmann equilibrium equation (9) that the radial component $i_{nr}(r, 0)$ of the electron current density goes to zero:

$$i_{nr}(r, 0) = e\mu_n n(r, 0) \cdot E_r(r, 0) + \mu_n kT \frac{\partial}{\partial r} n(r, 0) = 0. \tag{11}$$

In the case of holes, based on (9), the field and diffusion components will be equal also. But these are added to each other, since in case of the diffusion current density

$$-e\mu_p \cdot \frac{kT}{e} \cdot \frac{\partial}{\partial r} p(r, 0)$$

a change of the sign will take place relative to the electrons. Thus, for the radial current density component of the holes, the following relation is obtained:

$$i_{pr}(r, 0) = -2\mu_p kT \frac{\partial}{\partial r} p(r, 0). \tag{12}$$

When the doping of the emitter region is sufficiently heavy—and we assume such a heavy doping to be present—the hole injection into the emitter can be disregarded:

$$i_{pz}(r, 0) = 0. \tag{13}$$

[8] Compare, for instance, E. Spenke, "Electronic Semiconductors," McGraw-Hill Book Co., Inc., New York, N. Y., p. 100; 1958.

Finally, due to the radial symmetry of the whole arrangement,

$$i_{p\phi}(r, 0) = 0. \tag{14}$$

The continuity equation for the hole current density i_p is thus[9]

$$\operatorname{div} \vec{i_p} = \frac{1}{r} \cdot \frac{\partial}{\partial r} (r i_{pr}) + \frac{1}{r} \cdot \frac{\partial}{\partial \phi} i_{p\phi} + \frac{\partial}{\partial z} i_{pz} = - e\mathfrak{R}, \tag{15}$$

and with the help of (12) to (14) and with (5) and (4) we obtain

$$\frac{1}{r} \frac{\partial}{\partial r} \left\{ r \frac{\partial}{\partial r} p(r, 0) \right\}$$

$$= + \frac{e}{2\mu_p \mathrm{k} T \tau} p(r, 0) = \left(\sqrt{\frac{\mu_n}{\mu_n + \mu_p}} \frac{1}{L} \right)^2 p(r, 0). \tag{16}$$

Here L is the "diffusion length for strong injection[7, 10]"

$$L = \sqrt{2 \frac{\mu_n \cdot \mu_p}{\mu_n + \mu_p} \cdot \frac{\mathrm{k}T}{e} \tau}. \tag{17}$$

The solution of (16) which is adjusted to the physical conditions is, using the Bessel function of the zero order,

$$p(r, 0) = p(0, 0) \cdot J_0 \left(\mathrm{j} r \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} L \right). \tag{18}$$

Due to (4), the same radial distribution function is also valid for the electrons

$$n(r, 0) = n(0, 0) \cdot J_0 \left(\mathrm{j} r \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} L \right), \tag{19}$$

where j is the imaginary unit $\sqrt{-1}$.

B. The Effective Emitter Area

The computation of the emitter current is actually possible only after a complete determination of the concentration and potential distributions is available. But from the treatment of unidimensional models,[6,11,12] we know that the emitter current is proportional (with good approximation) to the concentration increase at the emitter. It can be assumed, therefore,[13] that the respective current share of an annular element $(r, r+dr)$ of the emitter is proportional to the increased concentration $n(r, 0)$ of this element, and we can thus assume that

$$i_e(r, 0) \sim n(r, 0), \tag{20}$$

or using (19)

$$i_e(r, 0) = \text{constant} \cdot J_0 \left(\mathrm{j} r \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} \cdot L \right). \tag{21}$$

If we now utilize the defining equation (3) of the effective emitter area A_{eff}, we obtain, with the help of the known integration rule for the Bessel functions,

$$\int \chi \cdot J_0(\chi) \cdot d\chi = \chi J_1(\chi) \tag{22}$$

and thus

$$A_{eff} = 2 \cdot \pi \cdot r_e \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} L \cdot \frac{-\mathrm{j} J_1 \left(\mathrm{j} r_e \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} L \right)}{J_0 \left(\mathrm{j} r_e \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} L \right)}. \tag{23}$$

For small and large emitter surfaces the following limit values are obtained from (23):

$$A_{eff} = \pi r_e^2 \qquad \text{for } r_e \ll \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} \cdot L \tag{24}$$

$$A_{eff} = 2\pi r_e \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} \cdot L \text{ for } r_e \gg \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} \cdot L. \tag{25}$$

These equations indicate that in the case of small radius the whole geometrical emitter surface, and in the case of large radius only the outer rim of the emitter, is active in a width

$$\sqrt{\frac{\mu_n + \mu_p}{\mu_n}} L.$$

The A_{eff} dependency upon r_e is shown in Fig. 7. It is evident that the boundary of validity between (24) and (25) is at

$$r_e = 2 \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} \cdot L. \tag{25a}$$

III. Experimental Results

The experimental problem resulting from the determinations in the previous section, consists in a quantitative determination of the value of A_{eff} as a function of r_e. But we will limit ourselves to a much simpler problem of a *relative* comparison of the effective emitter surfaces of transistors of different sizes. As a

[9] This equation for the unidimensional case was derivated, for instance, by A. Herlet and E. Spenke, "Rectifier with *p-i-n*, or *p-s-n* structure under direct current load," *Z. angew. Phys.*, vol. 7, pp. 99–107, 149–163, and 195–212; February–April, 1955. See particularly equation (27).

[10] A. Herlet, "Determination of the diffusion length L and of the inversion density n_i by means of forward characteristic of alloyed silicon surface rectifiers," *Z. angew. Phys.*, vol. 9, pp. 155–158; April, 1957.

[11] Hall, *op. cit.*, (16) and (17).

[12] F. H. Stieltjes and L. J. Tummers, "The transistor at high density," *Phillips tech. Rundschau*, vol. 18, pp. 44–51; February, 1956.

[13] The three-dimensional solution developed in the Appendix, which is rigorous under certain conditions, confirms this assumption.,

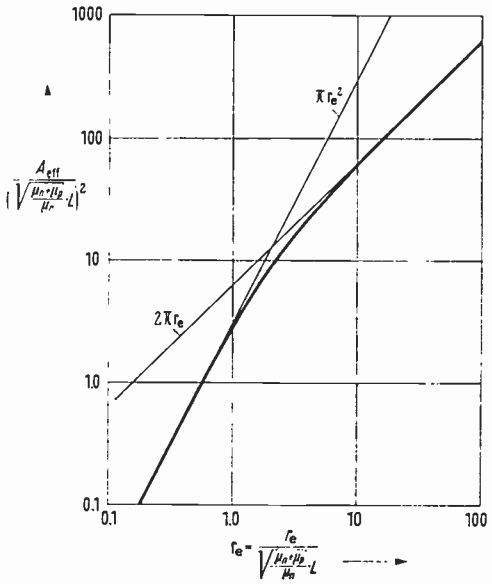


Fig. 7—The relation between the effective emitter area and the emitter radius in circular emitter electrodes.

measure for different effective emitter areas we consider the emitter currents of transistors of different sizes, since, as was determined from the defining equation (3),

$$I_e \sim A_{eff}, \tag{26}$$

as long as the emitter current density $i_e(r_e)$ is kept constant; or expressed differently, if the transistors which are compared with each other are loaded to the same extent. The determination of the "equal loading conditions" is rather problematic since the current density $i_e(r_e)$, as well as the corresponding junction voltage $U_e(r_e)$, cannot be directly measured. We must, therefore, try to determine the loading condition of the transistor by a quantity which can be measured directly; even if we are not able to define this condition exactly. For instance by the voltage U_{eb}, or by the current amplifying coefficient α, the value of which, as is known, depends substantially on the load conditions.

By keeping constant the magnitudes of α and of U_{eb}, we eliminate automatically another problem; namely, the material constant L is actually not a constant, but depends to a substantial extent upon the load.[10] But if the load is kept constant when comparing the transistors of different sizes, L will not vary, and in the range of validity of the approximate solution (25) it will follow from (26) for transistors with a circular emitter that

$$I_e \sim r_e. \tag{27}$$

As is known in the art, the value of L in an alloyed silicon rectifier[10] is in the order of magnitude of 0.1 mm in the case of strong injection. It therefore is expected that similar values will be obtained in the case of silicon transistors. Then out of (25a) it follows that we operate actually in the range of validity of the approximate solution (25) if the diameter of the emitter is greater than 0.2 mm, and we must, therefore, expect an increase of I_e

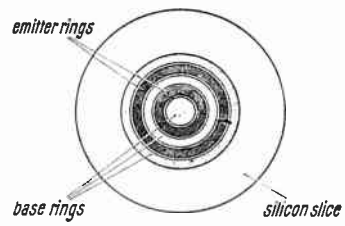


Fig. 8—An annular base-emitter pattern (schematic).

in accordance with (27). This could be confirmed also for transistors with a circular emitter surface in the range of 0.4 mm $< r_e <$ 2.5 mm.[14]

Using a relatively general formulation it can be said that the effective emitter area is proportional to the length of the rim of the emitter. In transistors with a circular emitter shape, the length of the rim of the emitter electrode, and thus the effective emitter area is relatively small. Larger rim lengths can be obtained when, instead of the above described circular transistor, a base-emitter pattern is alloyed with annular layers disposed one inside the other; for instance, in the manner shown in Fig. 8 wherein only the inner and outer rims of each emitter ring are portions of the effective emitter surface.[15]

The linear variation of the emitter current with the rim length (27) can be satisfactorily tested on transistors of this type. First, the middle base spot and the No. 1 emitter ring are contacted, so that only the inner rim of the No. 1 emitter ring is active. In the second step the No. 2 base ring is contacted, so that in this case the inner and outer rim of the No. 1 emitter ring are active. By switching on stepwise the No. 2 emitter ring, the No. 3 base ring, the No. 3 emitter ring etc., the emitter area can be increased stepwise. It must be assumed here that the width of the emitter ring is large relative to twice the length

$$\sqrt{\frac{\mu_n + \mu_p}{\mu_n}}\, L,$$

so that the inner and outer rims of each emitter ring can be active independent of each other.

The results of such an experiment on alloyed silicon *n-p-n* transistors with 3 emitter and 4 base rings is shown in Fig. 9. In this case the values of the currents are compared at the same value of the emitter base voltage $U_{eb} = 0.9$ volt. This value of U_{eb} corresponds approximately to the current amplification coefficient $\alpha \approx 8$ to 10. Fig. 9 shows the results of measurements for three transistors. Data are differentiated by the use of different designations for measured points. The I_e-straight lines for two of the transistors practically overlap each other. These collapsing lines for two transistors correspond substantially with the straight line shown

[14] Emeis and Herlet, *op. cit.*, in these experiments α was kept constant as a measure of the load.

[15] Another arrangement with which it is possible to obtain large emitter rim lengths is the grid specimen shown by Hall, *op. cit.*

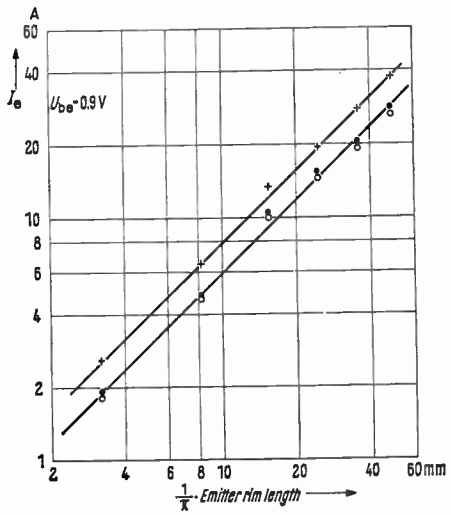


Fig. 9—The relation between the emitter current and the emitter periphery in the case of an alloyed silicon n-p-n transistor having a circular base-emitter pattern.

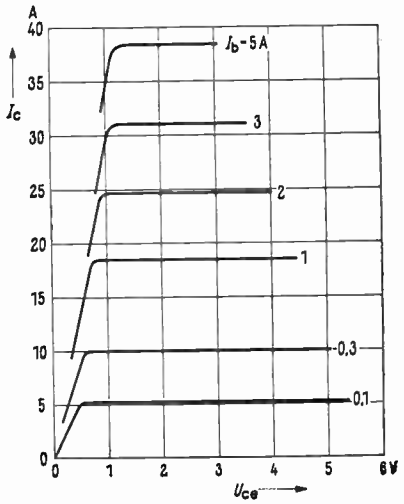


Fig. 10—Characteristics of an alloyed silicon n-p-n transistor having a periphery of 15 cm.

in Fig. 3 of Emeis and Herlet[5] for circular emitters and for a value of $\alpha = 10$.[16] The straight line for the No. 3 transistor in Fig. 9 is slightly higher. Such dispersions are not unusual due to the strong dependence of the emitter current upon U_{eb}. In this type of transistor we may, for instance, by reducing the value of U_{eb} to only 0.85 volt, obtain the same current values as for the other two transistors at 0.9 volt. In some of the transistors measured, the increase of I_e with the increase of the rim length is slower than linear. This flattening is relatively very small. Whether this flattening or this deviation from the linearity rule (27) is due to the imperfections in the production of the transistors or is due to other, deeper causes we do not know.

The full rim length of the experimental transistors with 3 emitter rings is 15 cm. This would correspond to a circular emitter surface having a diameter of 4.8 cm. Assuming the value of $L = 100\ \mu$, the effective emitter surface of transistors of this type would have, in accordance with (25), approximately 20 mm². The characteristics of a transistor of this type in the "on-condition" are shown in Fig. 10.

Appendix I

A "Rigorous" Solution for a Special Three-Dimensional Transistor Model

A. The Boundary Values for the Collector Junction

The determinations from (4) can be applied analogously to the collector. Also in this case the electrons arrange themselves in a radial direction in a Boltzmann distribution, and therefore, the radial electron current density i_{nr} will disappear due to the compensation of the diffusion portion by the field current

$$i_{nr}(r, W) = 0. \tag{28}$$

[16] Emeis and Herlet, *op. cit.*, in order to obtain a better comparison with Fig. 3 of Emeis and Herlet and Fig. 9 of this paper, the emitter length divided by π is plotted as abscissa.

In the case of holes, the field current will again be twice the diffusion current in the radial direction:

$$i_{pr}(r, W) = -2\mu_p kT \frac{\partial}{\partial r} p(r, W). \tag{29}$$

The hole current density will again disappear in the axial direction, since a high degree of doping is assumed also in case of the collector, so that the share of the holes in the collector current can be disregarded (collector yield coefficient $\gamma_c = 1$):

$$i_{pz}(r, W) = 0. \tag{30}$$

Due to the radial symmetry, the following equation is valid for the collector

$$i_{p\phi}(r, W) = 0. \tag{31}$$

Then the continuity equation is again limited to the sole remaining radial component i_{pr}, and thus leads to the solution

$$n(r, W) = p(r, W) = n(0, W) \cdot J_0\left(jr \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_n}} L\right). \tag{32}$$

B. The Boundary Values for the Base Junction

At the cylinder surface $r = r_b$, $0 < z < W$ the highly doped base ring passes into the weakly doped base region. Here just in front of the "base junction" determinations similar to those in Section II-A can also be made. Because the highly doped region in this case is not n-doped as at the emitter, but is p-doped, the electrons and the holes exchange roles as compared with Section II-A. Further since the base junction extends in the axial direction, while the emitter junction extends in the radial direction, the axial and radial directions will also exchange their roles.

In detail, according to Figs. 5 and 6 instead of (8), the valid relation will be

$$U_b(z) - U_b(0) = -[V(r_b, z) - V(r_b, 0)]. \tag{33}$$

The result of this will be that now not the electrons according to (9), but the holes will have a Boltzmann distribution along the junction under consideration, and we obtain.

$$p(r_b, z) = p(r_b, 0) \cdot e^{-e/kT[V(r_b,z)-V(r_b,0)]}. \quad (34)$$

In this case, for the *holes*, the axial field current will compensate the axial diffusion current, while for the *electrons*, the axial field current will double the axial diffusion current. Analogous to (11) and (7) it follows that

$$i_{pz}(r_b, z) = 0 \quad (35)$$

$$i_{nz}(r_b, z) = 2\mu_n kT \frac{\partial}{\partial z} n(r_b, z). \quad (36)$$

Due to the high degree of p doping of the base region, the electron share in the base current can be disregarded; therefore, the electron current laterally through the base junction will disappear, so that

$$i_{nr}(r_b, z) = 0. \quad (37)$$

Due to the cylindrical symmetry, naturally, we again obtain

$$i_{n\phi}(r_b, z) = 0. \quad (38)$$

Thus the continuous equation[17] for the electrons

$$\operatorname{div} \vec{i_n} = \frac{1}{r}\frac{\partial}{\partial r}(r i_{nr}) + \frac{1}{r}\frac{\partial}{\partial \phi} i_{n\phi} + \frac{\partial}{\partial z} i_{nz} = e\Re, \quad (39)$$

reduces to an axial differential equation for the electron concentration n:

$$\frac{\partial^2}{\partial z^2} n(r_b, z) = + \frac{e}{2\mu_n kT\tau} n(r_b, z). \quad (40)$$

When using now the prevailing "common" diffusion length in the case of strong injection, according to (17) we obtain

$$n(r_b, z) = \frac{n(r_b, 0) \sinh\left[(W - z) \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_p} L}\right] + n(r_b, W) \sinh\left[z \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_p} L}\right]}{\sinh\left[W \Big/ \sqrt{\frac{\mu_n + \mu_p}{\mu_p} L}\right]}. \quad (41)$$

C. A Special Three-Dimensional Transistor Model and Its "Exact" Concentration and Potential Distributions

By means of (11) to (14) in Section II-A it could be determined that the electron current in front of the emitter is purely axial and that the direction of the hole current is purely radial. The same applies to the plane $z = W$ in front of the collector according to (28) to (31). In front of the base junction the electrons and the holes exchange their roles, because the base ring, in contrast to the emitter and the collector, is highly p-doped and not strongly n-doped. But the axial and radial directions will be exchanged here too, relative to the emitter and collector, so that (35) to (38) will again indicate that at the base junction the electron current is again directed purely axially and the hole current purely radially.

It must therefore be asked, if there are arrangements and conditions under which it will be possible to have a purely axial electron flow and a purely radial hole flow cooperating with each other in the whole base region, and not only on parts of its surface. There are three reasons which may induce the electron flow to deviate from a straight line flow axially from the emitter to the collector. These reasons are:

1) an electron demand in the base ring.
2) a surface recombination on the free surfaces of the base region between the emitter and the base ring and between the base ring and the collector, and
3) a "too large" cross section of the base region in the middle between the emitter and collector. In the case $r_b > r_e$, and already for $r_c > r_e$, the electrons cannot flow purely axially any more.

Reason 1 is already eliminated by the assumption that the p doping of the base ring is very strong. In this case the electron component of the base current will disappear: "the γ_b factor of the base current will become unity." Reasons 2 and 3 will be eliminated simultaneously by the assumption that $r_c = r_b = r_e$. Instead of the transistor model of Fig. 2, we consider in the following a model given in Fig. 11, opposite.

In this model the electrons flow, in all exactness, purely axially and the holes purely radially (Fig. 12). But before writing down this solution and verifying it, we try to eliminate U_c and U_c,[18] and to replace them by the voltages U_{eb} between the emitter and base and U_{cb} between the collector and base.

For the electron concentration in front of the highly n-doped emitter, (6) is valid. In particular, on the rim $r = r_e$ we have

$$n(r_e, 0) = n_p \cdot e^{(e/kT)U_e(r_e)}. \quad (42)$$

In front of the highly p-doped base ring a relation similar to that of (6) is valid, which we write out, assuming $z = 0$, as follows:

$$p(r_e, 0) = p_p \cdot e^{(e/kT)U_b(0)}. \quad (43)$$

[17] This continuity equation can be found for the unidimensional case, for instance, in Herlet and Spenke, *op. cit.* See in particular (26).

[18] U_c and U_c are not accessible experimentally in contrast to U_{eb} and U_{cb}.

$$g(z) = \frac{\exp\left(\frac{e}{kT}\cdot\frac{U_{eb}}{2}\right)\sinh\left[(W-z)\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_p}}L\right] + \exp\left(\frac{e}{kT}\cdot\frac{U_{eb}}{2}\right)\sinh\left[z\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_p}}L\right]}{\sinh\left[W\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_p}}L\right]} \cdot \quad (50)$$

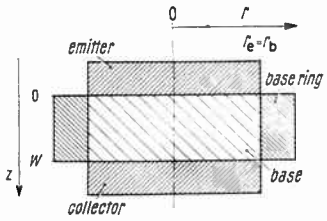


Fig. 11—A simplified cylindrical transistor model.

Multiplying these two equations we obtain

$$n(r_e, 0)\cdot p(r_e, 0) = n_p\cdot p_p\cdot e^{e/kT[U_e(r_e)+U_b(0)]}. \quad (44)$$

Due to the assumption of (4) of a strong injection, we can write on the left-hand side simply $n^2(r_e, 0)$. On the right-hand side we can write n_i^2 for the product of the equilibrium densities n_p and p_p according to the mass action rule. Finally, the distance at $r=r_e$ and $z=0$ between the circular shaped emitter and the base ring is negligibly small. Hence, the voltage drop U_{eb} between the emitter and the base ring is simply equal to the sum of the voltage drops $U_e(r_e)$ at $r=r_e$ of the emitter junction and the voltage drop $U_b(0)$ at $z=0$ of the base junction. Because of the above relations, we obtain from (44)

$$n(r_e, 0) = n_i\cdot e^{(e/kT)(U_{eb}/2)}. \quad (45)$$

At the rim of the collector, where $r=r_e$ and $z=W$, we have accordingly

$$n(r_e, W) = n_i\cdot e^{(e/kT)\cdot(U_{eb}/2)}, \quad (46)$$

whereby U_{eb} is now the voltage between the collector and the base ring.

It will now be verified that the following equation is a "rigorous" solution of the problem—naturally under the high injection conditions of (4) and (5):

$$n(r, z) = p(r, z) = n_i\cdot f(r)\cdot g(z) \quad (47)$$

$$V(r, z) = \frac{kT}{e}\ln\frac{f(r)}{g(z)}\cdot \quad (48)$$

Herein $f(r)$ and $g(z)$ are the following functions:

$$f(r) = \frac{J_0\left(jr\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_n}}L\right)}{J_0\left(jr_e\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_n}}L\right)} \quad (49)$$

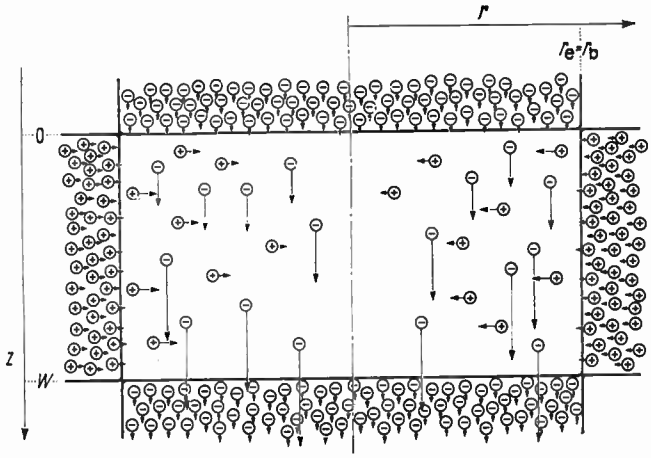


Fig. 12—A schematic representation of the flow of electrons and holes in a transistor model according to Fig. 11.

The solutions (47) to (50) must now be verified. First, it can be seen that due to the product form of (47), the radial increase of the concentration from the middle, $r=0$, to the rim, $r=r_e$, is independent of z. This increase repeats itself in each plane, $z=$constant, in the same manner as follows:

$$n(r, z) \sim J_0\left(jr\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_n}}L\right). \quad (51)$$

Due to the fact that this determination is also valid for the planes $z=0$ and $z=W$, the boundary conditions (19) and (32) will be fulfilled by the solution of (47) to (50). Correspondingly, the axial diminution of the concentration (see Fig. 1) is independent of r, and repeats in the same manner on each parallel to the cylinder axis $r=0$:

$$n(r, z) \sim \exp\left(\frac{e}{kT}\cdot\frac{U_{eb}}{2}\right)\sinh\left[(W-z)\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_p}}L\right] + \exp\left(\frac{e}{kT}\cdot\frac{U_{eb}}{2}\right)\sinh\left(z\Big/\sqrt{\frac{\mu_n+\mu_p}{\mu_p}}L\right). \quad (52)$$

Since this determination is also valid for the cylinder surface $r=r_e$ and furthermore, considering $r_b=r_e$ according to Fig. 11 and (45) and (46), the fulfillment of (41) is assured.

The solution of (47) to (50) thus has the required boundary values. The solution of the differential equations within the volume is confirmed in the following manner from (47) and (48):

$$n(r,z) = n_i \cdot g^2(z) \cdot e^{+(e/kT)V(r,z)} = \text{const}\, e^{+(e/kT)V(r,z)} \quad \text{for } z = \text{constant} \tag{53}$$

or

$$p(r,z) = n_i \cdot f^2(r) \cdot e^{-(e/kT)V(r,z)} = \text{const}'\, e^{-(e/kT)V(r,z)} \quad \text{for } r = \text{const}'. \tag{54}$$

Everywhere in the base region the electron concentration n in the radial direction is now in Boltzmann equilibrium, but the holes p are in axial direction. This means that everywhere in the base region the radial current density i_{nr} of electrons will disappear as will the axial current density i_{pz} of holes. Due to the axial symmetry, $i_{n\phi} = i_{p\phi} = 0$ anyway. From the continuity equation of electrons

$$\text{div}\, \vec{i_n} = + e \cdot \mathfrak{R}, \tag{55}$$

we have remaining only

$$\frac{\partial}{\partial z} i_{nz}(r,z) = + e \frac{n(r,z)}{\tau}. \tag{56}$$

and from the continuity equation of the holes

$$\text{div}\, \vec{i_p} = - e \mathfrak{R}, \tag{57}$$

we have remaining only

$$\frac{1}{r} \cdot \frac{\partial}{\partial r} [r \cdot i_{pr}(r,z)] = - e \frac{p(r,z)}{\tau}. \tag{58}$$

From the Boltzmann equilibrium (54), it follows that, in case of holes, the axial field current will just compensate the axial diffusion current. Since $p = n$ it follows also that in case of electrons the field current will be twice the diffusion current. We thus obtain from (56) that

$$\frac{\partial^2}{\partial z^2} n(r,z) = + \frac{\mu_p}{\mu_n + \mu_p} \cdot \frac{n(r,z)}{L^2}. \tag{59}$$

This equation will be fulfilled by (47) together with (50). The r dependence will be eliminated due to the fact that the coefficient $f(r)$ will be compensated on both sides. The fulfillment with respect to z can be proved simply by differentiating.

Utilizing analogous considerations, the fulfillment of (58) can also be obtained. Thereby it will be possible to solve the continuity equations of the electrons and holes in the whole base volume $0 < r < r_c$ and $0 < z < W$, by the suggested solution of (47) to (50).

There remains only the final confirmation of the assumption (48) made for the potential $V(r, z)$. The potential $V(r, z)$ not only occurs in the continuity equation (55) and in (57), for which $\vec{i_n}$ and $\vec{i_p}$ were just determined, but this potential must also fulfill the Poisson equation

$$\text{div grad}\, V = - \frac{4\pi}{\epsilon} \rho. \tag{60}$$

A "rigorous" transistor theory as described in this Appendix, in the same manner as all other rectifier and transistor theories, subdivides the n-s_p-n structure under consideration into junctions comprising space charges and neutral zones disposed in between said junctions. With regard to these neutral zones only an approximate solution of the Poisson equation (60) can be obtained, in the following sense: due to (60), the field strength divergency which is present, div grad V, is accompanied necessarily by a certain space charge ρ. If it is possible to produce this ρ by such a small deviation Δn from neutrality $n = p$,[19] so that

$$\Delta n \ll n = p, \tag{61}$$

then one will say that the Poisson equation is solved approximately. But in case div grad V requires a value of Δn which is of the same order as the concentrations n and p themselves, then the Poisson equation with $n = p$ is not satisfied even approximately.

For the purpose of confirmation, the divergence div grad V must be formed from (48) for the potential. It can be shown that this divergency is negative.[20] Instead of neutrality $n = p$[19] one should take n as being larger than the hole concentration p. As can be determined by a more thorough calculation, we obtain the maximum value of Δn in the middle and directly in front of the collector junction; *i.e.*, for $r = 0$ and $z = W$, and then we have

$$\frac{\Delta n}{p_p} = \left(\frac{x_0}{W}\right)^2 \cdot \exp\left(\frac{e}{kT} U_{ec}\right), \tag{62}$$

wherein the so-called Debye length will be

$$x_0 = \sqrt{\frac{\epsilon}{4\pi e p_p} \cdot \frac{kT}{e}} = \sqrt{\frac{\epsilon \mu_p}{4\pi \sigma_p} \cdot \frac{kT}{e}}. \tag{63}$$

In p silicon ($\epsilon = 12$, $\mu_p = 500$ cm²/volt/second) with a conductivity $\sigma_p = 1/20\Omega$cm at room temperature ($kT/e = 26$ millivolts) the order of magnitude is

$$x_0 = 0{,}1\mu, \tag{64}$$

and if we assume the thickness of the base to be

$$W = 10\mu, \tag{65}$$

we will obtain

$$\frac{\Delta n}{p_p} = 10^{-4} \cdot \exp\left(\frac{e}{kT} U_{ec}\right). \tag{66}$$

The whole determination which is carried out in this article is based on the "on condition" of a power transistor, in which the saturation of the transistor characteristic is avoided as far as possible, because saturation

[19] In the sense of approximation (4) of strong injection the doping n_A^- can be disregarded in this neutrality condition.

[20] In case of a potential curvature in the r direction according to Fig. 4, as well as in the z direction according to Fig. 1, a *negative* space charge is required.

will not bring any increase in the forward current. On the other hand the current then will flow with a steadily increasing inner voltage loss in the transistor. That is, U_{cc} should not go too high. A value of $U_{cc}=6(kT/e)$ ≈ 0.156 volt should be considered here[21] to be rather high. But even in this case we will have

$$\exp\left(\frac{e}{kT}\ U_{cc}\right) \approx 400, \tag{67}$$

and according to (66) there will still remain

$$\frac{\Delta n}{p_p} < 10^{-4}\cdot 400 = 4\cdot 10^{-2}. \tag{68}$$

[21] In case of real transistors, as shown in the characteristics in Fig. 10, the voltages U_{cc} which are necessary for the saturation of the transistor characteristics are substantially larger than the indicated $6kT/e=156$ millivolts. These voltages are actually approximately 1 volt. But in the scope of the theoretical model considered herein U_{cc} values of $6kT/e$ are more than sufficient to obtain the saturation.

But, actually, the comparison of Δn with p_p is much too restrictive. In the condition of strong injection the concentration will be

$$n \approx p \gg p_p \tag{69}$$

and for an approximate solution of (60) it will be sufficient to have

$$\frac{\Delta n}{p} \ll 1. \tag{70}$$

Based on (69), inequality (70) is satisfied *a fortiori* because the more restrictive inequality (68) is satisfied.

So we have seen that the approximate satisfaction of the Poisson equation (60) in the "on condition" of a power transistor can be assured.

ACKNOWLEDGMENT

The authors are greatly indebted to Dr. Angello who translated this paper.

The Electrical Characteristics of Silicon *P-N-P-N* Triodes*

I. M. MACKINTOSH†, MEMBER, IRE

Summary—**An investigation is made of the electrical characteristics of three-terminal silicon *p-n-p-n* structures, where electrical contact is made to both outer (emitter) regions and to one of the inner (base) regions.**

When the base current I_b is zero, the *V-I* characteristic naturally is identical to the two-terminal case, *i.e.*, ranges of high and low impedance separated by a region of differential negative resistance. Base current supplied from an independent, external circuit is found to decrease the breakover (peak) voltage and to decrease the turn-off current, *i.e.*, the current at which the device enters the low impedance state. In fact, the *p-n-p-n* triode is found to exhibit switching properties closely analogous to the conventional thyratron.

As an extension of earlier work, a general analysis of four-region structures is presented and is applied specifically to the *p-n-p-n* triode. Much of the detailed behavior of the device can be explained in terms of this analysis, and theoretical curves are given which are in good agreement with the experimental results.

I. INTRODUCTION

VARIOUS discussions of four-region semiconductor structures have appeared in the literature in recent years. Shockley's description[1] of a *p-n-p-n* hook collector transistor having a current amplification in excess of unity has been further developed and extended,[2] and equivalent circuits have been obtained[3] for various modes of operation of four-region structures. These discussions have been principally in terms of germanium devices. In silicon devices, however, the current gain factors are strong functions of the emitter currents, and a somewhat different behavior might be expected. Two-terminal silicon *p-n-p-n* transistors investigated recently by Moll *et al.*[4] were found to have the properties of an efficient electronic switch.

This paper presents an investigation of the electrical properties of three-terminal silicon *p-n-p-n* transistors of the type shown schematically in Fig. 1. Like the two-terminal device, the *p-n-p-n* triode exhibits switching properties, and in many respects is similar in behavior to the conventional thyratron. It also is analogous to the point-contact transistor switch, and it is possible that similar mechanisms may account for the behavior of both point-contact and three-junction devices. The

* Original manuscript received by the IRE, November 21, 1958; revised manuscript received, March 14, 1958.

† Bell Telephone Labs., Inc., Murray Hill, N. J.

[1] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y., pp. 112–113; 1950.

[2] W. Shockley, M. Sparks, and G. K. Teal, "*P-N* junction transistors," *Phys. Rev.*, vol. 40, pp. 151–162; July, 1951.

[3] J. J. Ebers, "Four-terminal *p-n-p-n* transistors," PROC. IRE, vol. 40, pp. 1361–1364; November, 1952.

[4] J. L. Moll, M. Tanenbaum, J. M. Goldey, and N. Holonyak, "*P-N-P-N* transistor switches," PROC. IRE, vol. 44, pp. 1174–1182; September, 1956.

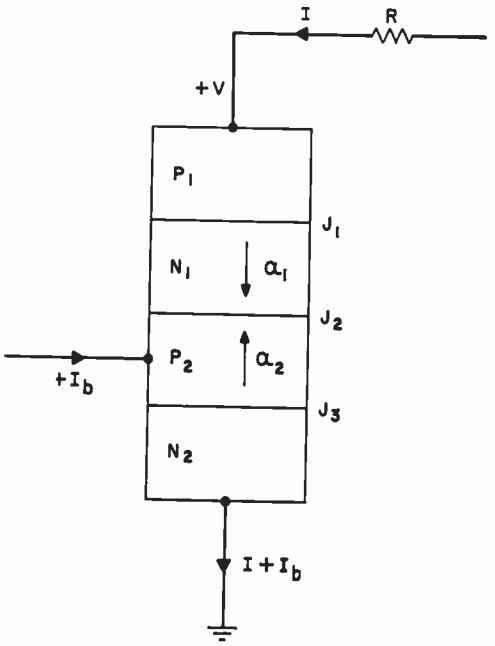


Fig. 1—Schematic of a *p-n-p-n* triode transistor.

analysis of the *p-n-p-n* triode differs from that for the hook collector transistor in that alpha is a strong function of emitter current; moreover, the hook collector transistor was considered in grounded base operation, whereas here the grounded emitter mode is more appropriate. A brief account of this work is presented elsewhere.[5] Germanium devices showing similar characteristics, but operating in a somewhat different fashion, have also been reported recently.[6,7]

If the base contact in Fig. 1 is left open-circuit the device becomes effectively two-terminal, *i.e.*, a *p-n-p-n* diode as investigated by Moll *et al.*[4] Their findings are discussed briefly here so that this mode of operation may be used as the norm to which triode behavior can be compared. The explanation is in terms of current gain factors (α_1 for the *p-n-p* section, α_2 for the *n-p-n* section) which are assumed to be functions of current but voltage independent,[8] and of current multiplication factors M_n and M_p at junction J_2 which are assumed to be functions of voltage only.[9]

A typical "forward" *V-I* characteristic for a silicon *p-n-p-n* diode is shown schematically in Fig. 2. By the forward direction, we mean that region P_1 is positive

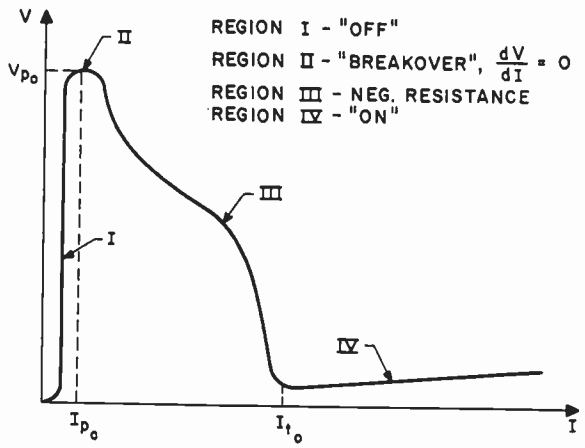


Fig. 2—A typical forward *V-I* characteristic of a *p-n-p-n* diode.

with respect to region N_2, so that junctions J_1 and J_3 act as emitters and J_2 as a collector. Let $\alpha_T = \alpha_1 M_p + \alpha_2 M_n$. The four principal regions of the characteristic can then be described as follows.

1) At low currents, $\alpha_T < 1$, and J_2 behaves essentially like a classical *p-n* junction in reverse bias. This is therefore the "off" or high impedance region.

2) As the voltage is increased the multiplication eventually increases to a point where $\alpha_T \approx 1$ and J_2 breaks down. This will be called the "breakover" region, defined by $\partial V/\partial I = 0$, at which the voltage and current in two-terminal operation have the values V_{p_0} and I_{p_0}, respectively.

3) The system then traverses a region of differential negative resistance, which usually consists of two ranges of high negative resistance separated by a range of lower negative resistance, as indicated schematically in Fig. 2.

4) The low impedance region is reached when $\alpha_1 + \alpha_2 \geq 1$ (at these low voltages M_n and M_p are essentially unity). The condition of current continuity requires that J_2 then becomes forward biased so as to emit holes and electrons back into the base layers. The low-current limit of this "on" region is designated the "turn-off" current, I_{t_0}.

II. Experimental Triode *V-I* Characteristics

At least for low-frequency applications, it is obvious from Fig. 2 that the *p-n-p-n* diode can be made to switch on only if the applied bias is increased sufficiently for the load line to pass through the breakover voltage V_{p_0}. The high multiplication effects necessary to meet the breakover condition $\alpha_T \approx 1$ could be avoided, however, if a way could be found of increasing α independently. One way of doing this is to apply the third terminal (Fig. 1) and, by means of a separate external circuit, to drive a base current I_b in the same direction as I across the emitter junction J_3. This base current can now increase α_2 independently of V and I. In other

[5] I. M. Mackintosh, "Three-terminal *p-n-p-n* transistor switches," IRE Trans. on Electron Devices, vol. 5, pp. 10–12, January, 1958.

[6] C. W. Mueller and J. Hilibrand, "The thyristor—a new high-speed switching transistor," IRE Trans. on Electron Devices, vol. 5, pp. 2–5; January, 1958.

[7] J. Philips and H. C. Chang, "Germanium power switching devices," IRE Trans. on Electron Devices, vol. 5, pp. 13–18; January, 1958.

[8] The effects of transition region widening on alpha are thus neglected.

[9] K. G. McKay and K. B. McAfee, "Electron multiplication in silicon and germanium," *Phys. Rev.*, vol. 91, pp. 1079–1084; September, 1953.

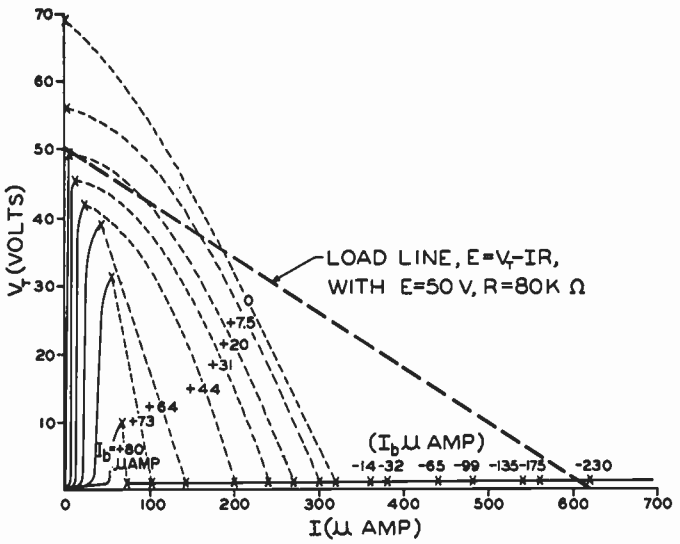


Fig. 3—The experimental V-I characteristics for a p-n-p-n triode at constant base current.

words, α_2 is a function of $(I+I_b)$, α_1 is a function of I only, and the value of α_T, which controls the shape of the V-I characteristic, can be modified by the flow of base current.[10]

The effect of base current on the V-I characteristic was observed experimentally on a diffused silicon p-n-p-n triode of the type shown in Fig. 1. I_b was maintained constant at various values and the main current I observed as the voltage applied to the device V_T was increased to breakover. After the device switched to the sustain point, V_T was gradually reduced and I observed until the turn-off point was reached. By this means the curves of Fig. 3 were obtained. Observations were not taken in the negative resistance regions and the breakover and turn-off points (indicated by the crosses) are simply joined by the dotted lines.

The most obvious effects of increasing I_b are to increase the "off" current and to decrease both the breakover voltage and the turn-off currents. These changes can be explained qualitatively in simple physical terms.[5] In the "off" state the device behaves essentially like a normal transistor, regions P_1 and N_1 acting together as a hook collector. Increasing the emitter forward bias, *i.e.*, increasing I_b, increases the collector or "off" current in the usual way. This larger current causes an increase in the alphas. The condition $\alpha_T \approx 1$ therefore is reached at lower values of M_n and M_p and the breakover voltage V_p is reduced. In the "on" state the flow of base current again increases the alphas. Thus, the current I can fall to lower values before the turn-off conditions are reached.

Fig. 3 also shows a load line representing a bias of 50 volts applied through a load of 80 KΩ. It is apparent

[10] The third contact also could be made to the N_1 base region, in which case I_b would directly influence α_1, not α_2. This case we shall arbitrarily call an "inverse" p-n-p-n triode, to distinguish from the "normal" case of Fig. 1.

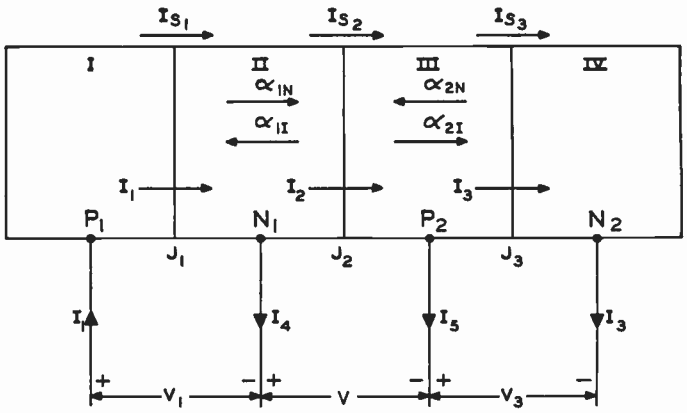


Fig. 4—Schematic of a generalized p-n-p-n structure.

that under these conditions the device can be switched on by a base current of about 20 μamp and off by a negative base current of 230 μamp.[11] In switching on, the device is quite similar to the thyratron in that a very small power is controlling a large current flow. The p-n-p-n triode has the added advantage that it can be turned off by operations on the base alone, although here the controlling currents are of the same order of magnitude as the currents switched.

So far, only the forward characteristic of the p-n-p-n structure have been discussed. The reverse characteristic (region N_2 positive with respect to region P_1) is very similar to that of the avalanche transistor.[12] The effect of the base current is simply to increase the collector current in the usual way.

III. General Analysis of Four-Region Structures

Simple physical explanations for some of the more detailed behavior of the p-n-p-n triode are not immediately obvious, and a more detailed analysis is required. The foundations have already been laid by Moll *et al.*,[4] and this section makes use of much of their work.

Fig. 4 represents schematically the structure to be considered, where it is assumed that electrical contact is made to each region. Let voltages V_1, V, and V_3 be applied across junctions J_1, J_2, and J_3, respectively, and taken as positive from left to right. Let the saturation current of each junction, when the other two junctions are short-circuited, be I_{s1}, I_{s2}, and I_{s3}. The various currents are defined as positive when in the directions shown.

Let the emission efficiencies of junctions J_1 and J_3 be γ_1 and γ_3, respectively, and let the transport factors

[11] This value of load was chosen for convenience in illustrating on Fig. 3 the turn-off action of I_b. With lower, more realistic loads the base current required to turn off the device would be much higher, of course.

[12] S. L. Miller and J. J. Ebers, "Alloyed junction avalanche transistors," *Bell Sys. Tech. J.*, vol. 34, pp. 883–902; September, 1955. M. G. Kidd, W. Hasenberg, and W. M. Webster, "Delayed collector conduction, a new effect in junction transistors," *RCA Rev.*, vol. 16, pp. 16–33; March, 1955.

	CURRENTS ARISING FROM V_1	CURRENTS ARISING FROM V	CURRENTS ARISING FROM V_3
I_1	$I_{s_1}(e^{\beta V_1}-1)$	$-\alpha_{1I}I_{s_2}(e^{-\beta V}-1)$	0
I_2	$\alpha_{1N}M_pI_{s_1}(e^{\beta V_1}-1)$	$-(M_pI_{p_s}+M_nI_{n_s})(e^{-\beta V}-1)$	$\alpha_{2N}M_nI_{s_3}(e^{\beta V_3}-1)$
I_3	0	$-\alpha_{2I}I_{s_2}(e^{\beta V}-1)$	$I_{s_3}(e^{\beta V_3}-1)$

$\beta = q/kT$; I_{p_s}, I_{n_s} = HOLE, ELECTRON COMPONENTS OF I_{s_2}

Fig. 5—Superposition equations for a generalized p-n-p-n structure.

across base regions N_1 and P_2 be β_1 and β_2, respectively.[13] Thus, when V_1 and V_3 are positive the flow of minority carriers may be described by the "normal" current gain factors

$$\left.\begin{aligned}\alpha_{1N} &= \gamma_1\beta_1\\ \alpha_{2N} &= \gamma_3\beta_2\end{aligned}\right\}. \tag{1}$$

If V is negative then J_2 will emit minority carriers in both directions. Let γ_2 represent the efficiency of this junction for the emission of holes. Then the "inverse" alphas may be written

$$\left.\begin{aligned}\alpha_{1I} &= \gamma_2\beta_1\\ \alpha_{2I} &= (1-\gamma_2)\beta_2\end{aligned}\right\}. \tag{2}$$

Note that since the β's have a maximum value of unity, $\alpha_{1I}+\alpha_{2I}\leq 1$. The definition of these alphas implicity assumes that they are independent of collector voltages.

Assume that J_1 and J_3 are always forward biased so that multiplication effects at these junctions may be ignored. Consistent with earlier work,[14] the multiplication factors at J_2 will be defined as

$$M_p = \frac{I_p(P_2)}{I_p(N_1)}, \quad \text{when} \quad I_n(P_2) = 0, \tag{3}$$

i.e., M_p is the ratio of the (majority) hole current emerging from the P_2-region side of J_2 to the (minority) hole current arriving at the N_1-region side of J_2, when the (minority) electron current arriving at the P_2-region side of J_2 is zero, and

$$M_n = \frac{I_n(N_1)}{I_n(P_2)}, \quad \text{when} \quad I_p(N_1) = 0. \tag{4}$$

The results of applying superposition to this system are seen in Fig. 5. The current components flowing due to the primary effect of one particular voltage are shown in the vertical columns.[15] The resultant currents I_1, I_2, and I_3 are obtained by adding horizontally. Note that the carriers produced by the avalanche process tend to flow as a field current (rather than a diffusion current) in the transition region. These additional carriers therefore constitute a majority carrier flow after leaving the transition region and do not influence adjacent junctions.

Solving the expressions for I_1, etc., for the factors $(e^{\beta V}-1)$ gives

$$I_{s_1}(e^{\beta V_1}-1) = \frac{[\gamma_2M_p+(1-\alpha_{2N}\alpha_{2I}-\gamma_2)M_n]I_1-\alpha_{1I}I_2+\alpha_{2N}\alpha_{1I}M_nI_3}{h} \tag{5}$$

$$I_{s_2}(e^{-\beta V}-1) = \frac{\alpha_{1N}M_pI_1-I_2+\alpha_{2N}M_nI_3}{h} \tag{6}$$

$$I_{s_3}(e^{\beta V_3}-1) = \frac{\alpha_{1N}\alpha_{2I}M_pI_1-\alpha_{2I}I_2+[(\gamma_2-\alpha_{1N}\alpha_{1I})M_p+(1-\gamma_2)M_n]I_3}{h} \tag{7}$$

where

$$I_{s_2} = I_{n_s} + I_{p_s}$$

and

$$h=(\gamma_2-\alpha_{1N}\alpha_{1I})M_p+(1-\alpha_{2N}\alpha_{2I}-\gamma_2)M_n. \tag{8}$$

These are the general expressions describing the electrical properties of the structure in Fig. 4. In principle, at least, the voltage-current relationships for any specific use of a p-n-p-n structure may be obtained from these equations.

IV. Analysis of Triode Behavior

A. Basic Considerations

A theoretical investigation of the electrical behavior of a p-n-p-n structure normally would begin by putting the applied voltage $V_T=V_1+V+V_3$, and relating V_T and I by means of (5) through (7). However, the expression so obtained proves to be intractable, and the simplifying assumption has to be made that the electrical behavior of the structure can be explained by reference to the voltage across J_2 alone. For any condition in which V_T is high, this will be a good approximation since the voltage drop across the two forward-biased junctions will be relatively small. The low-voltage point of greatest interest is "turn-off," which is described by the transition of junction J_2 from reverse to forward bias, *i.e.*, by the condition $V=0$. Thus, all the important parts of the characteristic may be considered merely in terms of (6).

Since (6) depends on the various α's and M's, it is

[13] Here the base region is defined as the space bounded by the edges of the two transition regions, *i.e.*, the β's do not include any effects which may occur in the transition regions. Also, it is assumed that the β's are effectively independent of whether the minority carriers flow to the left or right.

[14] K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, vol. 94, pp. 877–884; May, 1954.

[15] By consideration of the currents which flow in both cases, it may be verified easily that column 2 of Fig. 5 is valid for V both positive and negative. In particular, the same alphas appear whether J_2 is forward or reverse biased.

convenient at this point to consider the dependence of these parameters on current and voltage. The electron and hole multiplication factors (defined in Section III) have been found[16] to obey the empirical relationships

$$\left.\begin{aligned} M_n &= 1/(1-\eta^{n_1}) \\ M_p &= 1/(1-\eta^{n_2}) \end{aligned}\right\} \tag{9}$$

where $\eta = V/V_B$, V_B is the breakdown voltage of the isolated junction, and n_1 and n_2 depend on the type of junction considered. The silicon devices considered here are of the graded junction type for which $n_2 \approx 9$, and n_1 varies between about 1.4 and 2.5.[16]

In silicon transistors the increase in α with emitter current is found generally to be extremely rapid at low currents, becoming slower as the current increases. Typical experimental values for $d\alpha/dI$ are 10^4 to 10^5 amp^{-1} at 1 μamp and 10 to 10^2 amp^{-1} at 1 mamp. Two mechanisms are generally supposed to cause these variations. One is the effect on the emitter efficiency γ of recombination in the space charge region, as discussed by Sah, Noyce, and Shockley;[17] the other is the changes in the rate of recombination in the base region, and hence in the transport factor β, which arise in accordance with the Shockley-Read theory of saturable recombination centers.[18]

Relating the normal p-n-p-n triode of Fig. 1 to the generalized structure of Fig. 4, we have $I_1 = I_2 = I$, $I_3 = I + I_b$, $I_4 = 0$, and $I_5 = -I_b$. Eq. (6) therefore simplifies to[19]

$$I_{s_2}(e^{-\beta V} - 1) = fI_b - gI \tag{10}$$

where

$$f = \frac{\alpha_{2N} M_n}{h} \tag{11}$$

and

$$g = \frac{1 - \alpha_{1N} M_p - \alpha_{2N} M_n}{h}. \tag{12}$$

The simplification to the two-terminal case is obtained simply by putting $I_b = 0$.

Within the described limitations of V, (10) therefore should describe the behavior of the device throughout the entire V-I characteristic. Each part of the characteristic may now be considered in more detail.

B. High Impedance Region

In two-terminal operation,

$$g_0 I_0 = I_{s_2}(1 - e^{-\beta V}) \tag{13}$$

[16] S. L. Miller, to be published.
[17] C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in p-n junctions and p-n junction characteristics," Proc. IRE, vol. 45, pp. 1228–1243; September, 1957.
[18] W. Shockley and W. T. Read, Jr., "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835–842; September, 1952.
[19] For an "inverse" p-n-p-n triode, we get $I_{s_2}(e^{-\beta V} - 1) = kI_b - gI$, where $k = \alpha_{1N} M_p / h$.

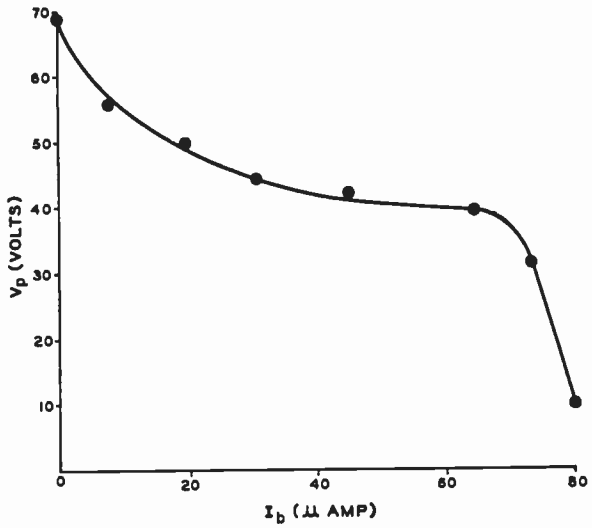


Fig. 6—The breakover voltage vs base current.

where g_0 and I_0 represent the values corresponding to $I_b = 0$. The current flowing in the triode at the same voltage V may therefore be written.

$$gI = fI_b + g_0 I_0. \tag{14}$$

Thus, the effect of I_b is to increase the "off" current by the factor $(fI_b + g_0 I_0)/gI_0$.

Consider the case where I_b is constant. The increase in I due to the base current causes increases in the alphas and reduces the value of g. The increase of I with V thus tends to be more rapid, and the "off" impedance is reduced. This effect can be seen in Fig. 3, and is discussed further in the Appendix.

C. Breakover Region

Fig. 6 shows the breakover voltage replotted as a function of base current. To explain this analytically, the condition $\partial V/\partial I = 0$ is applied to (10), as discussed in the Appendix. A relation is obtained between the voltage and currents at breakover which, to a good approximation, may be written

$$A_1 M_p + (A_2 + A_3) M_n = 1 \tag{15}$$

where

$$A_1 = \alpha_{1N} + I \frac{\partial \alpha_{1N}}{\partial I}$$

$$A_2 = \alpha_{2N} + I \frac{\partial \alpha_{2N}}{\partial I}$$

$$A_3 = I_b \frac{\partial \alpha_{2N}}{\partial I}.$$

If, when $I_b = 0$, the breakover current is extremely small (*i.e.*, comparable to I_s), then the alphas will be very low and (15) can be satisfied only by large values of M_n and M_p. In other words, V_{p_0} will lie close to the breakdown voltage V_B of junction J_2. Small increases in I_b (and hence in I) will therefore cause a relatively large in-

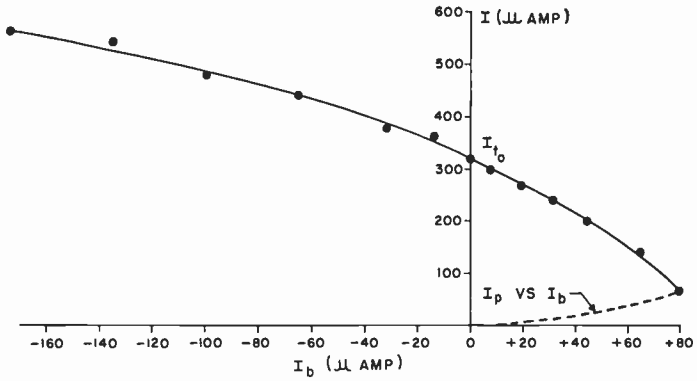


Fig. 7—The turn-off current vs base current.

crease in the alphas. Hence, there is a rapid decrease in the multiplication factors and V_p falls sharply. Further increases in I_b have less relative effect on the alphas, and the rate of change of the M's, and hence of V_p, with I_b tends to decrease. At high base currents the A factors have increased to the point where the M's must be very close to unity. In this range the multiplication factors are slowly varying functions of the voltage. Thus the required change in V_p with I_b becomes large and V_p fails abruptly towards zero. The base current at which this effect occurs will be designated I_b (max).

D. Negative Resistance Region

The Appendix also discusses the negative resistance region. The more interesting results are that the differential negative resistance theoretically should be infinite in the high current range (I very close to the turn-off current I_t), and that the concavity in this region of the V-I characteristic (see Fig. 2) occurs through the considerable decrease in the $\partial\alpha/\partial I$ terms in the current range covered by the negative resistance region.

E. Low Impedance Region

As discussed in Section IV-A, the turn-off point is defined by the condition that $V=0$. Eq. (13) then gives

$$\alpha_{1N} + \alpha_{2N} = 1 \tag{16}$$

as the condition defining the two-terminal turn-off current I_{t_0}.

In three-terminal operation, applying the condition $V=0$ to (10) gives

$$I_t = -\frac{\alpha_{2N} I_b}{\alpha_{1N} + \alpha_{2N} - 1}. \tag{17}$$

The main point of interest about this equation is that, unlike the two-terminal case, turn-off does not occur when $(\alpha_{1N}+\alpha_{2N})=1$. When I_b is positive, turn-off occurs with $(\alpha_{1N}+\alpha_{2N})<1$, *i.e.*, at lower values of I. In the "on" state the base current can be negative providing $|I_b|<I$. Under these conditions, turn-off occurs with $(\alpha_{1N}+\alpha_{2N})>1$, *i.e.*, at higher values of I.

Fig. 7 shows I_t replotted as a function of I_b. In accordance with (17), the curve has a variable negative gradient and intersects the ordinate $I_b=0$ at I_{t_0}. Fig. 7

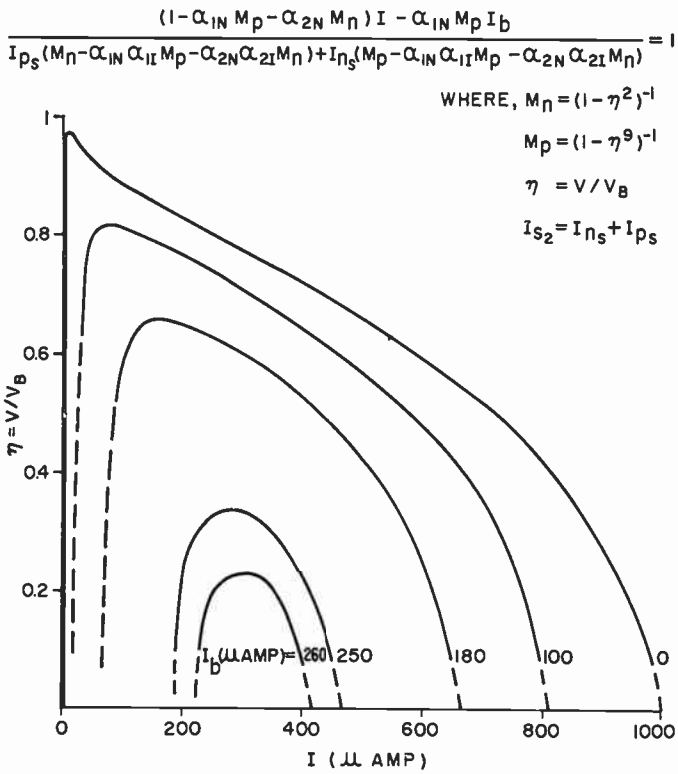


Fig. 8—The theoretical V-I characteristics for a p-n-p-n triode at constant base current.

also shows (dotted line) the observed variation of the breakover current I_p with base current. I_p increases because of the increase in the "off" current, discussed in Section IV-B. It can be seen that at I_b (max) the breakover and turn-off points coincide.

F. Theoretical V-I Characteristics

Detailed comparison between theory and experiment is not possible at present because of the practical difficulty of measuring all the parameters involved. As a further test of the theory, therefore, curves of V vs I at various fixed values of base current have been computed from the basic equations. An "inverse" p-n-p-n triode[10] was used as the theoretical model although the experiments were conducted on "normal" devices; the general effects of the base current should, of course, be the same in both types. The simplifying assumption was made that the voltage was high enough for the factor $e^{-\beta V}$ to be neglected; the calculations therefore are not valid in the "on" region. It was assumed that $\gamma_2=0.5$, that the multiplication factors had the analytical forms of (9), and that for the purposes of this computation the very approximate assumption could be made that the dependence of the alphas on current was parabolic, in accordance with the expression $\alpha=KI_e^{1/2}$, K being a constant. The theoretical curves, and the simplified expression from which they are computed, are shown in Fig. 8. It can be seen that the analysis is in good qualitative agreement with the observed behavior of the device.

Appendix

The Differential Resistance and the General Breakover Conditions

We shall only consider cases where the voltage V is high enough for the factor $e^{-\beta V}$ to be neglected (*i.e.*, V must be greater than about 1 volt). Simplifying (10) to these conditions and substituting for f and g gives

$$(1 - \alpha_{1N}M_p - \alpha_{2N}M_n)I - \alpha_{2N}M_nI_b = hI_{s_2}. \quad (18)$$

Differentiating with respect to I assuming I_b constant, and putting

$$\frac{\partial M}{\partial I} = \frac{\partial M}{\partial V}\cdot\frac{\partial V}{\partial I},$$

we have

$$1 - \alpha_{1N}M_p - \alpha_{2N}M_n - I\left[M_p\frac{\partial\alpha_{1N}}{\partial I} + M_n\frac{\partial\alpha_{2N}}{\partial I}\right] - I_bM_n\frac{\partial\alpha_{2N}}{\partial I} - \left[I\left(\alpha_{1N}\frac{\partial M_p}{\partial V} + \alpha_{2N}\frac{\partial M_n}{\partial V}\right) + I_b\alpha_{2N}\frac{\partial M_n}{\partial V}\right]\frac{\partial V}{\partial I} = \frac{\partial(hI_{s_2})}{\partial I}. \quad (19)$$

All the terms implicit in the right-hand side of (19) involve I_{s_2}, I_{r_s}, I_{p_s} or their differential coefficients with respect to I, and to a first-order approximation may be neglected. Rearranging then gives

$$\frac{\partial V}{\partial I} = \frac{\left[1 - M_p\left(\alpha_{1N} + I\frac{\partial\alpha_{1N}}{\partial I}\right) - M_n\left(\alpha_{2N} + (I + I_b)\frac{\partial\alpha_{2N}}{\partial I}\right)\right]}{\alpha_{1N}I\frac{\partial M_p}{\partial V} + \alpha_{2N}(I + I_b)\frac{\partial M_n}{\partial V}} \quad (20)$$

Eq. (20) may be used now to consider in more detail the breakover region and the differential resistance on each side.

Static breakover is defined by $\partial V/\partial I = 0$. Except in the indeterminate case when α_{1N} and α_{2N} are zero, breakover will be described by the general condition

$$M_p\left[\alpha_{1N} + I\frac{\partial\alpha_{1N}}{\partial I}\right] + M_n\left[\alpha_{2N} + (I + I_b)\frac{\partial\alpha_{2N}}{\partial I}\right] = 1. \quad (21)$$

This is the relation discussed in Section IV-C.

The differential resistance in the "off" region is given directly by (20). The terms in I_b decrease the numerator and increase the denominator, tending to reduce $\partial V/\partial I$. Although this effect will presumably be small, it can be seen in Fig. 3.

The low-current limit of the negative resistance region is described by $\partial V/\partial I = 0$, by definition. Near the high-current limit I_t, the multiplication factors are close to unity and $\partial M/\partial V$ approaches zero, *i.e.*,

$$\underset{I\to I_t}{\text{Lt.}}\frac{\partial V}{\partial I} = -\infty. \quad (22)$$

Both experiment and theory indicate that when the breakover current is very small (*i.e.*, I_b very small or zero), there is a sharp fall in voltage just beyond the breakover point (cf. Fig. 8). The explanation for this effect is similar to that given in Section IV-C for the initial rapid drop in breakover voltage as a function of base current. Thus, if I_b is very small the alphas will be very low, and (21) requires high values for M_n and M_p. Immediately after breakover, the small increase in I causes a relatively large increase in the alphas; the multiplication factors and hence the voltage must therefore fall sharply.

V. Acknowledgment

The author is grateful to I. M. Ross, F. M. Smits, and J. M. Goldey for helpful discussions, G. L. Baldwin and Mrs. W. Mammel for advice in computing theoretical curves of Fig. 8, H. H. Loar and L. A. D'Asaro for material from which experimental devices were made.

Multiterminal *P-N-P-N* Switches*

R. W. ALDRICH†, AND N. HOLONYAK, JR.†, ASSOCIATE MEMBER, IRE

Summary—A silicon *p-n-p-n* switch (two or three terminal) whose operation depends in part upon electric field transport of minority carriers is described. In a *p-n-p-n* structure relying upon an electric field to increase one internal alpha sufficiently to produce switching, the "turn-on" current is related to minority-carrier lifetime (diffusion length) and to the resistivity, area, and base width of the section depending upon field transport. It is shown that at least one base region can be of relatively large dimension. This increases "turn-on" current, "on" impedance, and "turn-off" time, but allows greater freedom in some aspects of device design and fabrication at not a great increase of the characteristics mentioned. In particular, two and three terminal *p-n-p-n* switches, dependent in part upon field transport, can be designed as signal devices or as power devices. Certain features of three-terminal *p-n-p-n* operation are discussed and experimental results presented.

INTRODUCTION

THE transistor and switching properties of semiconductor *p-n-p-n* structures have been extensively studied.[1–4] Moll *et al* described the characteristics of two-terminal silicon *p-n-p-n* switches and presented the conditions under which such devices exhibit switching behavior.[2] It was shown that increasing alphas in the *p-n-p* and the *n-p-n* sections of the device led to switching with the device biased positively at the exterior *p* region and negatively at the exterior *n* region. As the bias voltage is increased and approaches the avalanche voltage of the center junction, minority carriers injected by the exterior *p* region and the exterior *n* region are multiplied upon collection at the reverse-biased center junction. The resulting increased current causes the normal alphas of each section to increase until their sum is unity, whereupon switching to a low impedance state occurs; all three junctions become forward-biased.[2] Later Jonscher showed under what conditions a two-terminal *p-n-p-n* would switch when voltage is applied so that initially the center junction is forward-biased.[3] Moll *et al* further pointed out that a three-terminal *p-n-p-n* behaves much as a thyratron. A silicon "controlled rectifier," a high-current version of a three-terminal *p-n-p-n* switch, has been announced by the General Electric Company.[5] This device includes a relatively narrow interior *p* region and a wide interior *n* region [Fig. 1(a)]. The writers of the present paper have gathered evidence that the switching behavior of this device depends upon a hitherto undescribed mechanism.

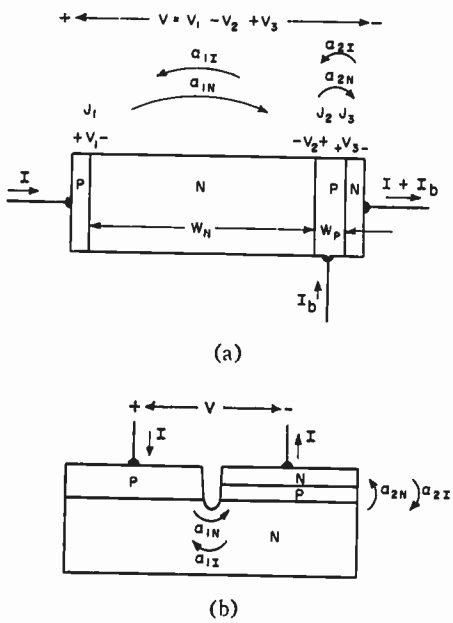


Fig. 1—Silicon *p-n-p-n* structures, two- and three-terminal.

The switching behavior of *p-n-p-n* devices results from alphas which increase with current. In silicon transistors changes in alpha with current have been widely observed and have been attributed to traps and/or generation-recombination in the junctions.[2,6] It is the purpose of this paper to demonstrate that increased minority-carrier transport in an electric field (resulting from ohmic current) plays a major role in the switching operation of certain types of *p-n-p-n* structures. Experimental evidence on two- and three-terminal silicon *p-n-p-n* switches is presented to demonstrate the role of the ohmic-current field. Some aspects of the behavior of three-terminal *p-n-p-n* switches are presented.

FIELD OPERATION OF SILICON *P-N-P-N* DEVICES

Fig. 1 shows two silicon structures, (a) and (b), which rely on field to increase α_{1N} to achieve switching. In Fig. 1(a) the dimension W_N is so large that minority carrier transport by diffusion from the *p* region at the left of the back-biased junction J_2 is negligible, or, in any case less than that due to drift. The section of the unit characterized by α_{2N} is generally constructed to be in the alpha region $\alpha_{2N} \sim 0.9$.[7] As the bias voltage V is increased

* Original manuscript received by the IRE, February 26, 1958.
† General Electric Co., Syracuse, N. Y.
[1] J. J. Ebers, "Four-terminal *p-n-p-n* transistors," PROC. IRE, vol. 40, pp. 1361–1364; November, 1952.
[2] J. L. Moll, M. Tanenbaum, J. M. Goldey, and N. Holonyak, "*P-N-P-N* transistor switches," PROC. IRE, vol. 44, pp. 1174–1182; September, 1956.
[3] A. K. Jonscher, "*P-N-P-N* switching diodes," *J. Electronics and Control*, vol. 3, pp. 573–586; December, 1957.
[4] I. M. Mackintosh, "Three terminal *p-n-p-n* transistor switches," AIEE-IRE Semiconductor Devices Res. Conf., Boulder, Colo.; July, 1957.
[5] Primarily responsible for the development were G. Hall and R. Frenzel.
[6] C. J. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in *p-n* junctions and *p-n* junction characteristics," PROC. IRE, vol. 45, pp. 1228–1243; September, 1957.
[7] This, however, is not a necessary requirement.

into the multiplication region of the junction J_2, appreciable electron current flows from junction J_2 to junction J_1. A field, $E = J\rho_N$ (current density x resistivity), then exists between J_1 and J_2 (in the direction J_1 to J_2). If the injection efficiency γ of holes at J_1 is considered unity and if diffusion is negligible, then

$$\alpha_{1N} \sim \exp - (t_d/\tau_p), \tag{1}$$

where τ_p is the lifetime of holes and t_a is the transit time of a hole across the region of width W_N. The quantity t_a is given approximately as

$$t_a \sim W_N(\mu_h \rho_N J)^{-1}, \tag{2}$$

where μ_h is hole mobility, ρ_N is the resistivity, and J is the current density in the n region of width W_N. In order that the device of Fig. 1(a) switch

$$\alpha_{1N} \sim \exp - (t_d/\tau_p) \geq 1 - \alpha_{2N}.^2 \tag{3}$$

The current density required for switching is

$$J_S \sim W_N(\mu_h \rho_N \tau_p \ln [1 - \alpha_{2N}]^{-1})^{-1}. \tag{4}$$

The dependence upon the reciprocal of the logarithm of $(1-\alpha_{2N})^{-1}$ makes J_S relatively insensitive to variations in α_{2N}.

When both diffusion and electric field are influential in minority-carrier transport in the low-alpha section (α_{1N}), the requirement for switching to occur can be given approximately as

$$(1 - \alpha_{2N}) \sim \exp (\beta W_N E/2)\{\beta L_p E/2(\beta^2 Lp^2 E^2/4 + 1)^{-1/2} \sinh (\beta^2 Lp^2 E^2/4 + 1)^{1/2} \cdot W_N/Lp + \cosh (\beta^2 Lp^2 E^2/4 + 1)^{1/2} \cdot W_N/Lp\}^{-1}, \tag{5}$$

where the right side of (5) represents α_{1N} when a constant field E is assumed in the region W_N, Lp is the diffusion length of holes, $\beta \equiv q(kT)^{-1} \sim 40$, and the other quantities have been previously defined. If the electric field E is zero, the right side of (5) becomes the conventional diffusion transport factor across the region W_N. If $LpE \ll kT(q)^{-1} = \beta^{-1}$ and $W_N > Lp$, the right side of (5) becomes

$$\sim 2(1 - \beta LpE/2) \exp (-W_N/Lp + \beta W_N E/2). \tag{6}$$

For low lifetime τ_p and small electric field E, (6) shows that α_{1N} increases essentially exponentionally with the voltage $W_N E$ across the base region. As larger fields are developed in the region of width W_N, (6) no longer applies, and the expression (5) must be used. For a structure in which α_{2N} is relatively large and α_{1N} is relatively small, the expression (5) is a reasonable approximation where again $E = \rho_N \cdot J$.

Fig. 2 is based upon (5) and shows in essence the electric field and hence current density J_S required for switching as a function of the ratio of the dimension W_N to the diffusion length Lp. α_{1N} at switching is a parameter. It will be noticed that for a given ratio

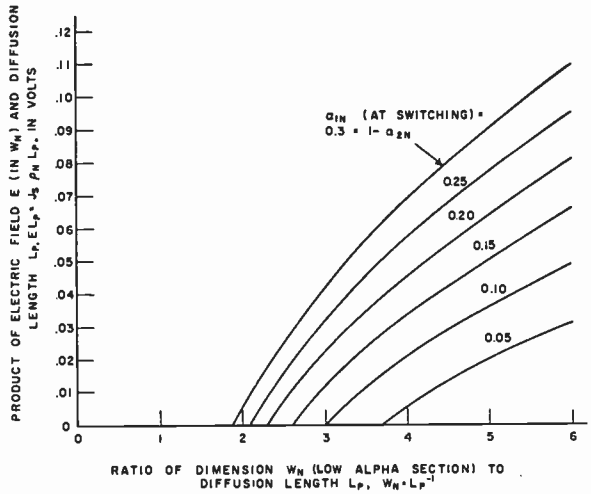


Fig. 2—Switching characteristics of low α_{1N} p-n-p-n structure.

W_N/Lp greater values $\alpha_{1N}(=1-\alpha_{2N})$ require larger current densities J_S. Likewise for a given α_{1N} larger values of W_N/Lp require larger current densities to achieve switching. For a fixed ratio W_N/Lp, lower "turn-on" current densities J_S are required for larger values of ρ_N. Qualitatively this behavior is in agreement with (4) as is readily seen when (4) is written in the form

$$J_S \rho_N Lp \sim W_N/Lp(\beta \ln [1 - \alpha_{2N}]^{-1})^{-1}$$

and compared to Fig. 2.

Fig. 3 represents the characteristics of a silicon p-n-p-n upon which it was possible to measure independently α_{2N} as a function of current through the device. Also via a fourth electrode $\sim$0.002 inch into the n region from the junction J_1 it was possible to measure the voltage V_1 plus the ohmic drop across the $\sim$0.002 inch region as a function of current (until switching occurred). From Fig. 3 it is reasonable to assume that at breakdown 0.2 v or more ohmic drop exists across the $\sim$0.002 inch region. The dimension W_N measured $\sim$0.007 inch. If it is observed further from Fig. 3 that $\alpha_{1N} \geq 0.12$ for switching to occur, from (1) τ_p must be of the order of 0.5 μsec which is found to agree with calculations of τ_p based on (5) (Fig. 2). Independent measurements of τ_p have agreed with this value. If parameters such as the resistivity, lifetime, and areas of the junctions J_1 and J_2 are held constant, then larger dimensions W_N necessitate correspondingly larger "turn-on" currents as has been discussed. Experimental devices have been built with W_N ranging from 0.005 inch (in some cases less) to 0.015 inch (hence negligible diffusion-alphas because of low lifetime) with "turn-on" currents ranging from about 1 ma to 10 ma. P-n-p-n switches of equal area and varying width W_N have been fabricated upon a single multiply diffused wafer and have displayed almost a linear dependence of J_S upon W_N, or more accurately the type of dependence shown in Fig. 2. If W_N is made sufficiently small, and diffusion transport of minority carriers from J_1 to J_2 predomi-

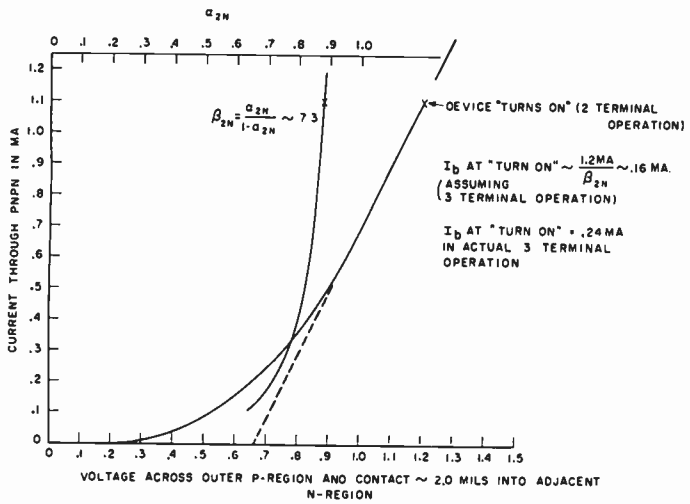


Fig. 3—Alpha and "turn-on" properties of an experimental silicon p-n-p-n switch, unit A.

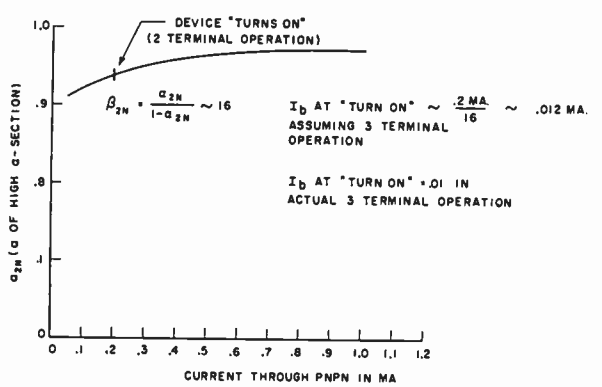


Fig. 4—Alpha and "turn-on" properties of an experimental silicon p-n-p-n switch, unit B.

nates, switching depends upon traps and/or generation-recombination. In many structures it is likely that transport due to both field and diffusion are important. A paper by I. A. Lesk[8] shows very directly the importance and necessity of an electric field to obtain switching in germanium p-n-p-n structures.

Fig. 1(b) depicts a p-n-p-n structure in which the device geometry was chosen to give negligible α_{1N}. Switching behavior was observed when sizeable currents were drawn (and fields created to assist carrier transport) through the device. As would be expected rather high "on" impedances are displayed by such devices.

The silicon structures of Fig. 1(a) have been operated in the manner of thyratrons. The bias applied to the device is kept below the avalanche region of the center junction, and switching to the "on" state is accomplished by means of a control current I_b introduced into the base of the high-alpha region, characterized by α_{2N}. Prior to switching the current I through the reverse biased junction J_2 is given as [see Appendix and Fig. 1(a)]

$$I=\{I_{S2}(1-\alpha_{1N}\alpha_{1I}-\alpha_{2N}\alpha_{2I})+I_b\alpha_{2N}\}(1-\alpha_{1N}-\alpha_{2N})^{-1}.$$

This expression may be rewritten as

$$I = \{I_{S2}(1 - \alpha_{1N}\alpha_{1I} - \alpha_{2N}\alpha_{2I}) + I_b\alpha_{2N}\}(1 - \alpha_{2N})^{-1} \cdot\left(1 - \frac{\alpha_{1N}}{1 - \alpha_{2N}}\right)^{-1}. \quad (7)$$

From (7) it is readily seen that if the inverse alphas of the structure are small and if $\alpha_{1N}<(1-\alpha_{2N})$,

$$I \sim (I_{S2} + I_b\alpha_{2N})(1 - \alpha_{2N})^{-1}. \quad (8)$$

At low currents, where $\alpha_{1N}<(1-\alpha_{2N})$, (8) indicates that the device operates as a grounded emitter transistor with gain $\alpha_{2N}(1-\alpha_{2N})^{-1}$. The exact expression (7) reveals that the structure, or an equivalent embodiment, will act as a high gain transistor provided $0.9<(\alpha_{1N}+\alpha_{2N})<1$. Furthermore, the gain will be essentially constant as long as $(\alpha_{1N}+\alpha_{2N})$ is constant.

I increases with I_b, and as previously discussed α_{1N} (and α_{2N}) also increases. This continues until $\alpha_{1N}\geq(1-\alpha_{2N})$ and switching occurs. Just as the voltage V_2 across the junction J_2 becomes positive (changes sign, see Appendix)

$$I_b \geq I(1 - \alpha_{1N} - \alpha_{2N})\cdot\alpha_{2N}^{-1}. \quad (9)$$

If the current I through the device at which this occurs is very roughly approximated by (8),

$$I \geq I_{S2}\cdot\alpha_{1N}^{-1}. \quad (10)$$

More meaningful calculations of "turn-on" currents, in the case of field operation, can be obtained from considerations of lifetime, resistivity, and geometry as discussed previously.

Figs. 3 and 4 show measurements of α_{2N} as a function of current I through the reverse-biased junction J_2 for the case of two experimental units A and B.[9] If the current I at which switching occurs in three-terminal operation is approximately equal to that at which switching occurs in two-terminal operation, the latter value divided by $\alpha_{2N}(1-\alpha_{2N})^{-1}$ approximates the control current I_b required for three-terminal operation. Physically this is reasonable. It would be expected that the same current through the device in three-terminal operation as in two-terminal operation would lead to switching. Also, in structures wherein α_{1N} is quite small until switching occurs, $I_b\sim I(1-\alpha_{2N})\cdot\alpha_{2N}^{-1}$ and cannot change much at the onset of switching. Figs. 3 and 4 indicate the validity of such an approximation for two units having high α_{2N} structures. The data of Fig. 4 were taken upon a device having drops of 0.54, 0.60, an 0.63 v when conducting currents of respectively 1, 5, and 10 ma in the "on" state.

In the four electrode p-n-p-n structures used in this work it was very easy to simulate a leaky junction J_2 with a resistance placed across the junction. As ex-

[8] To be published.

[9] These measurements were made on devices having electrodes upon all four regions.

pected, switching occurred at a much lower voltage than formerly. A somewhat higher device current was required as it was not possible to position electrodes directly on either side of J_2. The "off" impedance could be selected arbitrarily with no change in the breakdown voltage by bridging a resistance from n region to n region or from p region to p region.

Conclusions

Multiterminal silicon p-n-p-n switches have been described that depend very strongly upon the existence of electric fields to aid minority-carrier transport (increase alphas) and produce switching. It has been shown that at least one base region in a p-n-p-n structure can be of relatively large dimension. This increases "turn-on" currents, "on" impedances (but not excessively), and "turn-off" times in p-n-p-n switches but allows greater freedom in some aspects of device design and fabrication. In particular, thyratron-like three-terminal p-n-p-n switches of large dimensions can be made to operate at power frequencies and handle large currents, amperes or more. Also, smaller signal-type devices can be readily designed and built to depend upon field operation.

The fact that a field due to ohmic current may play a large role in the operation of semiconductor switches makes possible p-n-p-m switching structures, where the "m" refers to a metal contact or other type of contact which exhibits minority carrier injection at high current levels. Structures possessing one high alpha section are particularly susceptible to switching provided even slight injection occurs at what may be characterized as an "m" contact. This applies to deliberate p-n-p-m structures as well as to some categories of transistors which have somewhat emitting collector contacts.

Appendix

A brief analysis is presented below extending the equations of Moll *et al*[2] from the case of two-terminal p-n-p-n operation to that of three terminals. The notation used here follows that of the reference cited with the exception that the voltage across the junction J_2 (Fig. 1(a)) is considered positive when the polarity is plus on the p side and minus on the n side. The equations for the currents through the junctions J_1, and J_2, and J_3 are respectively

$$I = I_{S1}(\exp \beta V_1 - 1) - \alpha_{1I} I_{S2}(\exp \beta V_2 - 1) \quad (11)$$

$$I = \alpha_{1N} I_{S1}(\exp \beta V_1 - 1) - I_{S2}(\exp \beta V_2 - 1) + \alpha_{2N} I_{S3}(\exp \beta V_3 - 1) \quad (12)$$

$$I + I_b = -\alpha_{2I} I_{S2}(\exp \beta V_2 - 1) + I_{S3}(\exp \beta V_3 - 1). \quad (13)$$

They may be solved for the voltages and give for V_2

$$(\exp \beta V_2 - 1) = \{I_b \alpha_{2N} - I(1 - \alpha_{1N} - \alpha_{2N})\} I_{S2}^{-1} \cdot (1 - \alpha_{1N}\alpha_{1I} - \alpha_{2N}\alpha_{2I})^{-1}. \quad (14)$$

If $-|V_B| < V_2 < -kT/q$, the left side of (14) is equal to -1 and

$$I = \{I_{S2}(1 - \alpha_{1N}\alpha_{1I} - \alpha_{2N}\alpha_{2I}) + I_b \alpha_{2N}\}(1 - \alpha_{1N} - \alpha_{2N})^{-1}. \quad (15)$$

If the inverse alphas and α_{1N} may be ignored (a reasonable approximation at low currents),

$$I \sim (I_{S2} + I_b \alpha_{2N})(1 - \alpha_{2N})^{-1}. \quad (16)$$

From (16) it is obvious that at low currents, operation of the device is similar to grounded emitter operation of a conventional transistor.

When the device switches, all junctions are in forward bias. V_2 switches from a negative to a positive quantity, and the left side of (14) must be greater than zero. Eq. (14) gives

$$\alpha_{1N} \geq 1 - \alpha_{2N}(1 + I_b \cdot I^{-1}) \quad (17)$$

just as V_2 goes through zero and becomes positive.

Acknowledgment

The authors would like to express their gratitude to their many associates who have assisted in this work. Particular thanks are due to I. A. Lesk, H. A. Jensen, and G. Hall for supplying much useful information and for many stimulating discussions. Also, thanks are due to the General Electric Co., Metals Engineering Group, for the supply of crystals, and to B. G. Hess for the preparation of silicon wafers.

The Application of Transistors to Computers*

R. A. HENLE†, ASSOCIATE MEMBER, IRE, AND J. L. WALSH†, ASSOCIATE MEMBER, IRE

Summary—In recent years, the transistor has replaced the vacuum tube for most switching applications. This change has been brought about by the near-ideal switching characteristics of transistors, their low power requirements and long life. The purpose of this paper is to acquaint the reader with the basic principles of several circuit philosophies presently in use. Some advantages and disadvantages of these philosophies are given.

I. Introduction

THIS PAPER is a survey of some of the transistor switching circuit techniques now in use, and is concerned mainly with direct-coupled types of circuitry. We have attempted to present some of the design considerations and the advantages and disadvantages of various circuit philosophies in this class. This paper is not all-inclusive, and some circuit philosophies are not covered. What is covered is felt to be relevant at the present time. There are as many circuit building blocks as there are circuit designers. However, many of the variations are in details rather than in the broad aspects.

Sections II and III give a brief survey of some of the more important dc and ac transistor switching characteristics. Sections IV to VIII are concerned with resistance-coupled approaches. This covers the broadest class and embraces many possible combinations of circuits. Some designers may prefer transistor logic, others the diode logic approach. Some may make use of complementary transistor types, others may prefer circuits requiring only one type. We do not wish to minimize the differences, since seemingly slight variations can introduce significant economic and performance factors for a particular application. It is beyond the scope of this paper to consider all the effects of the various mutations in the resistance-coupled class. Rather, we have tried to discuss the general advantages and disadvantages of representative circuits. Section IX discusses briefly the direct-coupled transistor logic circuits, and Section X is concerned with the current switching approach.

On many of the circuits, the resistor, capacitor, and power supply values are included to give the reader a feeling for the magnitudes involved. Component values, of course, vary with the circuit application and the transistor specifications. We have not attempted to make these circuits a compatible set of building blocks.

II. Transistor DC Characteristics

This section consists of a very brief review of the dc switching characteristics of transistors, with the terminology that is used in later sections. The characteristics discussed are involved in the design of nearly all switching circuits. The importance of any one parameter varies with the circuit design philosophy.

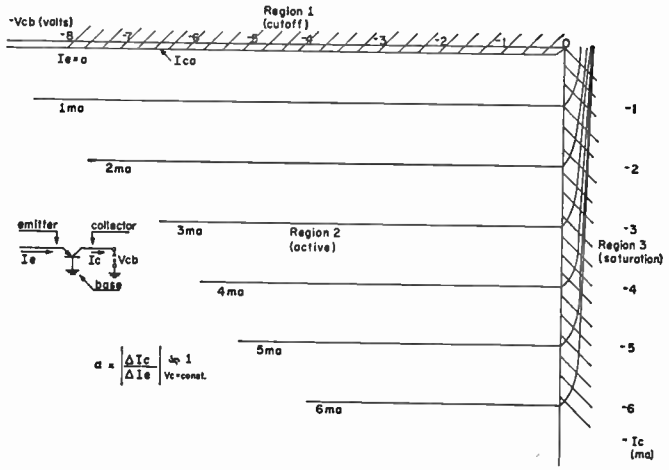


Fig. 1—Grounded-base collector V_c-I_c characteristic for constant values of I_e.

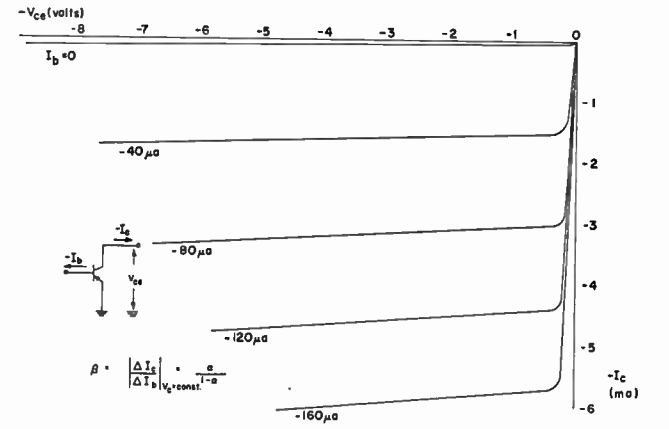


Fig. 2—Grounded-emitter collector V_c-I_c characteristic for contant values of I_b.

Fig. 1 is a plot of the grounded-base collector characteristic for constant values of emitter current. This plot has been divided into three operating regions: cutoff, active, and saturation. The current gain (α) for this connection is very close to unity; it is generally in the range of 0.95 to 0.99. This set of curves by itself is only informative in the saturation region. It generally differs appreciably between transistor types only in this region.

Fig. 2 shows the grounded-emitter collector characteristic for constant values of base current. The appearance of this plot will vary considerably from unit to unit because of the nonuniformity of β. β is readily determined from these curves, as is its uniformity over the active region. The saturation voltage drop and the decrease in this voltage with increasing base current are also evident.

* Original manuscript received by the IRE, April 21, 1958.
† Product Dev. Lab., IBM Corp., Poughkeepsie, N. Y.

Generally, transistors vary so widely from unit to unit that circuits are not designed from characteristic curves like those in Figs. 1 and 2. Designs are based on device specifications which detail the minimum and maximum limits of the individual transistor parameters. The more important of these parameters are discussed in the following parts.

I_{co}

This is the collector current which flows with zero emitter current. It is shown exaggerated in Fig. 1. For most low-power junction transistors, it will have a value under 5 μa at 30°C. I_{co} is relatively independent of collector voltage, but strongly temperature dependent. With some variation from transistor to transistor, I_{co} can be expected to double for every 8°C rise in junction temperature.

K Factor

This parameter represents the thermal resistance of the transistor.

$$K \text{ factor} = \frac{\text{degrees } C \text{ rise in junction temperature}}{\text{milliwatts of power dissipation}}.$$

The junction temperature for the determination of I_{co} is taken to be the ambient temperature plus the product of K factor and power dissipation. K factor is not a constant for a particular device, but depends on the mounting and environment of the transistor.

Saturation Voltage Drop

This voltage is measured from collector to emitter with the transistor in saturation. It is generally determined by applying the minimum operating base current and maximum operating collector current to the transistor and, under these conditions, measuring the collector-to-emitter voltage. The variation in this voltage represents the variation in the ON level of a saturated transistor. Measurement in this manner also insures a minimum saturation current gain. That is, if the transistor meets the saturation voltage specifications, it also has a minimum current gain at this operating point.

Emitter-Base Characteristic

Fig. 3 shows a plot of V_{be} (voltage from base to emitter) vs collector current for two alloy transistors at temperatures of 27°C and 70°C. This plot indicates that the voltage from base to emitter required to maintain a given collector current decreases with temperature. It also illustrates that to prevent base current amplification a reverse bias is required on the emitter-base diode. The collector current of these transistors increases by approximately a factor of three when going from several tenths of a volt reverse bias to zero reverse bias on the emitter-base diode. Plots of this type are generally considered when determining the necessary OFF conditions for a particular transistor type.

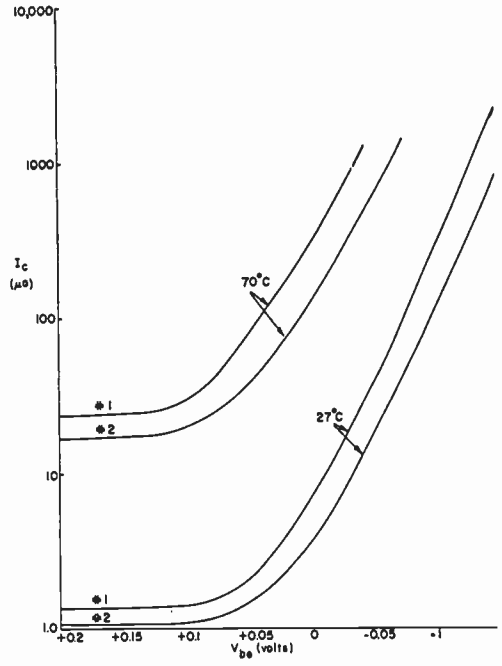


Fig. 3—Collector current vs emitter-base bias for two *p-n-p* alloy transistors.

Breakdown

Maximum collector and emitter voltages are determined by two factors, avalanche breakdown and punch-through. Either one or the other generally limits the maximum usable collector and emitter voltages at which the transistor may be operated. Although circuits have been designed which make use of these breakdowns, for most switching circuits the designer is not as interested in what type of breakdown is occurring as he is in what limitations these effects place on his allowable operating voltages.

III. AC Parameters of the Transistor

One of the first equivalent circuits[1] ever used to represent the transistor is shown in Fig. 4, with a load resistance R_L. If it is assumed that the collector generator αi_e can be represented by a minimum phase-shift network, a straightforward analysis shows that the response of the collector current i_c depends on two principal time constants. These are ω, the angular frequency cutoff of the collector generator, and $(r_b+R_L)C_c$, the time constant formed by the load and base resistances and the collector capacitance. By forming the ratio ω/r_bC_c, one could obtain a figure of merit which would give an indication of the relative speed of transistors. However, this figure of merit would be very dependent on the dc operating conditions of the transistor, because ω, r_b, and C_c can be marked functions of the dc operating point.

Contours of C_c and ω for a drift transistor are seen in

[1] R. F. Shea, ed., "Principles of Transistor Circuits," John Wiley and Sons, Inc., New York, N. Y., ch. 9; 1953.

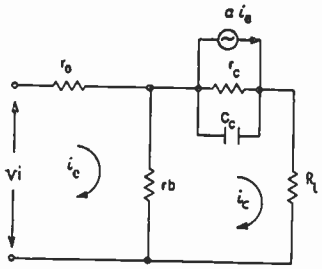


Fig. 4—Transistor equivalent circuit.

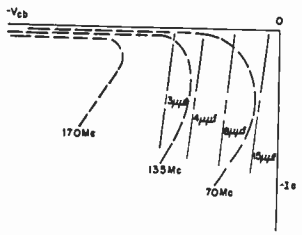


Fig. 5—Curves of frequency cutoff and collector capacitance.

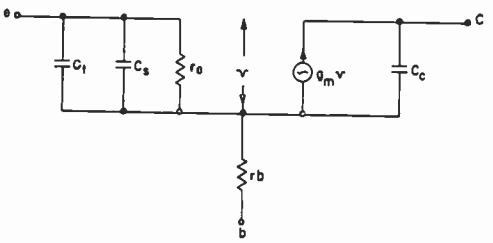


Fig. 6—Transistor equivalent circuit.

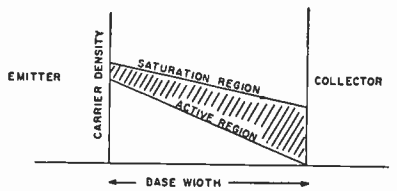


Fig. 7—Carrier density in the base region.

Fig. 5, which shows that a plot of constant frequency cutoff over the collector characteristic resembles a rectangular hyperbola. In general, at moderate values of collector current, frequency cutoff increases as collector reverse bias is increased. However, for a fixed value of collector voltage, the frequency cutoff is low at low values of collector current, increases as collector current is increased, but starts to fall off again after an optimum value of collector current is exceeded. Collector capacitance contours, also shown in Fig. 5, indicate that this parameter is quite dependent on the collector voltage but relatively independent of collector current. In general, the collector capacitance of most transistors varies inversely as the $\frac{1}{3}$ to $\frac{1}{2}$ power of the collector voltage.

An insight into why the frequency cutoff is low at low currents may be gained from the transistor equivalent circuit[2] of Fig. 6. Here the frequency response of the collector generator $g_m v$ depends on the two input capacitors, C_T and C_S. C_T is the transition capacitance and represents the barrier capacitance formed by the emitter and base material of the transistor. The transition capacitance depends on the emitter-base bias, and follows an inverse power law similar to the one for the collector capacitance. The storage capacitance C_S represents the number of carriers in the base region. As such, it is directly proportional to the emitter current, so that at very low currents the transition capacitance is larger than the storage capacitance. Note from Fig. 6 that the transition capacitance and the storage capacitance are in parallel with the emitter resistance. For low values of emitter current, the emitter resistance is high ($r_e = kT/qI_e$). A portion of the input current goes into charging the transition capacitance, and is not amplified nor does it contribute significantly to the output. This initial charging of the transition capacitance causes the frequency response to be poor at low currents.

When a transistor is in saturation and a pulse is applied at the input to turn the transistor OFF, a storage delay occurs before the output current begins to change. This problem of minority carrier storage has been investigated and a report made.[3] It may be understood by referring to Fig. 7. This shows a plot of the density of carriers on a line across the base region for the active and the saturation regions of operation. When a turn-off pulse is applied at the input, the carriers represented by the shaded region of Fig. 7 must be cleared from the base before the output current will begin to decrease. Ebers and Moll[4] have shown that the time constant for minority carrier delay is

$$\left[\frac{1}{\omega_N} + \frac{1}{\omega_I}\right]\frac{1}{1 - \alpha_n\alpha_I}$$

where

ω_N = normal angular frequency cutoff,
α_n = normal emitter-to-collect current gain of the transistor,
ω_I = inverted angular frequency cutoff, and
α_I = inverted current gain.

This time constant is of interest because it involves inverted as well as normal transistor parameters.

This problem of minority carrier delay is serious where high-speed operation is desired.

IV. Saturating Resistance-Coupled Circuits

The circuit in Fig. 8 consists of a transistor with a resistive load R_L and a resistive input network R_1 and R_2. The input network and load are chosen so that one

[2] A. L. Kestenbaum and N. H. Ditrick, "Design, construction, and high-frequency performance of drift transistors," *RCA Rev.*, vol. 18, pp. 12–23; March, 1957.

[3] J. L. Moll, "Large signal transient response of junction transistors," Proc. IRE, vol. 42, pp. 1773–1784; December, 1954.

[4] J. J. Ebers and J. L. Moll, "Large signal behavior of junction transistors," Proc. IRE, vol. 42, pp. 1761–1772; December, 1954.

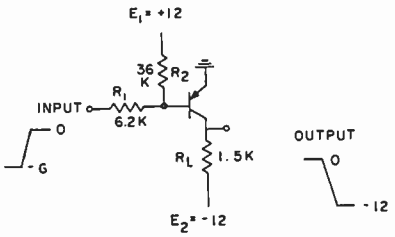


Fig. 8—Resistance-coupled inverter.

circuit is capable of driving three or more similar circuits. Despite the apparent simplicity of this circuit, many factors must be considered in its design. As an illustration, the following paragraph considers the various factors involved.

The circuit as used in digital computers has two output states. When the transistor is turned ON, it is operated in saturation and the output is at ground potential minus the saturation voltage drop of the transistor (see Fig. 2). In order to turn the transistor ON, the maximum value for its collector current must be known. This is determined by the most negative value the power supply E_2 can have and by the minimum value for the load R_L. The difference between the most negative value of E_2 and the minimum saturation voltage drop divided by the minimum value for R_L gives the maximum value of the collector current. The minimum saturation current gain of the transistor divided into this maximum collector current determines the minimum amount of base current that must be supplied to the transistor. The divider R_1 and R_2 at the input must be capable of supplying this current with the input voltage at its minimum negative value. Also considered are the ON emitter-to-base drop of the transistor, the maximum value for E_1, the minimum value for R_2, and the maximum value for R_1. This combination of these parameters gives the minimum value of the input base current, which in turn must be at least as great as that necessary to drive the transistor into saturation. Simultaneously, the input divider also must be designed to supply the required OFF current when the input is at its most negative UP value. The OFF current required at the base is the I_{co} the transistor requires at its maximum junction temperature. The maximum junction temperature is determined in turn by the maximum ambient in which the circuit is expected to operate, plus the rise in junction temperature above this ambient caused by the power dissipated in the transistor. This rise is the product of the K factor for the transistor and its maximum power dissipation in milliwatts. The maximum OFF current is a function of the reverse bias, and curves similar to those shown in Fig. 3 must be used in the design. This current also must be available when E_1 has its minimum value, R_2 is at its maximum value, and R_1 is at its minimum value. It is evident that the required characteristics for the design of this circuit include most of the dc switching characteristics for the transistor.

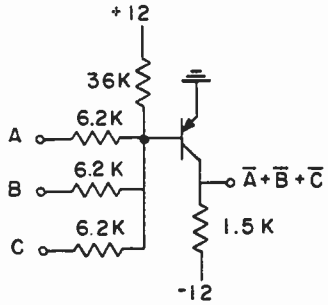


Fig. 9—NOR circuit.

The circuit of Fig. 9 is similar to that of Fig. 8, except that a number of resistor inputs have been provided. If the ground level is defined as a binary "1" and the down level as a binary "0," the circuit performs the function $\overline{A}+\overline{B}+\overline{C}$ and is called a NOR circuit. Additional factors must be considered in the design of this circuit.[5] The NOR circuit of Fig. 9 is logically complete, that is, any logical expression can be achieved by a combination of NOR's.

The NOR circuit has many advantages, including low cost, simplicity, and high reliability. Reliability is felt to be excellent because of the minimization of semiconductor components that are inherently more complex and prone to failure than simple resistors. The serious disadvantages of this circuit are the small amount of overdrive available to the transistor and the high degree of minority carrier storage brought about by the deep saturation of the transistor when two or more inputs are negative. Both of these factors tend to lower the operating speed. The NOR circuit appears quite competitive with diode logic from a cost standpoint. With low-frequency transistors ($f_{co} < 5 \times 10^6$ cps), circuit delays less than 10 μsec per stage are achieved. This would appear to limit the circuit to low-frequency applications. However, the use of higher frequency transistors with low storage time makes this type of circuitry quite competitive with more complex circuit types.

These resistance-coupled circuits have the advantage of permitting inexpensive resistors to be used for logic but, as has been discussed, have the disadvantage of low-frequency operation caused by lack of substantial overdrive at the input to the transistor. The circuit shown in Fig. 10 is used where higher operating speeds are required. The capacitor at the input to the circuit provides overdrive for turn ON and turn OFF operation. Ebers and Moll[3,4] of the Bell Telephone Laboratories have discussed the effect of current overdrive on the rise, storage, and fall times of the transistor. Their analysis relates the effects of the normal and inverted current gain and angular frequency cutoff to the response times. The overdrive provided by the input capacitor can increase the upper operating frequency by a factor of three or more.

[5] For a discussion of these the reader is referred to W. D. Rowe, "Transistor NOR circuit design," *Electronic Design*, vol. 6, pp. 2–6 29; February 5, 1958.

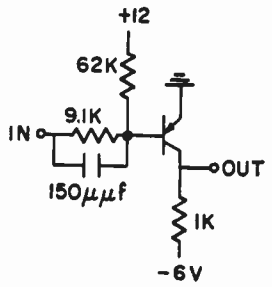


Fig. 10—Resistance capacitance-coupled inverter.

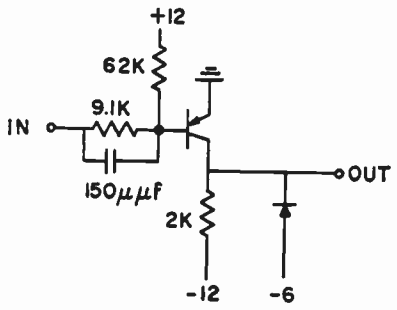


Fig. 11—Resistance capacitance-coupled inverter with diode-clamped load.

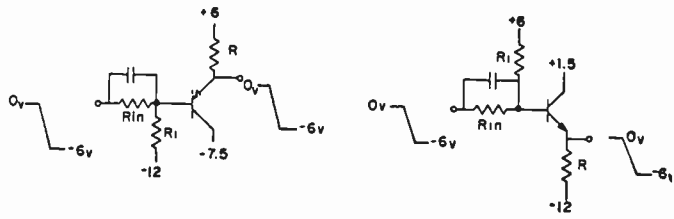


Fig. 12—Emitter follower circuits.

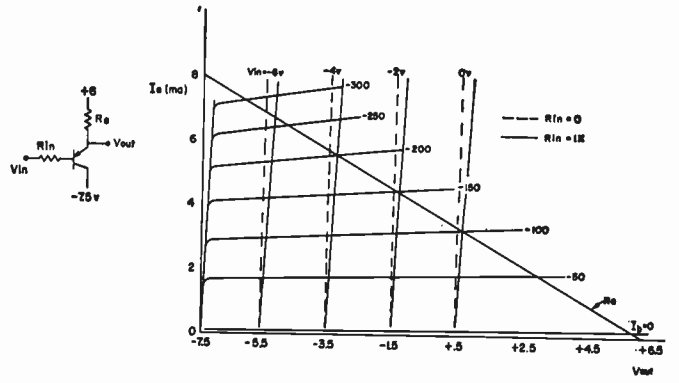


Fig. 13—Characteristics of a *p-n-p* emitter follower.

Fig. 11 shows the circuit of Fig. 10 with a diode-clamped load. This clamp provides the circuit with a well-defined OFF output level. It also reduces the fall time by having the load discharge toward −12 volts and clamping it at −6 volts.

The capacitor-coupled circuits have found wide usage in computer applications. They are used in the Lincoln Laboratory TX-2 and the IBM 608. The trigger and some of the logical connections of this circuit are discussed later in this paper. One of the main limitations of this type of circuit is that the transistor can be driven deep into saturation. As a result, minority carrier storage time is a definite consideration in determining operating times. A second disadvantage is the increased noise sensitivity of the circuits. This is brought about by the high transient currents the circuits draw and by the direct coupling of noise "spikes" into the base of the transistor through the capacitor. The circuits generate more noise and at the same time are more susceptible to it. In spite of these disadvantages, the increased performance obtained by this type of operation insures its future use in many applications.

V. Emitter Followers

Emitter followers find frequent use in computer circuits as logical switches and also as power drivers. A *p-n-p* and an *n-p-n* emitter follower are shown in Fig. 12. Although the circuit has a voltage gain of less than unity, it still is a valuable circuit because of its very low output impedance. Large-signal output impedances of 10 to 50 ohms are obtained with commercially available transistors.

The dc characteristics of a typical *p-n-p* circuit are plotted in Fig. 13. One can see from the slope of the curves of constant input voltage that the circuit has a low output resistance. This output resistance depends on the driving point resistance at the input, as is shown by the dotted curves of Fig. 13, which were plotted with the input resistor (R_{in}) short-circuited. The output potential of the *p-n-p* circuit is shifted up by the voltage drop across the emitter base diode. Since the emitter follower circuits discussed here are always conducting, this drop is relatively constant.

Thus, it is common practice to shift the input level with a voltage divider (such as R_{in}, R_1 in Fig. 12). This compensates for the emitter-base diode drop, and the output potential approximately equals the input potential.

The emitter follower frequently is used to drive capacitive loads or to drive a load of many inverters. A typical situation for a capacitive load is shown in Fig. 14. Here the dotted line represents the path of the operating point when a pulse is applied at the input. Note that in this case when the input goes positive, the emitter follower is driven to cutoff because the rise time at the input is less than the rise time at the load (2.3 RC). The cutoff condition lasts until the load capacitance C_L is charged to the input potential V_1 through the load resistor R.

When an emitter follower is used to drive a load of several inverters, the switching speed of the inverter loads depends on the value of the emitter follower load resistance. This situation, which is similar to the problem of driving capacitive loads shown in Fig. 14, is sketched in Fig. 15. When the input to the emitter follower goes negative, the inverter turns on quite rapidly. This is because the emitter follower presents a low output impedance, and a sizable transient overdrive current is drawn from the base of the inverter. A limit is reached for the turn-on condition when the steady-state currents I_{on} plus I_R cause the emitter fol-

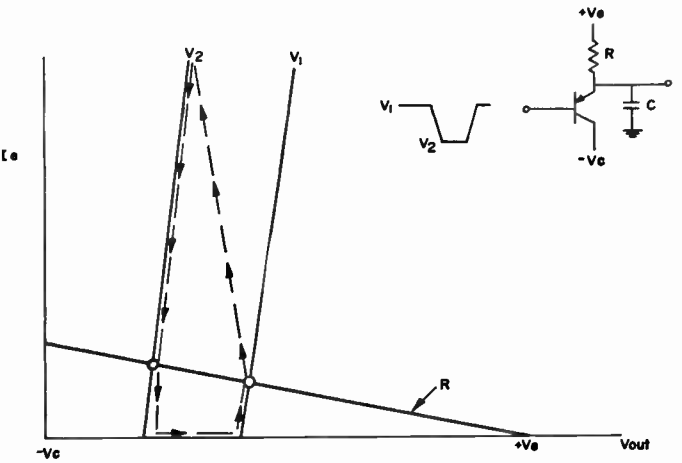


Fig. 14—Emitter follower driving capacitance.

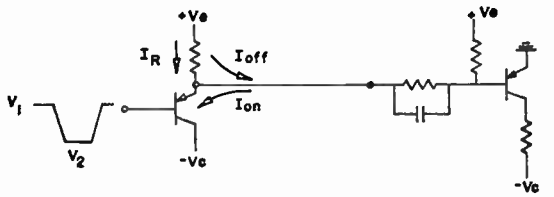


Fig. 15—Emitter follower driving an inverter.

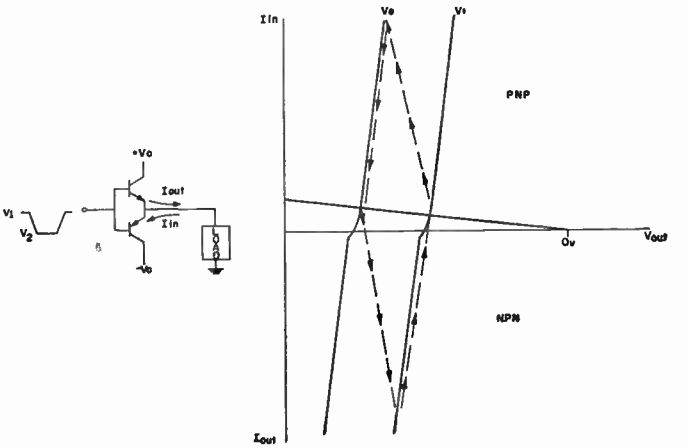
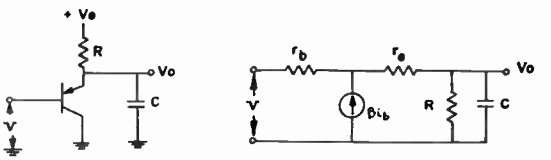


Fig. 16—Complementary emitter follower.



$$\beta = \frac{\alpha}{1-\alpha}$$

$$\alpha = \frac{\alpha_0}{1 + j\frac{\omega}{\omega_c}}$$

Fig. 17—Simplified equivalent circuit.

lower power dissipation to rise to a maximum allowable value. When the input to the emitter follower goes positive, the emitter follower cuts off and all of the transient overdrive current I_{off} must be supplied through the load resistor R. If R is made large, the value of I_R is small so that I_{on} can be made larger and more inverter loads can be driven. However, the overdrive current supplied in the off direction is also small, and the inverter loads turn off slowly. Clearly, then, if single emitter followers are used, a compromise must be made between the number of loads that can be driven and the turn-off times that are required.

The inability of a single emitter follower to drive in both directions can be overcome by using the complementary emitter follower, which is shown in Fig. 16. When the input goes negative in this circuit, the *p-n-p* emitter follower conducts, and current is drawn from the load. The *n-p-n* emitter follower is cut off. When the input goes positive, the *p-n-p* emitter follower is cut off, the *n-p-n* emitter follower conducts, and current is supplied to the load. In Fig. 16, the dotted line represents the path followed by the operating point. The complementary emitter follower is a very good power driver. It is commonly used to drive large loads, such as a shift line in a register or the pulse input line in a ring counter.

When emitter followers are used in logic or in power-driving applications, trouble is frequently encountered with oscillations or with large overshoots on the output waveform. The trouble is caused by a capacitive load on the output. Under the right conditions, the input impedance may have a negative real part. The emitter follower and a simplified equivalent circuit are shown in Fig. 17. The input impedance of this circuit will be of the form $R - JX_c$, and the real part may be written

$$R' = r_b + \frac{\left[1 - \alpha_0 + \left(\frac{\omega}{\omega_c}\right)^2\right] R - \frac{\alpha_0 \omega^2 R^2 C}{\omega_c}}{\left[(1-\alpha_0)^2 + \left(\frac{\omega}{\omega_c}\right)^2\right][1 + \omega^2 (RC)^2]} + \frac{\left[1 - \alpha_0 + \left(\frac{\omega}{\omega_c}\right)^2\right] r_e}{(1-\alpha_0)^2 + \left(\frac{\omega}{\omega_c}\right)^2}$$

where

ω = angular frequency of the input signal,
ω_c = angular frequency of the transistor, and
α_0 = low-frequency current gain.

The numerator of the second term has a negative term and can be rewritten as

$$R\left[1 - \alpha_0 + \left(\frac{\omega}{\omega_c}\right)^2 (1 - \alpha_0 \omega_c RC)\right].$$

In the expression for the input impedance, damping is provided by the base resistance r_b and the emitter resistance r_e. However, both of these can be very small, particularly in the higher speed drift transistors. If this is the case, and if the term $\alpha_0 \omega_c RC$ exceeds one by a sufficient amount, the circuit has a negative input component over the higher frequency range. The input impedance is plotted in Fig. 18.

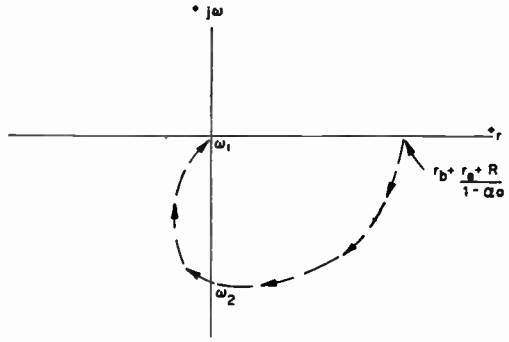


Fig. 18—Input impedance vs frequency.

If this emitter follower circuit is connected to a line which has an inductive output impedance, then the circuit will oscillate at a frequency determined by the line inductance and the input capacitance of the circuit, if this frequency is in the range of ω_1 to ω_2. The circuit can be stabilized by putting a resistor larger than the negative resistance in series with either the base or the emitter. The divider networks at the input to the emitter follower circuits shown in Fig. 12 serve also to damp out the negative resistance component of the circuits. Fortunately, the line inductance encountered in practical circuits is generally quite small. This makes the frequency of oscillation high and the resistor required for damping small. Under these conditions, the frequency of oscillation is close to ω_1 in Fig. 18. The emitter follower oscillation problem is serious, and makes it difficult to design and package emitter follower logic.

Emitter followers also are capable of generating overshoots and damped oscillations independent of the line conditions at the input. This problem has been reported in the literature.[6] Consider the equivalent circuit of Fig. 17, in which the emitter resistance is neglected and the collector capacitance is made a part of the load capacitance. A solution for the output voltage $V_0(t)$ for an input step function has the following roots:

$$S_{1,2} = \frac{-\left[\frac{1}{R_x C} + \omega_c(1-\alpha_0)\right] + \sqrt{\left[\frac{1}{R_x C} + \omega_c(1-\alpha_0)\right]^2 - \frac{4\omega_c}{R_x C}}}{2}$$

where

$$\frac{1}{R_x} = \frac{1}{r_b} + \frac{1}{R}.$$

The roots, $S_1 S_2$, indicate that there are three conditions of interest.

Overdamped case:

$$\left[\frac{1}{R_x C} + \omega_c(1-\alpha_0)\right]^2 > \frac{4\omega_c}{R_x C}.$$

[6] L. P. Hunter, "Handbook of Semiconductor Electronics," McGraw-Hill Book Co., New York, N. Y., ch. 15, pp. 36–37.

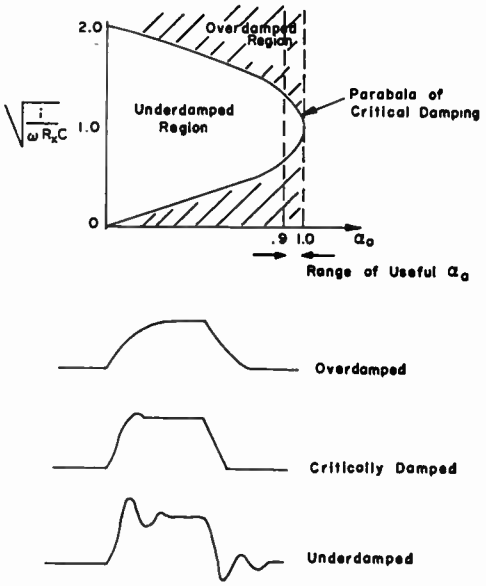


Fig. 19—Emitter follower stability.

Critically damped case:

$$\left[\frac{1}{R_x C} + \omega_c(1-\alpha_0)\right]^2 = \frac{4\omega_c}{R_x C}.$$

Underdamped case:

$$\left[\frac{1}{R_x C} + \omega_c(1-\alpha_0)\right]^2 < \frac{4\omega_c}{R_x C}.$$

A better understanding of how parameter variation affects the circuit and the three cases defined above may be gained by a further study of the critically damped case. This may be written

$$(1-\alpha_0) = (2-X)X$$

where

$$X = \sqrt{\frac{1}{\omega RC}} \quad \text{and} \quad \frac{1}{R_x} = \frac{1}{r_b} + \frac{1}{R}.$$

The expression for critical damping is plotted in Fig. 19. Typical waveforms which illustrate the three cases are also shown. The input network on the emitter followers of Fig. 12 will hold these circuits in the overdamped region of Fig. 19.

VI. Power Drivers

The circuits discussed here have the feature of being able to act as buffer amplifiers between low-power circuits and large resistive and/or capacitive loads. Several of the circuits also are useful for charging or discharging capacitive loads, an essential feature for fast operation. Nearly all of these circuits make use of the saturating

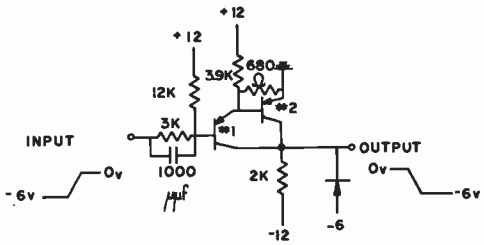


Fig. 20—Power inverter No. 1.

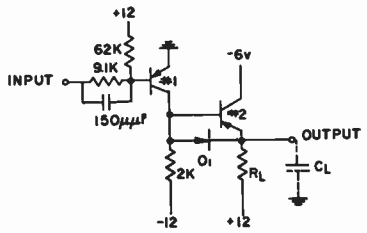


Fig. 21—Power inverter No. 2.

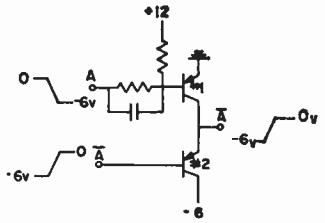


Fig. 22—Cascode circuit.

resistor capacitance-coupled inverter discussed in Section IV.

Fig. 20 shows the power inverter No. 1 circuit, which has some very useful characteristics. It has a large current gain from input to output. This approaches the product of the common emitter current gains of the two transistors used. The high current gain permits the circuit to have a higher input impedance when driving heavy loads than would be possible with a single transistor inverter. A second advantage of the circuit is that the grounded-emitter transistor is not driven into saturation. This follows from the method of coupling between No. 1 and No. 2 transistors. Transistor No. 2 cannot reach saturation because its base-to-collector voltage can never be less than the emitter-to-collector voltage of No. 2 transistor which drives it. Transistor No. 2 can saturate, of course, but this transistor is operated at a relatively low current level and is more easily driven out of saturation. The input circuit is designed so that when the input is at its most negative UP level both transistors No. 1 and No. 2 are biased OFF. Transistor No. 1 is held OFF by the voltage divider action of the 3K and 12K resistors, while transistor No. 2 is biased OFF by the divider network composed of the 3.9K and 680-ohm resistors. The 3.9K resistor supplies the turn OFF current of transistor No. 2. The 1000-$\mu\mu$f capacitor aids in supplying overdrive to transistor No. 1. When the circuit input is down, both transistors No. 1 and No. 2 are in conduction. The input network is designed so that No. 1 transistor is capable of conducting the base current of transistor No. 2 and the current through the 3.9K resistor. The collector currents of No. 1 and No. 2 transistors add at the output.

This circuit has application where large currents with good rise times must be delivered to a load with a minimum input loading. The fall time for the circuit is determined largely by the RC time constant of the collector load. A diode clamp as shown can be used to reduce this time.

Fig. 21 shows a power circuit which is very useful where capacitive loads are to be driven and only one type of transistor is available. It has the capacity for supplying large output currents into or out of the driven load. When the input to the circuit goes negative, transistor No. 1 conducts heavily and supplies current through diode D_1 to the load capacitance C_L, charging it to ground potential. A positive input to the circuit cuts No. 1 transistor OFF. The collector of No. 1 transistor falls negative toward -12 volts. Transistor No. 2 acts much like an emitter follower and charges the load capacitance C_L to the potential -6 volts. The -6-volt and -12-volt supplies could be made the same potential. However, if connected as shown, some improvement in fall time is achieved because of the clamping action of the collector diode of No. 2 transistor on the fall of collector potential of No. 1 transistor. The resistor R_L is not required if the Thevenin's equivalent of the load is returned to a potential more positive than -6 volts.

Fig. 22 shows a cascode arrangement which has been used by Lincoln Laboratory in the TX-0 and TX-2 circuitry. The cascode circuit requires complemented inputs and therefore is most suitable for use as a trigger buffer amplifier. In the TX computers it has been used in this manner for driving from triggers to logic circuitry or to other triggers. Transistor No. 1 is connected as an inverter, while No. 2 transistor is used essentially as an emitter follower. With the input A at zero volts and $\overline{A}$ at -6 volts, No. 1 transistor is cut off and No. 2 transistor holds the output at -6 volts. When input A goes to -6 volts and $\overline{A}$ to 0, No. 1 transistor conducts and brings the output to ground potential, where it is held by the combined action of No. 1 and No. 2 transistors. With No. 1 transistor conducting for a rising output (-6 volts to 0) and No. 2 transistor conducting for a falling output (0 to -6 volts), the circuit can drive capacitive loads positively in both directions.

The complemented inverter shown in Fig. 23 is composed of two saturating resistance capacitance-coupled circuits. The p-n-p inverter is identical to the circuit discussed previously. The n-p-n inverter is also identical in operation, except for the use of a transistor of the opposite type.

The p-n-p transistor is ON during a negative-going input and the n-p-n ON during a positive-going input. These two transistors act in a push-pull fashion to provide a positive drive at the output. No load resistor is

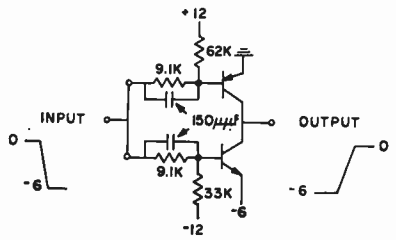


Fig. 23—Complemented inverter.

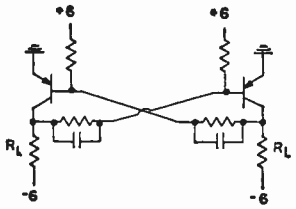


Fig. 24—Basic trigger.

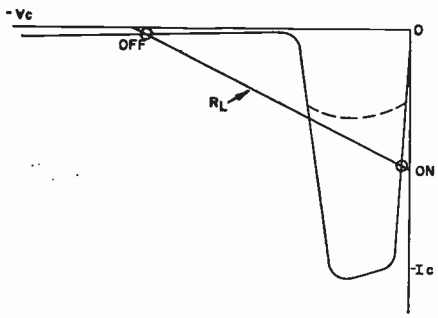


Fig. 25—Collector V_c-I_c characteristic for basic trigger circuit.

needed at the output except the load being driven. Therefore, all output current is available for driving the load itself.

VII. Trigger Circuits

Two of the capacitively coupled inverter circuits of Fig. 10 may be cross-coupled, as shown in Fig. 24, to construct a symmetrical Eccles-Jordan trigger circuit. This particular trigger configuration is widely used, and many different methods of triggering the circuit have been devised. In general, if a collector load resistor is removed, the volt-amp characteristic looking into the collector would be of the form shown in Fig. 25. Triggering may be accomplished either by turning OFF the ON side of the circuit, as represented by the dotted volt-amp plot in Fig. 25, or by turning ON the OFF side of the circuit.

Perhaps the most widely used gating circuits are of the resistance-gated diode type, one example of which is shown in Fig. 26. Here the resistors R_s sense the state of the trigger. The sensing resistors are cross-coupled so that the input diode on the ON side of the trigger is close to conduction. The input diode at the OFF side of the trigger is reverse biased because its sensing resistor is returned to the OFF side of the trigger. When the input signal goes negative, the diode at the ON side conducts and the OFF transistor is turned on through

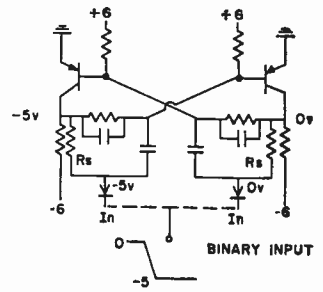


Fig. 26—Negative input binary trigger.

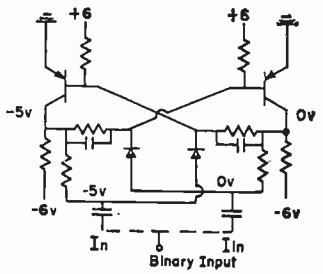


Fig. 27—Positive input binary trigger.

the coupling capacitor. A binary trigger may be formed simply by joining the two gating diodes together.

Another version of the resistance-gated diode is shown in Fig. 27. Here the positions of the diode and the capacitor are interchanged, and the circuit now triggers on a positive-going input. In this circuit the sensing resistors gate the diode at the OFF side of the trigger, and a positive input shuts OFF the ON side of the trigger. As in the circuit of Fig. 26, the two inputs may be joined to form a binary input.

The resistance-gated diode trigger is widely used because it is economical and is easily applied to the problems of rings, counters, and shifting registers. The speed of operation of the circuit (the time from the start of the trigger input to the start of the last moving collector output) generally is slightly slower than two of the inverter circuits of Fig. 10 in series. Speed is limited chiefly by minority carrier storage or by the recovery time of the diode gates. Collector peaking coils and collector clamping diodes are used to improve rise and fall times. The recovery time of the diode gates may also be improved considerably by using emitter followers in the cross-coupling networks. Many other gating circuits have been devised and applied to the basic trigger of Fig. 24.

With the exception of the cascade circuit, all the circuits in Section VI may be used in trigger connections. Two complemented inverters coupled together make a useful trigger which drives heavy loads and operates with fast transition times. A trigger composed of two power inverter No. 2 circuits has similar characteristics.

VIII. Logical Circuits

The resistance capacitance-coupled circuits of Sections IV and V may be combined in a number of ways

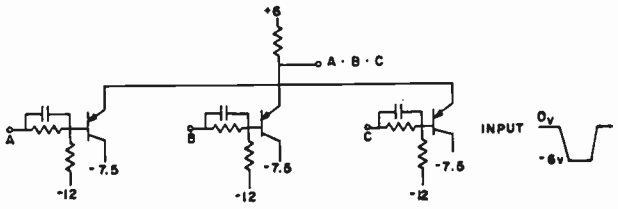


Fig. 28—*P-N-P* emitter follower AND circuit.

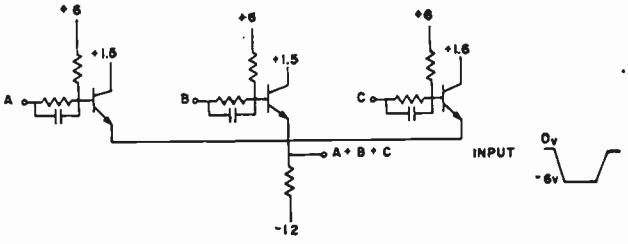


Fig. 29—*N-P-N* emitter follower OR circuit.

o perform logic. In general, they are used in logic much ike vacuum tubes. However, this type of transistor ircuitry offers logical advantages over the vacuum-ube equivalent, because there are two types of transis-ors (*p-n-p* and *n-p-n*). In most logical circuit configura-ions, simply rearranging the potentials to use the pposite type transistor also changes the logical state-nent at the output to a different but useful logical con-nective. In the following discussion, positive logic is ised. A binary "1" is defined as the most positive signal potential, and a binary "0" as the most negative signal potential.

If the above definitions are followed, an AND circuit may be formed by paralleling the basic *p-n-p* emitter follower circuit, as shown in Fig. 28. The output of the emitter follower AND circuit is positive only when all inputs ($A \cdot B \cdot C$) are positive. If one input is negative and all others are positive, then the output is negative and all transistors with positive inputs are cut off. The output of this circuit is controlled by the most negative input. If negative logic is used, the circuit performs an OR function.

When *n-p-n* emitter followers are paralleled, an OR circuit is formed. This arrangement is shown in Fig. 29. In this circuit the output is positive if one or more of the inputs are positive. The output potential follows the most positive input. If negative logic is used, the circuit is an AND circuit.

Emitter followers are commonly used in chains of logic. Since these circuits have no voltage gain and also since there is a shift in output potential, depending on the tolerances of the input network and the transistor, there is a limit to the number of stages that can be connected in series. This limit is determined by the characteristics of the circuit that is used to re-establish the signal level. An inverter or a power driver circuit may be used to perform this function.

The AND and OR connectives may be combined with the INVERSION connective by paralleling the basic

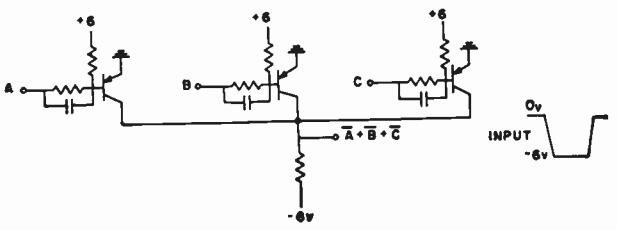


Fig. 30—*P-N-P* inverter NOT AND circuit.

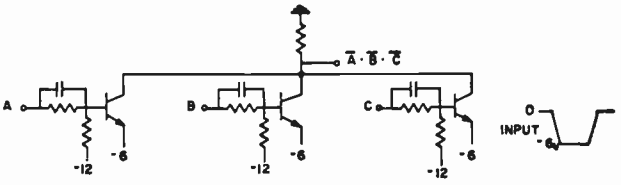


Fig. 31—*N-P-N* inverter NOT OR circuit.

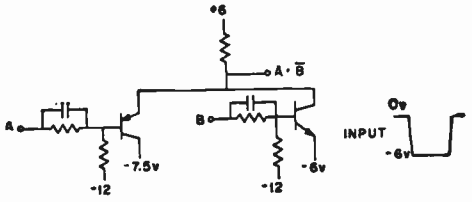


Fig. 32—*N-P-N* inverter, *p-n-p* emitter follower combination.

inverter circuit of Fig. 30. Here the basic inverters are paralleled to generate the NOT AND connective. In this circuit, a negative input at one or more inputs drives the transistor into saturation and the output potential rises to ground. The output can be negative only if all of the inputs are positive. Therefore, the arrangement is logically a NOT AND circuit.

An *n-p-n* version of the circuit is shown in Fig. 31. Here the output is positive only if all the inputs are negative. However, the output is negative if one or more of the inputs are positive. Therefore, the circuit performs a NOT OR operation.

Further logical combinations may be formed by combining the emitter follower and inverter circuits. Since the load resistors of *p-n-p* emitter followers and *n-p-n* inverters are returned to a positive voltage, they may be paralleled as shown in Fig. 32 to form a different logical statement. In this circuit the output can be positive only when the input to the emitter follower is positive and the input to the inverter is negative. When the input to the inverter is negative, that circuit is cut off and the output potential rises until it is clamped by the emitter follower.

The complementary version of the above circuit is shown in Fig. 33. This circuit can invert and perform an OR function. The output is positive if the input to the emitter follower is positive or the input to the inverter is negative. When the input to the inverter is negative, that circuit is cut off and the output potential falls until it is clamped by the emitter follower.

An EXCLUSIVE OR circuit could be formed from two of the emitter follower, inverter circuits of Fig. 33

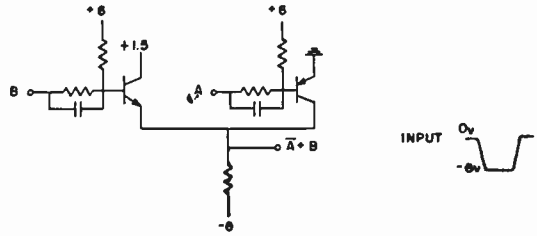


Fig. 33—*P-N-P* inverter, *n-p-n* emitter follower combination.

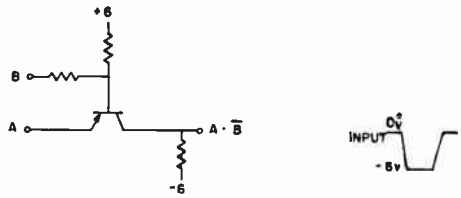


Fig. 34—Single transitor $A\cdot\overline{B}$ circuit.

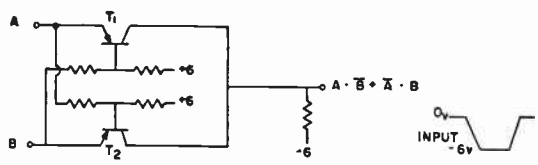


Fig. 35—Two-transistor EXCLUSIVE OR circuit.

and a two-way *n-p-n* emitter follower OR circuit. Another way to perform the EXCLUSIVE OR function is suggested by the circuit shown in Fig. 34. Here both the emitter and the base are used as inputs. The output is positive only when the emitter input is positive and the base input is negative. An EXCLUSIVE OR circuit can be formed by cross-coupling two of the circuits of Fig. 34, and this is shown in Fig. 35. Because the circuits are cross-coupled, the transistors can conduct only when there is a difference between the input signals. That is, if both A and B are either negative or positive, the transistors are cut off. When the input at A is positive and the input at B negative, transistor T_1 conducts. When the input at B is positive and the input at A is negative, transistor T_2 conducts. The disadvantage of this circuit is that the emitters must be driven and there is no current gain.

Diode logic is often used in combination with the circuits discussed in this section, because it offers possible economic advantages over transistor circuitry. Emitter followers, inverters, or power drivers may be used to drive diode logic. Several ways of driving diode circuitry are shown in Figs. 36 and 37. In Fig. 36, an *n-p-n* emitter follower is shown driving a diode AND circuit. The ratio of R_A and R_L should be such that the emitter follower conducts when the input is negative. In the inverter circuit, diode C is used to clamp the OFF level of the inverter to ground and establish the input level to the diode AND circuit. A clamped *p-n-p* inverter also may be used to drive a diode AND circuit.

In Fig. 37, a *p-n-p* emitter follower and a *p-n-p* clamped inverter are used to drive a diode OR circuit.

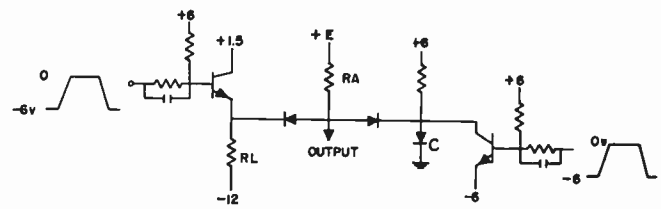


Fig. 36—Inverter and emitter follower driving a diode AND circuit.

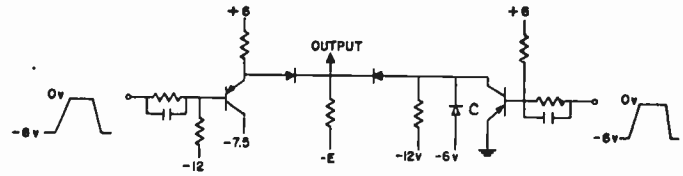


Fig. 37—Inverter and emitter follower driving a diode OR circuit.

A clamped *n-p-n* inverter also could be used as a driver. There would be little point in using *p-n-p* emitter followers to drive diode and circuits, or *n-p-n* emitter followers to drive diode OR circuits. This is because simply paralleling these emitter followers would form the necessary logical circuit without using the diodes. After a signal has passed through several stages of logic, it is shifted positive or negative, depending on the diode tolerances, and it is necessary to re-establish the correct signal level to switch other circuits reliably. This is the same problem encountered when using emitter follower logic. The circuits used to re-establish the signal levels are generally designed to accept signal swings much smaller than those normally applied at the input of the diode chain. This is done to allow several diode stages to be connected in series.

IX. Direct-Coupled Transistor Logic

This system of circuits, abbreviated as DCTL, has found relatively wide acceptance in the computer field.[7] The basic circuits have a high degree of simplicity and enable transistors to perform in a manner similar to relay contacts, *i.e.*, simple parallel and series connections to perform OR and AND operations.

Fig. 38 shows a series chain of the basic amplifiers. The circuits are characterized by the absence of series-coupling networks. The load for transistor No. 1 is the base input characteristic of transistor No. 2 in parallel with the 1K load resistor returned to -3 volts. Fig. 39 shows this composite load plotted on the collector characteristic of transistor No. 1. The ON state of transistor No. 1 (high current state) brings its collector voltage into the region of 20 to 60 mv. In the OFF state, the output voltage of transistor No. 1 is at approximately -0.39 volt. It is evident from Fig. 3 that a transistor remains relatively OFF with small forward biases from emitter to base. Therefore, when transistor No. 1 is driven into saturation, its collector-to-emitter voltage

[7] A. L. Cavalieri, Jr., "What's inside Transac-1," *Electronic Design*, vol. 4, pp. 22–25; July 1, 1956.

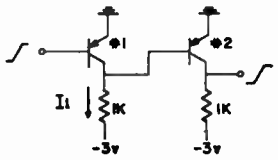


Fig. 38—Series-connected direct-coupled transistors.

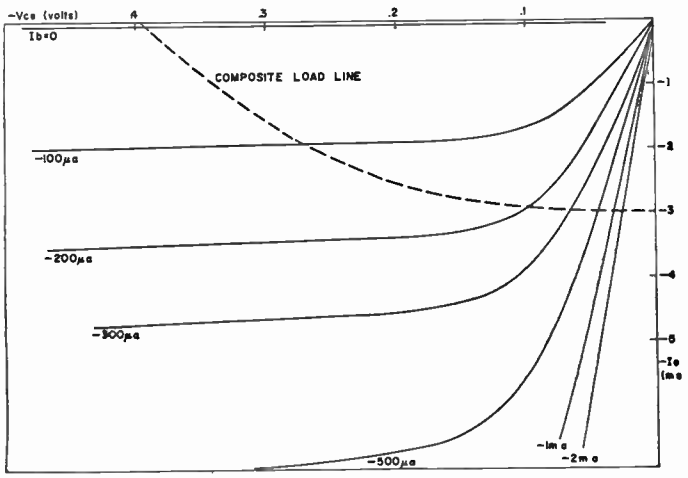


Fig. 39—Collector characteristic of transistor No. 1 with composite load line formed by the base input characteristic of transistor No. 2 and 1K returned to −3 volts.

must become sufficiently small to hold transistor No. 2 OFF. Most alloy transistors can be used in this manner. When transistor No. 1 is held OFF, the current I_L must flow from the base of transistor No. 2, turning it ON. Because of the very small signal swing at the collector of transistor No. 1, the 1K load resistor and the −3-volt supply essentially form a constant current source, and the current I_L is switched between the collector of No. 1 transistor and the base of No. 2 transistor. It is seen, therefore, that the dc characteristics of these transistors are such that by turning transistor No. 1 ON and OFF it is possible to couple its collector directly to the base of No. 2 transistor and turn this transistor OFF and ON. The only requirements for operation of this chain are that the saturation voltage drop of the ON transistor be sufficiently small to hold the following transistor in the OFF state, and that the base-to-emitter current gain of the transistors be such that the base current I_L minus any leakage current from the collector of No. 1 transistor (when OFF) is sufficient to drive No. 2 transistor into saturation.

In order to construct useful computer circuits, it is essential that the output of one transistor be able to drive more than one circuit. Fig. 40 shows the simple manner in which branching is achieved in DCTL circuits. The number of bases that one transitor can drive, in a configuration such as shown in Fig. 40, is fixed by the current gains of the transistors and the degree of similarity of the base input characteristics. For example, if (as shown in Fig. 40) transistor No. 2 saturated with −0.3 volt at its base and transistor No. 3 did not saturate until its base potential reached −0.4 volt,

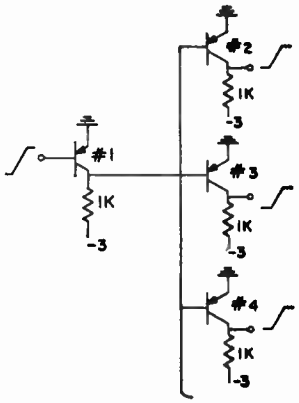


Fig. 40—Branching of DCTL amplifiers.

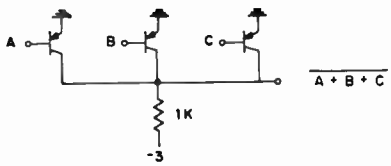


Fig. 41—DCTL NOT OR circuit.

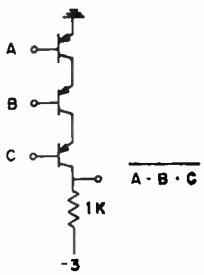


Fig. 42—DCTL NOT AND circuit.

then it is evident that transistor No. 2 would "rob" base current from transistor No. 3 and could possibly keep it from reaching full conduction. It is evident that a controlled uniformity of input characteristics is a requirement for DCTL connections. Fortunately, transistors of a single type are relatively uniform in this respect.

Transistors connected as shown in Figs. 38 and 40 perform inversion. An input at the base of transistor No. 1 would be inverted at the collector of No. 1 and reinverted at the collector of No. 2. If we use negative logic, that is, the most negative potential at the base of No. 1 transistor would be represented as a "1" and the most positive signal (which turns "1" transistor OFF) as a "0," then we can perform a NOT OR function with the parallel connection of transistors shown in Fig. 41 and a NOT AND function with the series connection shown in Fig. 42. The operation of the circuit of Fig. 41 is straightforward and noncritical. Whenever a "1" (most negative signal) is present at inputs A or B or C, the output is at its most positive or "0" state. The output can only be a "1" if we do *not* have A or B or C present.

The operation of the circuit of Fig. 42 is more restrictive. Here the inputs to A and B and C must all be at

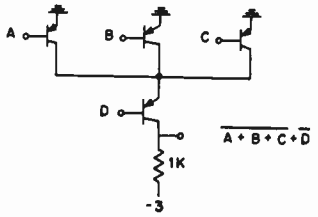


Fig. 43—Parallel-series transistor connection.

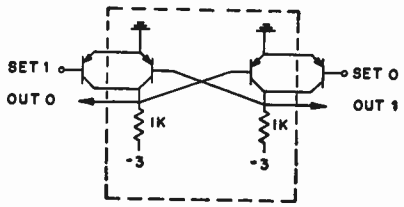


Fig. 44—DCTL flip-flop.

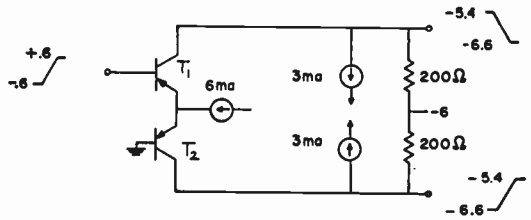


Fig. 45—Basic *p-n-p* current switch.

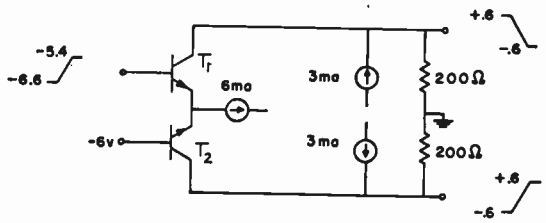


Fig. 46—Basic *n-p-n* current switch.

their "1" (most negative) state to bring the output to a "0" state. If A and B and C are *not* all present, the output is "0." The restriction on this circuit is the limited number of transistors which can be put in series and still assure that the output will be sufficiently positive, when all transistors are ON, to hold OFF any transistors this output may be driving. Since the transistors are in series, the "0" output voltage is the sum of the saturation drops of the transistors in the string. It is evident that the saturation voltage drops of the transistors used should be as small as possible. Further complications arise because input C requires a more negative signal to turn its transistor ON than is necessary for this operation at input B. Input B also requires a slightly more negative signal than input A. For the above reasons, the use of the series circuit is generally restricted to a series chain of two or possibly three transistors. Other restrictions may be placed on what the output of this circuit is allowed to drive.

Logical connections need not be limited to simple series or parallel connections. Combinations of these circuits can be used to give more complex logical expressions. An example of this is given in Fig. 43.

The basic flip-flop is formed in this system, as in most others, by a series connection of two inverters. The circuit for this is shown within the dashed lines of Fig. 44. The two transistors outside of the dashed lines are pullover transistors and are used for setting and resetting the circuit. A "1" input into the set "0" side of the flip-flop causes the pullover transistor to conduct and pull the OUT "1" output to near ground potential. This in turn causes the OUT "0" output to go to its most negative or "1" state. If the set "0" input is returned now to "0," the flip-flop will hold this condition. To reverse the state of the flip-flop, a "1" is inserted into the set "1" input, thereby causing the OUT "1" output to go to the "1" state.

The salient advantages and disadvantages of DCTL are summarized as follows.

Advantages

1) Minimum number of component types.
2) Single power supply.
3) Simple configurations.
4) Low-power dissipation.

Disadvantages

1) Close transistor tolerances are required.
2) The small-signal swings make very good grounding necessary.
3) Speed is limited by transistor saturation.

X. Current Switching Circuits

A basic current switching circuit is shown in Fig. 45. This is a differential amplifier in which the signal is applied to one base and the other base is returned to a reference voltage which is ground for the circuit shown here. When the input potential is at -0.6 volt, transistor T_1 is conducting and 6 ma from the emitter current source flows into transistor T_1, causing its output potential to rise from -6.6 to -5.4 volts. The input potential of -0.6 volt is sufficient to bias transistor T_2 off, and its output potential falls to -6.6 volts. When the input potential is raised to $+0.6$ volt, transistor T_1 is shut off and conduction switches to T_2. This circuit illustrates a method by which a well-defined current is switched by a relatively small input signal. The input signal need only be large enough to insure that the potential at the emitter node rises above or below ground.

From Fig. 45, it is seen that the collector is returned to -6 volts through a small load resistor, and a suitable current bias is applied to obtain a signal swing about -6 volts. Because of the 6-volt difference between input and output, a *p-n-p* current switch cannot drive another *p-n-p* current switch. This difficulty can be overcome by constructing the complementary *n-p-n* switch shown in Fig. 46.

In the *n-p-n* circuit of Fig. 46, the reference input is returned to -6 volts so that this circuit can be driven

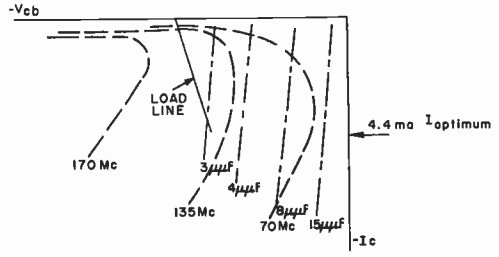


Fig. 47—f_{ro} and C_r contours.

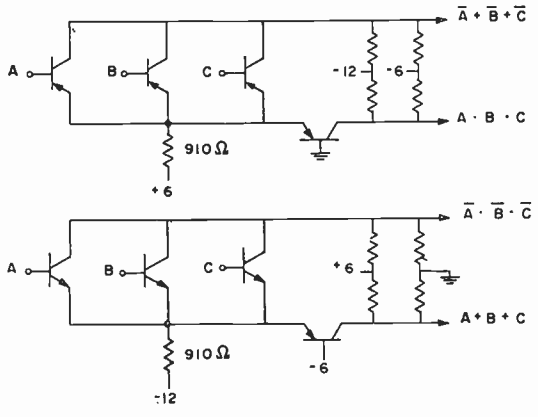


Fig. 48—*P-N-P* and *n-p-n* logical circuits.

by the *p-n-p* switch. Also, the output of the *n-p-n* switch is referenced to ground and is capable of driving a *p-n-p* switch. The *n-p-n* and *p-n-p* switches are basic to all the current switching circuits discussed here. A basic requirement of current switching circuits is that *p-n-p* circuits drive *n-p-n* circuits, which in turn always drive *p-n-p* circuits.

The basic circuits of Figs. 45 and 46 can be made to switch very rapidly by optimum choice of the dc operating conditions. As discussed in Section III and shown in Fig. 4, the frequency response and the collector capacitance can be very much a function of the dc operating conditions. The curves of Fig. 5 are repeated in Fig. 47, and a load line X is added which represents the situation in a current switching circuit. Note that, except for the region of low-frequency response at low currents, the operating point always remains in a region where frequency response is high and collector capacitance is low. Also, by proper choice of the emitter current source, it is possible to switch an optimum transistor current, although stray circuit capacitance and load capacitance generally dictate a somewhat higher current for optimum circuit speed.

The basic switches of Figs. 45 and 46 may be extended to perform logic by the addition of transistors in parallel with the top transistor. The circuits are shown in Fig. 48. Here a binary "1" is defined as the most positive input to a switch, regardless of whether the signal line in question is referenced to ground or to -6 volts. One logical feature of current switching circuits apparent in Fig. 48 is that there are two logical outputs

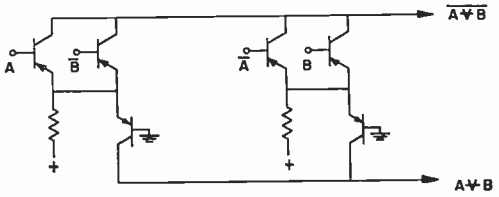


Fig. 49—EXCLUSIVE OR circuit.

and they are complements of each other. As a logical "1" is defined here, the top outputs are inverted and the bottom outputs are normal. The fact that inverted outputs are always available eliminates the need for a separate inverter building block in a system. This can be an advantage by reducing the over-all delay in high-speed systems where long chains of logic require frequent inversion.

Although the *p-n-p* and *n-p-n* logical circuits of Fig. 48 perform the necessary connectives required in a computer, a separate EXCLUSIVE OR circuit is often desirable, particularly if it is more economical or faster. Such is the case for the EXCLUSIVE OR circuit of Fig. 49. This EXCLUSIVE OR circuit consists of two parallel AND circuits which generate $A \cdot \overline{B}$ and $\overline{A} \cdot B$. With the four inputs connected as shown, only one of the AND circuit outputs can be conducting at any one time. Therefore, the AND circuit outputs can be connected to form the OR circuit required to complete the EXCLUSIVE OR function. Under the input conditions $\overline{A} \cdot \overline{B}$ or $A \cdot B$, the inverted outputs of both AND circuits are conducting and twice the normal current flows into the load network. This network is designed to give a normal output only when both sides are conducting. In this way, the inverted EXCLUSIVE OR statement is obtained. The inverted outputs supply only one unit of current when the normal EXCLUSIVE OR inputs ($\overline{A} \cdot B$ and $A \cdot \overline{B}$) are present.

No discussion of a set of switching circuits would be complete without reference to some means of storage and some means of generating a well-defined pulse. The storage circuit in this case is shown in Fig. 50. It consists of two basic switches cross-coupled to form a symmetrical bistable flip-flop. The flip-flop can be set in through the pull-over transistors on either side. An OR function can be built into the flip-flop by paralleling the pull-over transistors. Operation is in no way different from the logical block circuits of Fig. 48.

A basic current mode single shot is shown in Fig. 51. The circuit again consists of the two basic switches cross-coupled together. However, in this case, to obtain monostable operation one side is coupled through a short-circuited delay line. The bottom *n-p-n* transistor is biased on by the network at its base. The short-circuited delay line is terminated at the sending end, and the pulse width at the output is determined almost entirely by the time required for the wave front at the delay line input to travel down and then back up the delay line.

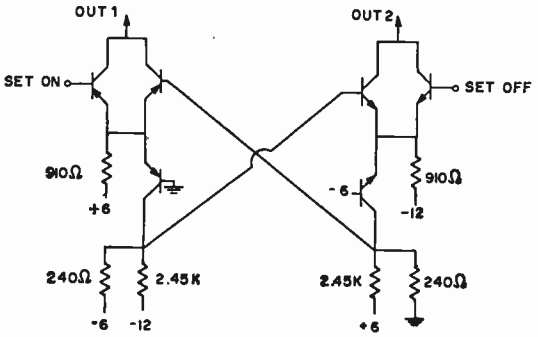


Fig. 50—Current switching trigger.

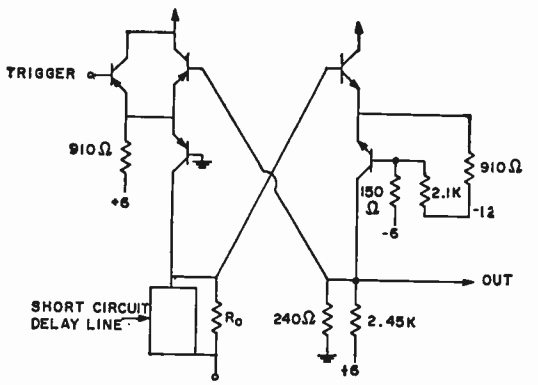


Fig. 51—Current switching single shot.

In any computing system, it is sometimes necessary to drive loads located at a considerable distance from the driving source. This may be done by driving conventional coaxial transmission line and terminating the coaxial line with either of the circuits seen in Fig. 52.

Two arrangements are shown here, both driven by a basic switch. The *n-p-n* line terminator shown in the top of Fig. 52 is a Class-A grounded-base amplifier. When the top output of the basic switch driving the line is off, the *n-p-n* transistor conducts and 6.5 ma flows into the current sink formed by the 660-ohm resistor and the −12-volt supply. When the top transistor conducts, the emitter current of the *n-p-n* grounded-base stage is reduced to 0.5 ma. The input impedance of this stage has a small inductive component and an impedance of 11 ohms. The 82-ohm resistor is added to increase the total impedance to 93 ohms, and match the characteristic impedance of the line. The small capacitor compensates for the inductive input component. The value of the series resistance can be changed to match lines of different characteristic impedance, if desired. The *p-n-p* line terminator shown at the bottom of Fig. 52 operates in the same way. In this circuit, the base is biased to −3 volts so that the output signal will be referenced to −6 volts. The *n-p-n* circuit differs from the *p-n-p* circuit in that it translates the output of the basic *p-n-p* switch from −6 volts up to ground level. Because of this, the *n-p-n* grounded-base amplifier can also be used as a coupling means between two *p-n-p* logical blocks.

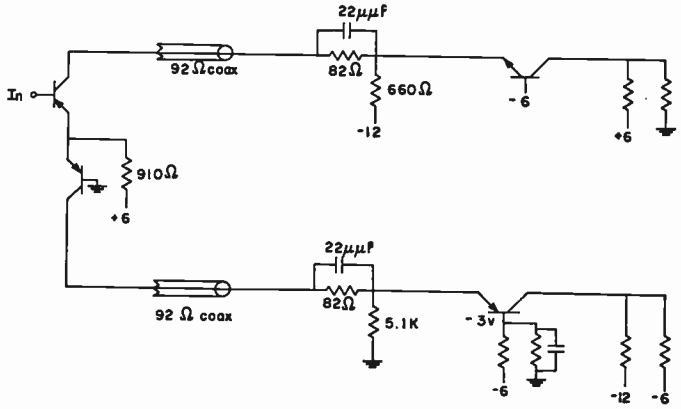


Fig. 52—Transmission line terminators.

Only a few of the variations of current switching have been discussed here. As with other basic circuit approaches, many different circuits are possible. The transistor count can be reduced by replacing the grounded-base transistors with diodes and sacrificing the in-phase output. One transistor type can be used by inserting voltage-translating blocks between circuits of a single type. Such a block might be a reverse-biased Zener diode.

Current switching has the disadvantages of using a relatively large number of transistors, and of requiring more than one power supply. The compensating factors are the high operating speed and the elimination of saturation transistor specifications. The circuits exhibit good noise tolerance and tend to generate less noise because of the current switching feature.

XI. Conclusions

As we have seen, transistors can be applied in many ways, and no single approach has all the advantages or disadvantages. In the problem of selecting a circuit type, the primary considerations are the performance level required and the transistor types available. For applications where operating frequencies are well within the capabilities of the available devices, the resistor logic of the NOR circuit, or variations on it, offers attractive economic advantages. In the intermediate speed range, resistance capacitance-coupled circuits with diode or transistor logic generally offer a satisfactory compromise between cost and performance. The simplicity of DCTL will also appeal to many designers. In the high-speed area, current switching techniques offer attractive speed advantages.

Application of Transistors in Communications Equipment*

DAVID D. HOLMES†, MEMBER, IRE

Summary—**Communications equipment in its various forms poses design problems which encompass nearly all facets of the electronics art. Transistors, as they exist today, can perform a large percentage of the functions required in comminications systems.**

A brief discussion of a number of circuit techniques which are useful in the application of transistors to communications equipment is presented. Representative equipments, which indicate the state of the art, are described.

Introduction

THE relatively long design-production cycle which applies to communications equipment, in contrast to the rapid introduction of innovation in the home entertainment field, has been a mild deterrent to the large-scale appearance of transistors in commercial and military communications apparatus. In addition, the design of communications equipment encompasses more problems in that the equipment must ofttimes meet more stringent requirements in respect to performance, reliability, environment, etc. Transistors are only beginning to come into their own in this area which is perhaps the senior element of the electronics art.

Included in this paper is a brief discussion of certain circuit-design techniques which are pertinent to the application of transistors in communications equipment. This presentation is by no means complete; the intention is to provide some helpful clues and references. A small sampling of transistorized communications equipment is presented to serve as an indication of the types of application which are practicable, and of the kind of performance to be expected at this point in time.

Digital Communications Systems

Transistor circuits which are required to perform logical operations in digital communications systems are similar in form and function to those employed in digital computers. A detailed discussion of such circuits is not given here, but can be found elsewhere in this issue.[1]

A by-product deriving from utilization of transistors has been the development of a circuit-module concept wherein a relatively small number of circuit types can be arrayed to perform many different operations. Transistors have helped in paving the way toward a digital systems design approach where "standard" circuits which perform AND, OR and NOT logical functions, and which can be interconnected to serve as binary flip-flops for timing and counting, are basic components, rather than resistors, capacitors, or transistors alone. With the circuit building blocks at hand, one works with a functional systems diagram rather than with detailed circuit diagrams.

In addition to obvious size, weight, and power advantages, transistors provide a degree of reliability heretofore not realizable with vacuum tubes. Properly designed transistor circuits afford long-term reliability which is orders of magnitude superior to their vacuum-tube counterparts, and which tends to offset potential first-cost disadvantage through drastically reduced down-time and maintenance expense.

Recent developments indicate that transistors are only the beginning in the exploitation of solid-state phenomena which can be used to advantage in digital circuits. For example, a "shift-register transistor," which employs a single semiconductor block and multiple junctions, appears feasible. Here, many of the circuit functions would be built into the device.

Switching speeds required for the generation and processing of the bulk of digital signals found in contemporary communications systems are relatively slow, principally due to the bandwidth limitations of existing radio and wire circuits and speed limitations of mechanical input-output devices. Transistors having response in the 100-kc range can perform a majority of digital communications systems functions satisfactorily. On the other hand, there is a growing need for high-speed communication of digital information between computers. Many types of devices are being developed now which can switch at speeds in the millimicrosecond range. One promising device is the "Thyristor,"[2,3] shown in Fig. 1. The Thyristor is a *p-n-p*-like structure which exhibits a negative-resistance characteristic; this characteristic is the result of electron injection at the collector contact. The Thyristor functions as a conventional high-speed transistor when the collector current is kept below a critical value. If the collector current exceeds this value, the device rapidly switches to a high-conductance mode, where collector currents in the order of 100 ma can be conducted at a

* Original manuscript received by the IRE, March 21, 1958. The AN/TCC-26, AN/PRC-25, AN/PRC-34, and Marker Beacon Receiver discussed in this paper were developed by RCA Defense Electronic Products, Camden, N. J., and sponsored wholly or in part by the U. S. Army Signal Engineering Labs.

† RCA Laboratories, Rocky Point, N. Y.

[1] R. A. Henle and J. L. Walsh, "Application of transistors in computers," this issue, p. 1240.

[2] C. W. Mueller and J. Hilibrand, "The Thyristor—a new high-speed switching transistor," IRE TRANS. ON ELECTRON DEVICES, vol. ED-5, pp. 2–5; January, 1958.

[3] L. E. Barton, "The Thyristor—some of its characteristics and applications," *Electronic Design*, vol. 6, pp. 48–51; March 19, 1958.

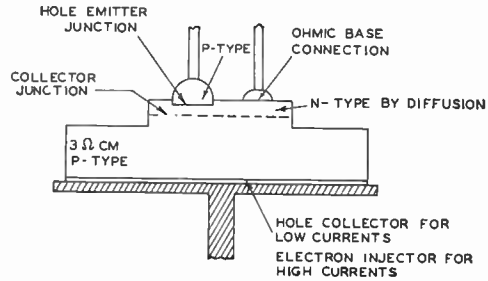


Fig. 1—The "Thyristor."

collector-to-emitter voltage drop in the order of 0.5 volt. While the Thyristor resembles a thyratron in turn-on action, it differs from a thyratron in that it can be turned off by application of a reverse base voltage, without removal of collector voltage. A typical Thyristor characteristic is shown in Fig. 2. Devices of this type have shown rise times in the order of 10 mμsec.

An example of a transistorized pulse-position-modulated time-division multiplex equipment, the AN/TCC-26, is seen at the left in Fig. 3. This is a 23-channel multiplexer-demultiplexer unit capable of being paired with another identical unit to provide a 46-channel system. The ratio of signal-to-noise plus crosstalk between fully modulated channels exceeds 60 db. The equipment utilizes 289 transistors; its input power requirement is only 65 watts.

While the AN/TCC-26 represents a very considerable size reduction in comparison with its vacuum-tube predecessor (the AN/TCC-13), the development of a Micro—Module concept indicates the possibility of a further drastic size reduction, as illustrated at the right of Fig. 3.

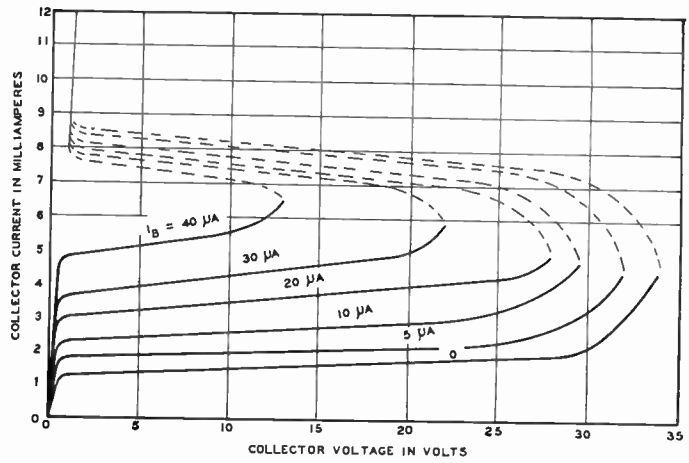


Fig. 2—Typical Thyristor collector family.

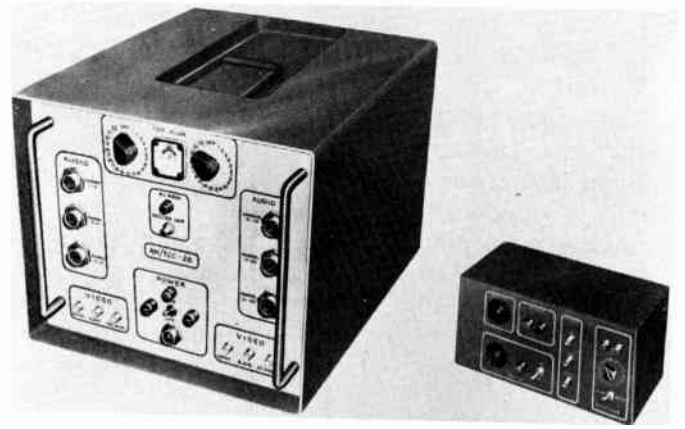


Fig. 3—AN/TCC-26 time-division multiplex equipment (left) and its projected Micro—Modular equivalent (right).

Signal Generation and Modulation

Transistor oscillators can assume the same variations of form as do vacuum-tube oscillators. The particular circuit combination which is chosen for a given application depends upon requirements such as frequency and amplitude stability and power output. At signal frequencies which are an appreciable fraction of the α-cutoff frequency, feedback paths within the device become increasingly important and are likely to influence the choice of an oscillator-circuit configuration.

The extent to which the oscillator frequency is susceptible to variations in power supply voltage and ambient temperature depends upon the degree of coupling between transistor and circuit, insofar as the effect of transistor parameter variation is concerned. The degree of decoupling which can be achieved is dependent upon the maximum available gain of the transistor at the frequency in question. Generalizing, maximum frequency stability is obtained when circuit parameters are judiciously chosen and apportioned in such a way that the oscillator feedback-coupling circuit loss is approximately equal to the maximum available transistor gain. Insight to the basic principles which apply here can be derived from a subsequent section in which factors influencing the instability of a tuned amplifier are mentioned. In that section, a proportionality between transistor loading (ratio of operating circuit Q to unloaded coil Q) and coupling-circuit loss is indicated.

Frequency or amplitude modulation of a transistor oscillator can be achieved by various means; care must be exercised in order that both types of modulation do not occur simultaneously. Lin[4] has shown the effects of collector-voltage and emitter-current modulation upon both frequency and amplitude of oscillator output, and has suggested means for obtaining relatively pure AM or FM through simultaneous modulation of both parameters.

Other compensating methods can be employed. For example, consider the oscillator circuit shown in Fig. 4(a). Feedback from collector to base is obtained via a secondary winding which is closely coupled to the tuned collector circuit. The oscillator output amplitude can be made very nearly a linear function of dc emitter current. However, when AM is imparted through variation of I_e, a degree of attendant FM will obtain. This follows

[4] H. C. Lin, "Modulated Transistor Oscillators and Their Applications" in "Transistors I," RCA Labs., Princeton, N. J., pp. 547–560; 1956.

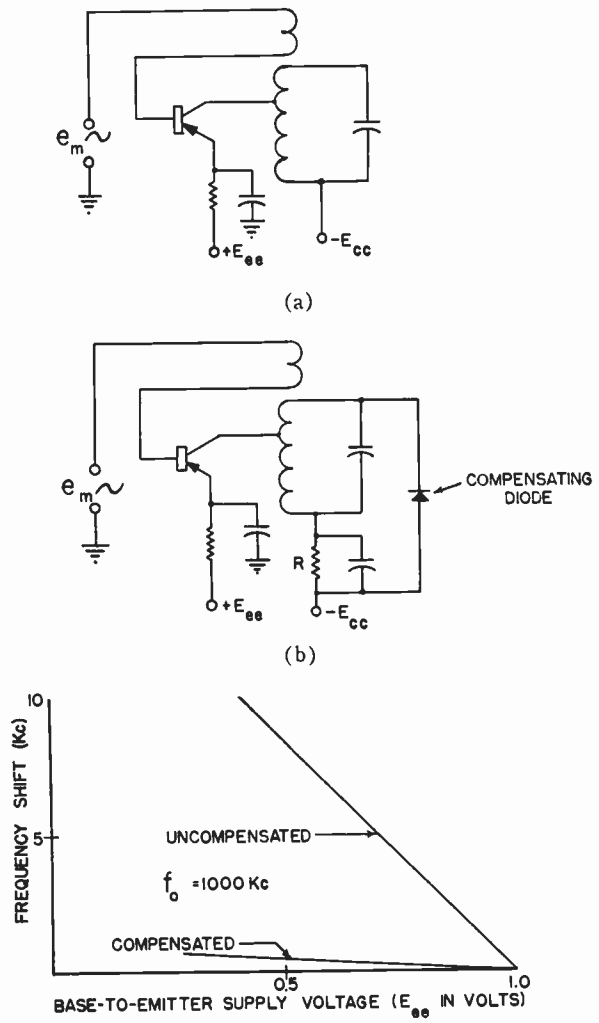


Fig. 4—Diode frequency-shift compensation. (a) Uncompensated oscillator. (b) Oscillator with compensation added. (c) Illustration of degree of compensation.

from the fact that as I_e is increased, the emitter diffusion capacitance increases; a proportionately larger capacitance thus is reflected across the tuned circuit.

An arrangement which utilizes a compensating junction diode is seen in Fig. 4(b). The compensating diode is reverse biased by the voltage developed across resistor R in the collector circuit; this voltage is a direct function of I_e. The diode capacitance variation, which is an inverse function of the voltage appearing across R, will tend to compensate for the frequency shift produced by variation of the emitter junction diffusion capacitance. An indication of the degree of compensation which can be attained is shown in Fig. 4(c).

RF Power Amplification

Nearly ideal power amplifiers can be realized with junction transistors because of their characteristically low collector-voltage knee. The dissipation ratings of transistors designed for operation at VHF and UHF thus far have been predominantly tens of milliwatts, so that available power output is relatively small.

Power amplifier design procedure is straightforward at frequencies which are a small fraction of the α-cutoff frequency. As the operating frequency is increased to a value which is an appreciable fraction of the α-cutoff frequency, the driving power for a given power output must be increased. Input circuit dissipation can become an appreciable fraction of the total allowable device dissipation and must be taken into account. An illustration of the state of the art, insofar as RF power output capability is concerned, is the AN/PRC-35,[5] a completely transistorized FM transmitter-receiver which provides a power output of several hundred milliwatts in the frequency range of 30 to 70 mc.

Receiver Circuits

Improvements in transistors, and the development of circuit design techniques, have progressed to the point where excellent transistor receiver performance, in terms of noise factor, selectivity, automatic gain control, and cross modulation, can be attained for signal frequencies extending into the VHF range. Results obtained, for example, with experimental VHF television receivers have been quite promising, noise factors in the 10- to 15-db range over the VHF band having been realized.

In this section, design techniques relating to various receiver functions are discussed briefly.

RF Amplifiers

Compromises which can be made between selectivity and noise performance of an RF amplifier have been outlined by Freedman.[6] While it is noted, for example, that near-optimum noise performance is obtained when the receiver input tuned circuit is matched to the transistor input resistance (operating circuit Q equal to one-half unloaded coil Q), an advantageous trade can be made wherein a considerable gain in selectivity can be obtained at the expense of only a small sacrifice in noise performance. Where two RF tuned circuits (antenna and interstage) are employed, there exists a unique apportionment of selectivity between the two circuits which provides maximum noise performance for a given over-all selectivity requirement.

It has been shown[7] that the performance of a transistor RF stage with respect to cross modulation can be comparable to that of a variable-mu pentode. Cross modulation in any device depends upon the slope of its log gm-vs-input-circuit-bias characteristic.[8] Such a

[5] This equipment is being developed under contract with the U. S. Army Signal Eng. Lab. by RCA Surface Communications Dept., Defense Electronic Products, Camden, N. J.

[6] L. A. Freedman, "Design considerations in the first stage of transistor receivers," 1957 IRE National Convention Record, pt. 3, pp. 182–192.

[7] D. D. Holmes, "Cross Modulation in Transistor RF Amplifiers" in "Transistors I," RCA Labs., Princeton, N. J., pp. 422–430; 1956.

[8] "Application of the Electronic Valve in Radio Receivers and Amplifiers," Philips Technical Library, Book IV, p. 325.

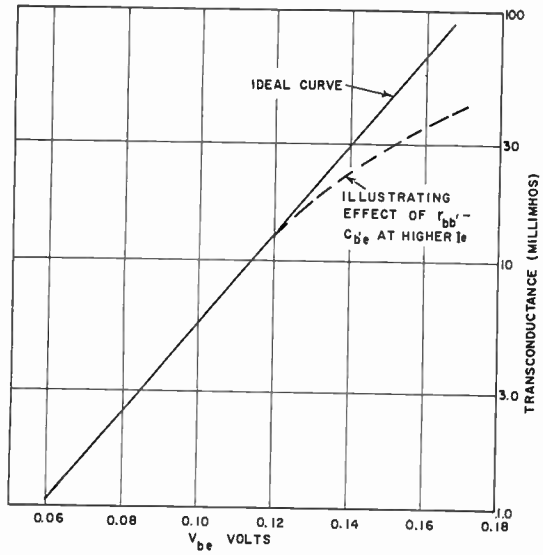


Fig. 5—Transconductance vs base-to-emitter voltage for ideal transistor (solid curve) and illustration of effect of $r_{bb'}$–$c_{b'c}$ at higher emitter currents (dashed curve).

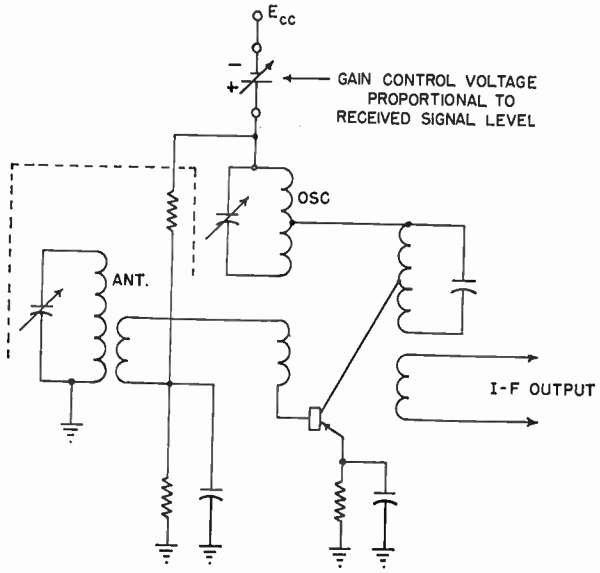


Fig. 6—Conversion gain control by simultaneous reduction of collector voltage and emitter current.

characteristic for an alloy-junction transistor is illustrated in Fig. 5. It can be seen that this is a straight line for an ideal transistor (no base-lead resistance), indicating that cross modulation will be independent of operating point over a wide range of input circuit bias. At higher values of *Vbe* (higher *Ic*) cross-modulation performance improves due to the larger voltage-dropping effect of the base-lead resistance. Under very strong signal conditions, improved performance can be obtained by arranging the AGC circuit in such a way that the base-emitter junction is reverse biased by an amount in excess of the peak signal voltage. As in vacuum-tube receivers, the distribution of gain control among several stages should be arranged to provide a best compromise among noise performance, cross modulation, and overload characteristics.

Mixers and Converters

Mixing action can be obtained by impressing signal and local oscillator voltages in series across the base-to-emitter junction of a transistor. Conversion gain increases rapidly with increasing oscillator injection, ultimately reaching a maximum value which, interestingly, is very nearly equal to the signal-frequency gain of the transistor operating as an RF amplifier. Beyond the point of maximum conversion gain, the gain decreases slowly with increasing oscillator voltage.

A converter can be obtained by combining oscillator and mixer functions in a single transistor. Examples of an oscillator mixer and a converter circuit can be found in the literature.[9,10]

A modicum of conversion gain control can be obtained by reducing oscillator injection, or (in the case of a converter) by reducing the dc collector voltage until the oscillator voltage swing in the collector circuit carries over the knee of the collector characteristic, *i.e.*, "bottoming" occurs. A control range in the order of 30 db can be obtained by these means.

Precautions should be taken against incidental oscillator frequency shift with application of conversion gain control. When gain control is effected through reduction of converter dc emitter current, diode compensation similar to that described in the second section can be employed. An arrangement which embodies a combination of emitter current and collector voltage control is shown in Fig. 6. Here, the collector and base return to a common negative supply which is made to vary in proportion to the input signal level. The collector voltage and emitter current, then, decrease with increasing input signal level so that the effect upon oscillator frequency of increasing collector capacitance is offset by the opposite effect of decreasing emitter-to-base diffusion capacitance.

IF Amplifiers

One of the most important design problems in transistor IF amplifiers, or in tuned amplifiers generally, is stability in the signal-feedback sense. The general problem was recognized early in vacuum-tube art,[11] and more recently a technique for dealing with the problem of reactive feedback in transistor amplifiers has been developed.[12] This technique relates tuned amplifier stability (or tendency toward instability) to interstage

[9] L. A. Freedman, T. O. Stanley, and D. D. Holmes, "An experimental automobile receiver employing transistors," PROC. IRE, vol. 43, pp. 671–678; June, 1955.

[10] D. D. Holmes, T. O. Stanley, and L. A. Freedman, "A developmental pocket-size broadcast receiver employing transistors," PROC. IRE, vol. 43, pp. 662–670; June, 1955.

[11] B. J. Thompson, "Oscillation in tuned radio-frequency amplifiers," PROC. IRE, vol. 18, pp. 421–437; March, 1931.

[12] D. D. Holmes and T. O. Stanley, "Stability Considerations in Transistor Intermediate-Frequency Amplifiers" in "Transistors I," RCA Labs., Princeton, N. J., pp. 403–421; 1956.

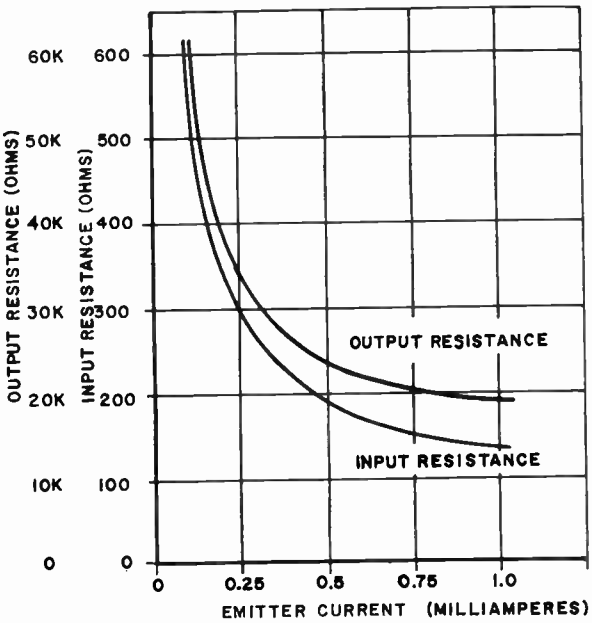


Fig. 7—Variation of input and output resistance with emitter current.

coupling-circuit loss and the ratio of unloaded coupling transformer Q to operating Q. It is shown, for example, that stability and coupling-circuit loss can be directly proportional.

In general, it can be stated that tuned amplifier interstage coupling-circuit design is determined by stability and selectivity requirements rather than by a need for impedance matching.

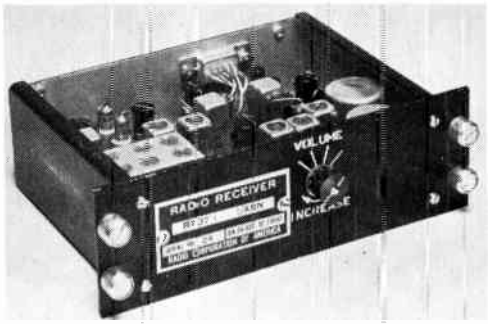
Automatic Gain Control

Automatic gain control of tuned-amplifier stages generally is accomplished by means of emitter current variation. While the gain of a transistor is relatively independent of emitter current, its input and output impedances are a strong function of emitter current, as shown in Fig. 7. Thus, gain reduction is predominantly produced by mismatch loss. The amount of gain control which can be obtained is determined by the difference between maximum usable gain and residual feedthrough when the transistor is biased to cutoff. A control range of 40 db or more can be readily obtained.

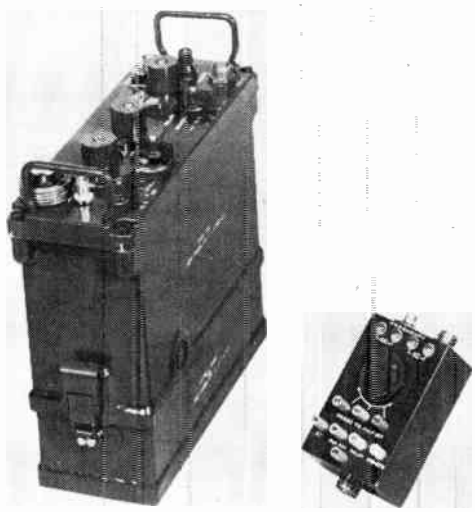
As mentioned in the discussion on cross modulation, it is sometimes desirable to bias a gain-controlled RF stage beyond cutoff under strong signal conditions. In order that this may be accomplished, a condition of amplified signal and fedthrough signal balancing must be avoided. This condition can arise at a low value of emitter current if the amplified and fedthrough signals are of the same magnitude and of opposite phase; a "sticking" of the emitter current at some small finite value can result.

Audio Amplifiers

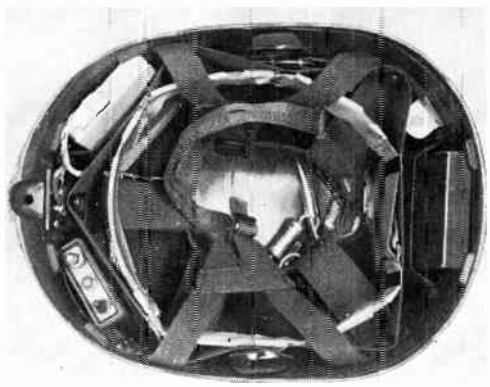
One of the most striking advantages of the transistor is its utility in Class-B audio amplifier circuits, and



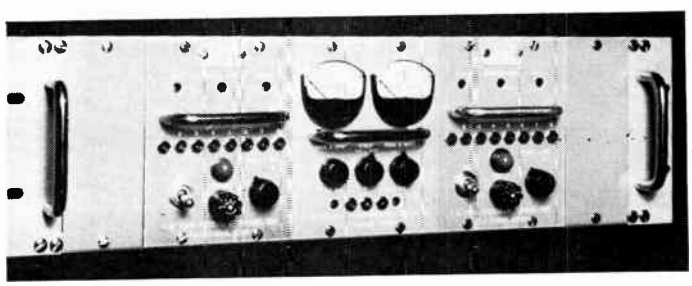
(a)



(b)



(c)



(d)

Fig. 8—Illustrative equipments. (a) 75-mc marker beacon receiver. (b) AN/PRC-25 transceiver (left) and projected Micro—Modular equivalent (right). (c) AN/PRC-34 helmet transceiver. (d) Frequency-shift tone receivers and diversity combining equipment.

most importantly in Class-B complementary symmetry circuits.[13] It is likely that the Class-B complementary symmetry circuit, which requires no audio transformers, will eventually make the audio transformer a component of the past, at least in receiver applications.

In using Class-B audio circuits, care must be taken to insure that output stage current flowing through the battery or other common impedance does not produce a voltage which is fed back to vulnerable portions of the receiver such as the local oscillator, gain-controlled stages, or early audio stages. A gain-controlled stage is particularly susceptible to small audio voltages, which may be fed back on the AGC bus or via a biasing network, when it is biased near cutoff under strong-signal conditions. Frequently, much can be accomplished in reduction of brute-force filtering by judicious return of emitter and base-bias bypasses.

An example of the use of a Class-B complementary symmetry circuit in a broadcast receiver can be found in the literature.[14]

Representative Hardware

Transistorized equipments which are indicative of the present state of the art are shown in Fig. 8 (p. 1259). A marker beacon receiver, Fig. 8(a), uses ten transistors in a crystal-controlled double-superheterodyne circuit. No shock mounting is required in installation. The AN/PRC-25, Fig. 8(b), is a man-pack FM transceiver providing two-way voice communication in the 30- to 76-mc frequency range. Through the use of a frequency synthesizer, this transceiver is crystal controlled on any one of 920 instantly available channels. Fig. 8(c) is a microminiature FM transceiver, AN/PRC-34, mounted in a combat infantryman's helmet. It operates in the frequency ranges of 38 to 39.2 mc and 47.8 to 51 mc. The receiver consists of an RF stage, crystal-controlled oscillator, mixer, two IF stages, and an AF amplifier. The transmitter includes AF circuits, variable-capacitance-diode modulator, and combined oscillator-power amplifier and tone generator.

An illustration of commercial point-to-point communications equipment is given in Fig. 8(d). This is a complete dual diversity system comprising two frequency-shift tone receivers and a dual-diversity comparator. A 1-db differential in the two-diversity channel amplitudes results in better than 30-db suppression of the weaker signal. Such transistorized equipment provides a high degree of reliability, and its small power dissipation is a welcome relief in large installations where problems of equipment-generated heat is severe.

[13] T. O. Stanley and T. M. Scott, "Design Considerations in Class-B Complementary Symmetry Circuits" in "Transistors I," Princeton, N. J.; 1956.
[14] D. D. Holmes, "A six-transistor portable receiver employing a complementary symmetry output stage," 1957 IRE National Convention Record, pt. 3, pp. 193–198.

Acknowledgment

The author is indebted to Anatole Minc, of the Tele-Signal Corp., Roslyn Harbor, N. Y., for information on diversity combining and related equipment.

Transistor Monostable Multivibrators for Pulse Generation*

J. J. SURAN†, SENIOR MEMBER, IRE

Summary—**This paper deals with the analysis and design of transformerless pulse generators which utilize clamped transistors as active nonlinear elements and passive RCL components as linear timing networks. RC and RCL monostable multivibrators are considered and approximate design procedures are developed for each of the basic types. Line-driving applications are discussed and experimental results are described.**

Circuit Considerations

THE problem of generating pulses of given amplitude and given width is a basic one in most digital systems. For example, flip flops may readily fail to operate if the driving pulse waveform is not constrained within proper limits. Pulse generators which develop pulses of precise width may also be used as digital delay networks and offer considerable advantages over delay lines, particularly when the delay interval is long.

A popular form of pulse generator is the blocking oscillator, which utilizes a single transistor and transformer in the basic configuration [1], [2]. In the blocking oscillator the transistor is used as an active switching element, *i.e.*, a switching device with power gain, and the energy storage properties of the transformer primarily determine the pulse waveform. Although the blocking oscillator is the most economical type of pulse

* Original manuscript received by the IRE, February 28, 1958.
† Electronics Lab., General Electric Co., Syracuse, N. Y.

generator from the point of view of the number of components used and although the transformer permits convenient impedance transformation between generator and load, several problems arise directly from the use of a transformer. For example, if a narrow pulse having a rise time equal to or less than 20 mμsec is required, the circuit bandwidth must be at least 25 mc wide to pass the required energy components. This requires a high-quality video transformer, since the flat top of the pulse must be more or less preserved, and transformers of this type are difficult to construct. A second problem arises from the fact that the energy stored in the transformer must be discharged at the termination of the pulse. If a low-resistance clamp circuit, *e.g.*, a diode, is used to discharge the transformer, a considerable time constant is introduced (L/R) and hence the circuit recovery time is long, thus limiting the pulse-repetition frequency. If the resistance level of the clamping network is increased, the discharge-time constant is reduced but "flyback" effects may become appreciable.

Some of the disadvantages inherent in blocking oscillators using "linear" transformers may be overcome by using square-loop ferrite cores. For example, the flyback problem is greatly reduced. However, driving square-loop cores at high-pulse repetition rates (*e.g.*, 500 kc) requires significant energy expenditure due to the losses encountered in traversing the B-H loop and hence pulse generators of this type lose much of their practicability as the frequency of operation is increased. In addition, the generation of narrow pulses requires very small cores and since these are difficult to wind with many turns, considerable current is required for driving purposes.

In view of these considerations, it seems judicious to investigate the "one-shot" multivibrator as a pulse source. The following advantages make the monostable multivibrator attractive:

1) The circuit is easily designed with readily available components, not requiring specially-designed transformers.

2) Variable pulse widths and variable pulse amplitudes are readily achieved by simple capacitor and resistor variations, respectively.

3) "Cleaner" pulse waveforms may be achieved due to the absence of flyback effects in the output circuit.

4) No special trigger circuits are required (input transformers are often used with blocking oscillators).

5) Multivibrator circuits may be designed to be "fast" in leading and trailing edge response whereas blocking oscillators usually require strong feedback and consequently lower loading capabilities to achieve high-speed switching (it should be pointed out here that the extra gain of the second mmv transistor makes this possible).

6) Higher-pulse repetition frequencies are possible due to the absence of transformer flyback effects.

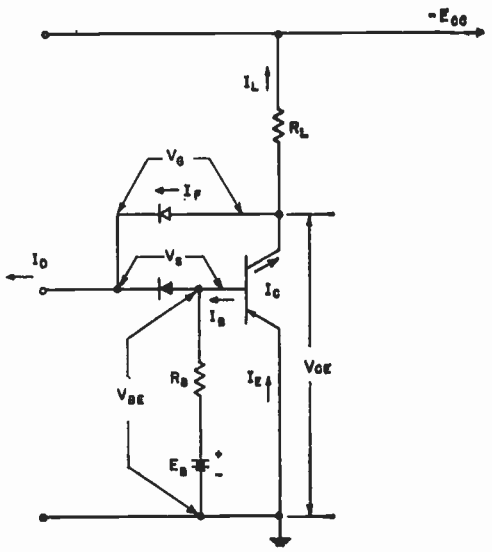


Fig. 1—Clamped transistor-switching stage.

Basic Switching Circuit

The basic transistor-switching element which is used in the circuits described below is the degeneratively clamped common-emitter stage illustrated in Fig. 1. It has been described sufficiently in the literature [1], [3], and hence any analysis here will be redundant. However, some of its basic advantages are summarized below.

1) Low-clamping voltage leads to high-power efficiency.

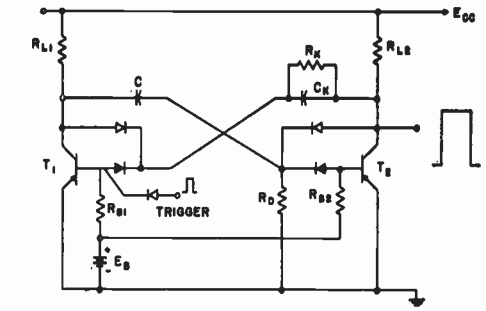
2) Degenerative feedback clamp minimizes the dissipation effects of high overdrive and makes possible a considerable loading margin without excessive collector and clamp-diode current ("current-demand" property).

3) Degeneration action reduces delay, or pulse "stretching" effect, to such a negligible amount that in most cases the turnoff delay is completely masked by the rise time of the driving signal.

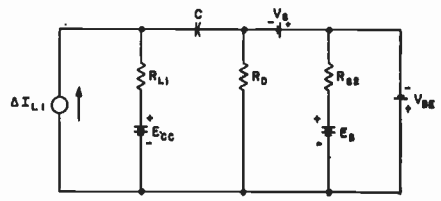
RC Monostable Multivibrator

An RC-type monostable multivibrator circuit is illustrated in Fig. 2(a). Under quiescent conditions T_1 is biased on and T_2 is held off. When a positive trigger pulse is applied to the base of T_1 the latter is turned off and the negative transient generated at the collector of T_1 forces T_2 into a conducting state. The regenerative action provided by the closed-loop coupling through $R_K - C_K$ keeps T_1 cut off even after the trigger pulse has disappeared, providing that T_2 is still conducting. Transistor T_2 will continue to conduct until capacitor C charges sufficiently to block the conduction current in the base of T_2. At that point T_2 begins to cut off, due to the action of E_B, and the circuit is carried back to its initial state in a regenerative manner.

The time duration of the regenerated pulse may be calculated approximately from the equivalent circuit illustrated in Fig. 2(b). The resulting equation is



(a)



(b)

Fig. 2—(a) Monostable multivibrator. (b) Equivalent circuit of mmv during pulse generation.

$$I_{L1}\exp(-t/R_{L1}C) = (I_{L2}/\alpha_{E\min}) + (E_B + V_{BE})/R_{B2} \quad (1)$$

where I_L is the load current of the conducting transistor, $\alpha_{E\min}$ is the minimum value of current gain for the transistor connected in the common-emitter configuration, and V_{BE} is the base-to-emitter drop of the conducting transistor. The right-hand side of (1) represents the *minimum* current flowing out of the driving point (junction between R_D and C) which is required to maintain T_2 in a clamped state. The left-hand side of (1) represents the current transient flowing out from the driving point due to T_1 shutting off in a step manner. Validity of (1) depends upon the following inequalities:

$$[R_D R_{B2}/(R_D + R_{B2})] \gg r_{B2}'$$
$$R_{L1} \gg r_{B2}'$$

where r_{B2}' is approximately the common-emitter input resistance of transistor T_2 (neglecting reactive effects). Under these conditions, the pulse duration T_D may readily be calculated from (1):

$$T_D = R_{L1}C \ln\left[\frac{\alpha_{E\min}}{(I_{L2}/I_{L1}) + \alpha_{E\min}(E_B + V_{BE})/R_{B2}I_{L1}}\right]. \quad (2)$$

It is now possible to formulate a design procedure for the circuit. As a starting point, it should be recognized that half of the circuit of Fig. 2(a), consisting of T_2, R_{L2}, R_K, C_K, and R_{B1}, is identical to that of a bistable flip flop. Hence, the design for determining these elements may be applied directly to symmetrical flip-flop configurations.

In the design procedure, it will be assumed that the supply voltages, E_{CC} and E_B, are given and that the desired collector voltage swing, ΔV_C, and the nominal load current, I_L, are specified or selected. In addition, R_{L1} is made equal to R_{L2} to insure a strong base drive into T_2. The problem is to find suitable values for R_L, R_{B1}, R_K, and C_K. For convenience, the following design parameters are defined:

$$Z = R_L/(R_L + R_K), \quad (3)$$
$$A = \Delta V_C/E_{CC}, \quad (4)$$
$$\sigma = \alpha_{E\min}(I_{C1} - I_{L1})/I_{L1}. \quad (5)$$

The difference in collector voltage between the conducting and nonconducting states of T_1 is readily shown to be

$$\Delta V_C = [R_K/(R_L + R_K)]E_{CC} + [R_L/(R_L + R_K)](V_S + V_{BE}) - V_{CE}$$

where V_{CE} is the clamped collector potential given by

$$V_{CE} = V_{BE} + V_S - V_G. \quad (6)$$

Consequently, the design parameter Z, given by (3), may be written in terms of the circuit specifications:

$$Z = \frac{1 - A - V_{CE}/E_{CC}}{1 - (V_S + V_{BE})/E_{CC}}. \quad (7)$$

Since V_{CE} and V_S are characteristics of the clamping diodes and transistors, it is seen that as soon as the ratio $|\Delta V_C/E_{CC}|$ is specified, the design parameter Z is determined. However, it can be shown that for temperature stabilization of the circuit under the worst conditions (which occurs when the maximum leakage current, I_{com}, of one transistor is much greater than the leakage current of the other), the following inequality should be satisfied:

$$Z \gg AI_{com}/I_L.$$

Thus, the particular set of initial specifications which have been chosen inherently governs the temperature stability of the circuit. If the temperature stability is not adequate, the initial specifications must be relaxed, *e.g.*, by decreasing A or increasing I_L. Furthermore, it can be shown from (7) that if Z is to be made fairly independent of tolerances associated with V_{CE}, V_S, and V_{BE}, the following inequality should be satisfied:

$$(V_{CE}/E_{CC}) < [0.5(1 - A)]$$

Consequently, the specification of A, given by (4), may have to be relaxed in order to obtain reasonable values of the design parameter Z.

The design parameter σ, as defined by (5), is referred to as the *loading factor* and is a measure of the degree by which the load current can be increased beyond the nominal design value before the current in the clamp diode is reduced to zero. The quantity $\alpha_{E\min}(I_{C1} - I_{L1})$ represents current which is *potentially* available to the load, R_{L1}, when that portion of the drive current in the clamping diode is transferred to the base of transistor T_1. In the Appendix, two important relationships involving σ are derived. The first shows that the maximum value

ich the loading factor can have is

$$\sigma_{max} = \alpha_{Emin}\left[\frac{E_{CC} - (V_S + V_{BE})}{E_{CC} - V_{CE}}\right]\left(\frac{Z}{1 - Z}\right). \quad (8)$$

ıe second relates the loading factor to the resistor

$$R_{B1} = \alpha_{Emin}R_L(E_B + V_{BE})/(\sigma_{max} - \sigma)(E_{CC} - V_{CE}). \quad (9)$$

ɔw it is apparent that the value of R_{B1} depends solely the designer's choice of the loading factor, σ, and at this choice is the first degree of freedom thus far countered since R_L and R_K are automatically deter-ined by the initial circuit specifications, *viz.*,

$$R_L = \Delta V_C/I_L, \quad (10)$$

$$R_K = (1 - Z)R_L/Z. \quad (11)$$

large value of R_{B1}, corresponding to the selection of a ading factor approaching σ_{max}, results in maximum rcuit stability but poor trigger sensitivity since the ırn-off drive contributed by E_B is very small. A small alue of R_{B1}, corresponding to the selection of a loading ctor approaching zero, results in maximum turn-off ·ive on T_1, during the unstable condition of the circuit, ut also results in poor circuit stability since the clamp-ıg action of T_1 becomes marginal. Hence, the choice of should be such that a reasonable value of R_{B1} is ob-ıined.

Steps 1–5 in Table I are predicated upon (9), (10), nd (11). The selection of C_K (step 6 in Table I) is based pon the requirement that the discharge time constant ssociated with C_K should in no way affect the pulse ·idth T_D, as given by (2). The design problem now is arrowed to finding suitable values for R_{B2}, R_D, and C. Constraints may be imposed upon R_{B2} by the require-ıent that the logarithmic argument of (2) should as-ume a reasonable value. If the argument is too close to ınity, the logarithmic term approaches zero; conse-ıuently, inordinately high values of C will be required to ɔbtain reasonable pulse widths. On the other hand, if he logarithmic term becomes too large, excessively mall values of C will result and insufficient turn-on drive ıt the base of T_2 may occur during the initiation of the ınstable state of the circuit. Thus it is desirable to con-ıtrain the value of the logarithmic argument in (2) to a ·ange not to exceed some upper limit, Q, nor to fall be-ow some minimum value, P. Experience has shown that P values corresponding to approximately 0.5 $R_{L1}C$(P = 1.7) and Q values corresponding to approximately 1.5 $R_{L1}C$(Q = 4.5) are reasonable. These values are used to derive the constraints imposed upon the selection of R_{B2} in step 7 of Table I.

In establishing a method for selecting R_D, it should be noted that the role of this resistor in the mmv circuit of Fig. 2(a) is simply to provide a return path to ground for the bias current from E_B. One criterion for selecting R_D is to make certain that the cutoff potential at the base of T_2, due to the division of E_B across R_D and R_{B2}, is not so large that the peak inverse base-to-emitter voltage rating of T_2 is exceeded. Another criterion for selecting R_D is to specify that C is at least 90 per cent discharged within the minimum expected period, T_{Pmin}, of the trigger pulses. The latter constraint forms the basis for selecting R_D in step 9 in Table I.

DESIGN TABLE I

Circuit Configuration: Fig. 2(a).
Given: E_{cc}, E_B, ΔI_{L2}, ΔV_{c2}, T_D, T_{pmin}
Find: R_{L1}, R_{L2}, R_{B1}, R_{B2}, R_D, R_k, C_k, C

Procedure:

1) Calculate $R_{L2} = \Delta V_{c2}/\Delta I_{L2}$
   Then select $R_{L1} = R_{L2}$
2) Calculate Z:

$$Z = \frac{1 - A - V_{CE}/E_{cc}}{1 - (V_s + V_{BE})/E_{cc}}\Bigg\} A = \Delta V_{c2}/E_{cc}$$

3) Calculate R_k: $R_k = (1 - Z)R_{L2}/Z$
4) Select σ:

$$0 < \sigma < \left[\frac{\alpha_{Emin}Z}{A}\left(1 - \frac{V_{BE} + V_s}{E_{cc}}\right) - 1\right]$$

5) Calculate R_{B1}:

$$R_{B1} = \frac{\alpha_{Emin}E_B}{\Delta I_{L2}\left[\frac{\alpha_{Emin}Z}{A}\left(1 - \frac{V_{BE} + V_s}{E_{cc}}\right) - 1 - \sigma\right]}$$

6) Calculate C_k:

$$C_k \leq \frac{(R_k + R_{L2} + R_{B1})(T_D)}{5R_k(R_{L2} + R_{B1})}$$

7) Select R_{B2}:*

$$\frac{4.5\alpha_{Emin}(E_B + V_{BE})}{\alpha_{Emin}\left(\frac{E_{cc} - V_{CE}}{R_{L1}}\right) - 4.5\left(\frac{E_{cc} - V_{CE}}{R_{L2}}\right)} \geq R_{B2} \geq \frac{1.7\alpha_{Emin}(E_B + V_{BE})}{\alpha_{Emin}\left(\frac{E_{cc} - V_{CE}}{R_{L1}}\right) - 1.7\left(\frac{E_{cc} - V_{CE}}{R_{L2}}\right)}$$

8) Calculate C:

$$C = \frac{T_D}{R_{L1}\ln\left[\frac{\alpha_{Emin}R_{B2}(E_{cc} - V_{CE})}{R_{B2}(E_{cc} - V_{CE}) + \alpha_{Emin}R_{L1}(E_B + V_{BE})}\right]}$$

9) Select R_D:* If $2.5\,R_{B2}C > (T_{pmin} - T_D)$

$$R_D \leq \frac{(T_{pmin} - T_D)R_{B2}}{2.5CR_{B2} - (T_{pmin} - T_D)}$$

* R_{B2}, $R_D \gg r_{b2}'$ for proper triggering action.

Example—Suppose that it is desired to generate a pulse of 25 ma, 8-volt amplitude and of 0.5-μsec duration at a maximum repetition frequency of 500 kc. The battery supplies E_{CC} and E_B are 10 and 5 volts, respectively, and $V_S \cong 0.5$ volt, $V_{BE} \cong 0.2$ volt, $V_{CE} \cong 0.5$ volt. Transistors having the following parameters are used:

$$\alpha_{Emin} = 20,\ \omega_{ab} = 100 \times 10^6 \text{ rad/sec},\ r_b' = 100 \text{ ohms}.$$

Following the procedure outlined in Design Table I:

1) $R_{L1} = R_{L2} = 8/25 \times 10^{-3} = 320$ ohms.

2) $$Z = \frac{1 - 0.8 - 0.05}{1 - 0.07} = 0.16.$$

3) $R_K = \left(\frac{0.84}{0.16}\right)(320) = 1.7\text{K ohms.}$

4) Select σ:

$$0 < \sigma < \left\{\left[\frac{(20)(0.16)}{0.8}\right](1 - 0.07) - 1\right\},$$

$0 < \sigma < 2.7$; for example, select $\sigma = 1$.

5) $R_{B1} = \dfrac{(20)(5)}{(25)(10^3)(1.7)} = 2.3\text{ K ohms.}$

6) $$C_k \leq \left[\frac{(1.7 + 0.32 + 2.3)(10^3)(0.5)(10^{-6})}{(5)(1.7)(0.32 + 2.3)(10^6)} = 100\ \mu\mu f\right],$$

Select $C_k = 100\ \mu\mu f$.

7) Select R_{B2}:

$$\frac{(4.5)(20)(5.2)}{(30 \times 10^{-3})(20 - 4.5)} \geq R_{B2} \geq \frac{(1.7)(20)(5.2)}{(30)(10^{-3})(20 - 1.7)}$$

$760 \geq R_{B2} \geq 320$ Select $R_{B2} = 750$ ohms

8) $$C = \frac{0.5 \times 10^{-6}}{320 \ln\left[\dfrac{(20)(750)(9.5)}{(750)(9.5) + (20)(320)(5.2)}\right]} = 1250\ \mu\mu f.$$

9) $R_D \leq 1300$ ohms.

Select $R_D = 750$ ohms.

The circuit designed above was tested experimentally using the closest standard component values to those calculated. The results which were obtained are illustrated in Fig. 3. Close agreement between theoretical and experimental data exists until the pulse width begins to approach the rise time of the output transistor. Since transistor reactive effects were neglected in the analysis, this divergence between theory and experiment is to be expected.

Waveform of the generated pulse is good for small pulse widths but as the pulse is increased, by increasing C, the trailing edge begins to show a pronounced rounding effect at the beginning of the fall transient. This anomaly, illustrated in Fig. 4, is due to the delay between the time the exponentially-decaying drive current falls to that level just required to sustain the clamps and the time the collector potential falls enough for the circuit to become regenerative. It is apparent from Fig. 4 that the trailing edge anomaly becomes worse as C is made larger. The same cause which leads to the trailing edge rounding, *viz.*, the decreasing derivative of the exponential form of the drive current, may also be responsible for considerable trailing edge jitter. For example, slight changes in the clamp or regeneration levels could cause great variations in the pulse width by shifting the point (in time) where the fall transient begins. Such changes may occur as the result of noise, transistor thermal effects, or bias-supply variations. In order to *reduce* the jitter effect, the logarithmic argument of (2) was constrained to a maximum value given by Q (which was chosen as 4.5), but it should be noted that this does not *eliminate* jitter. For the circuit illustrated in Fig. 3, trailing-edge jitter of the order of 3 per cent of the generated pulse width was observed.

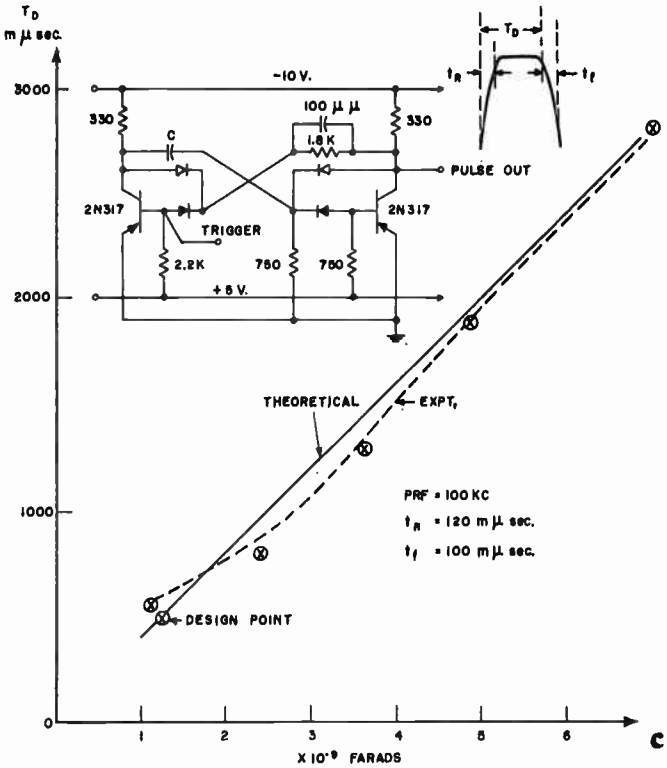


Fig. 3—Experimental vs theoretical results.

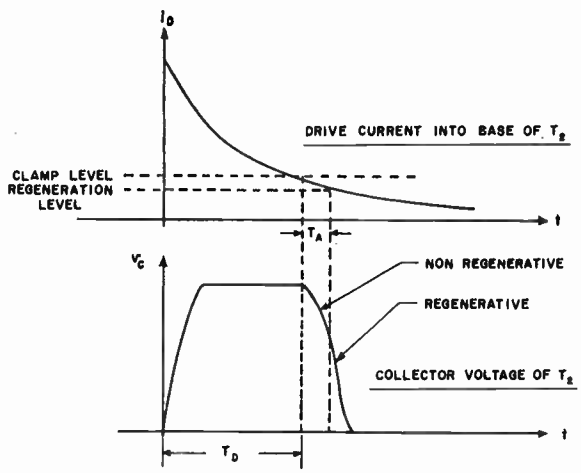
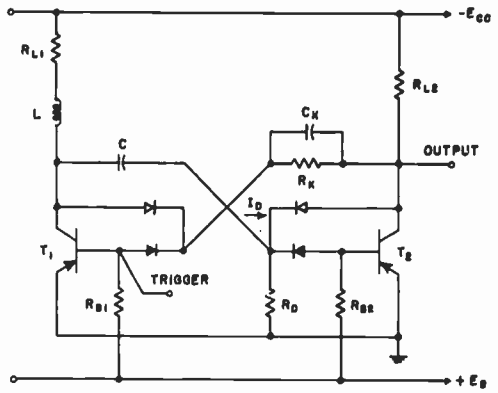


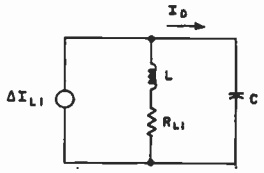
Fig. 4—Trailing edge anomaly.

RLC Monostable Multivibrator

Most of the deficiencies which may be attributed to simple RC pulse generators of the type described in the



(a)



(b)

Fig. 5—RLC regenerative pulse generator.

previous section are generally attributable to the exponential timing waveform. The suggestion is very strong, therefore, to use a different type of timing network. In order to obtain a substantial improvement, the new timing network should fulfill the following requirements.

1) Exhibit an *increasing* derivative as the circuit-recovery point is approached in order to minimize trailing-edge jitter.

2) Allow more efficient transistor-base drive, particularly in high-current and high-speed circuits, in order to decrease both rise and fall times of the generated pulse.

A circuit configuration which satisfies these requirements is illustrated in Fig. 5. The configuration is substantially the same as the RC pulse generator of Fig. 2(a), but includes an inductance element, L, in series with the collector-load resistance of T_1. In the RLC circuit shown, transistor T_1 is normally conducting and T_2 is normally cutoff. When a positive trigger pulse is applied to the base of T_1, the latter is driven off. Consequently, the collector potential of T_1 changes very rapidly and the current transient is directed through C toward the base of T_2 thus causing T_2 to turn on rapidly. Assuming that the parallel combination of R_D and R_{B2} is considerably greater in resistance than the input impedance of T_2 when the latter conducts, the waveform of the decaying drive current, I_D, will be governed by the resonant response of the $R_{L1}LC$ network. The approximate equivalent circuit during this interval is illustrated in Fig. 5(b). The current transform, $I_D/\Delta I_{L1}$, in terms of the Laplace operator s, is

$$\frac{I_D(s)}{\Delta I_{L1}(s)} = \frac{s + R_{L1}/L}{s^2 + (R_{L1}s/L) + 1/LC}. \tag{12}$$

Now if $\Delta I_{L1}(s)$ is assumed to be a step function, I_{L1}/s, the transient solution of (12) is

$$i_D(t) = I_{L1}\,[\exp\,(-R_{L1}t/2L)][\cos\,(\omega t + \theta)] \tag{13}$$

where

$$\omega = \sqrt{(1/LC) - (R_{L1}/2L)^2} \tag{14a}$$

$$\theta = -\tan^{-1}\,(R_{L1}/2L\omega). \tag{14b}$$

It is seen from (14a) that as long as $(1/LC)$ is larger than $(R_{L1}/2L)^2$, the waveform of the drive current will be a damped sinusoid. Furthermore, if θ is small, the derivative of the drive current will initially be very small but will increase as a sine function as i_D decreases.

In order to calculate the duration of the pulse generated at the collector of T_2, $i_D(t)$ is equated to the current value which is just necessary to sustain the clamps on T_2. Thus

$$I_{L1}\left[\exp\left(-\frac{R_{L1}t}{2L}\right)\right][\cos\,(\omega t + \theta)] = \frac{I_{L2}}{\alpha_{E\min}} + \frac{E_B + V_{BE}}{R_{B2}} \tag{15}$$

if

$$(1/LC) \gg (R_{L1}/2L)^2 \tag{16a}$$

and

$$(2L/R_{L1}) \gg T_D. \tag{16b}$$

Eq. (15) may be solved for the pulse duration T_D as follows:

$$T_D \cong \sqrt{LC}\,\cos^{-1} \cdot \left\{\frac{[R_{B2}(E_{cc} - V_{CE})/R_{L2}] + \alpha_{E\min}(E_B + V_{BE})}{\alpha_{E\min}R_{B2}I_{L1}}\right\}. \tag{17}$$

It is important at this point to consider carefully the conditions under which (17) is valid. Inequality (16a) effectively reduces the initial rate of change of $i_D(t)$, with respect to time, to a very small value and makes it possible to write (15) as

$$[\exp\,(-t/\tau)]\,\cos\,\omega t = H \tag{18}$$

where

$$\tau = 2L/R_{L1}$$

and

$$H = [(I_{L2}/\alpha_{E\min}) + (E_B + V_{BE})/R_{B2}]/I_{L1}.$$

However, the equation from which (17) was derived is

$$\cos\,\omega t = H. \tag{19}$$

It is apparent that the solutions for t of both (18) and (19) are identical at $\omega t = \pi/2$ and that the difference between the two solutions increases as ωt approaches zero. But one of the advantages which is supposed to be gained by using the RLC instead of the RC circuit is to decrease jitter at the trailing edge of the generated pulse, which should occur as the result of the increasing derivative of the cosine function. Consequently, the solutions of (18) and (19) should be constrained to the neighborhood of $\omega t = \pi/2$. Now if the solutions are arbitrarily constrained to the region

$$\pi/3 \leq \omega t < \pi/2 \tag{20}$$

then, from (19),

$$0 \leq H \leq 0.5. \tag{21}$$

By using (19) instead of (18) to calculate t, the error will always be such as to yield a greater theoretical value of t, *i.e.*, from (19)

$$\omega t_1 = \cos^{-1}(H)$$

and from (18)

$$\omega t_2 = \cos^{-1}[H \exp(t_2/\tau)].$$

From inequality (20) the *maximum* value t_2 can have is

$$t_2 = \pi/2\omega.$$

Hence, the *maximum error* which can result by using (19) instead of (18) to calculate t is

$$\Delta t_{\max} = t_1 - t_2 = (1/\omega)\{\cos^{-1} H - \cos^{-1}[H \exp(\pi/2\omega t)]\}.$$

Defining the per cent error, δ, as

$$\delta = (t_1 - t_2)/t_1,$$

it is seen that

$$\delta = \{1 - \cos^{-1}[H \exp(\pi/2\omega t)]/\cos^{-1} H\}. \tag{22}$$

Solving (22) for ωt results in:

$$\omega t = \pi/2 \ln\left\{\frac{1}{H}\cos[(1-\delta)\cos^{-1} H]\right\}. \tag{23}$$

If inequality (16a) is valid, then $\omega \cong 1/\sqrt{LC}$ and (23) may be written as

$$\sqrt{\frac{L}{C}} = \frac{\pi R_{L1}}{4 \ln\left\{\frac{1}{H}\cos[(1-\delta)\cos^{-1} H]\right\}}. \tag{24}$$

Thus it is seen that if the error is to be kept below a specified amount, δ, the ratio of $\sqrt{L/C}$ must be equal to or greater than the right-hand side of (24). Eq. (24) is plotted in Fig. 6 for $H = 0.5$. This curve represents the *maximum error* which can be made by using the approximation of (17), instead of the transcendental form given by (15), if the constraints imposed by inequalities (16a) and (21) are applied, for a given $\sqrt{L/C}/R_{L1}$ ratio. Hence, if the maximum approximation error is to be less than 50 per cent, the ratio $\sqrt{L/C}/R_{L1}$ should be equal to or greater than 1.4.

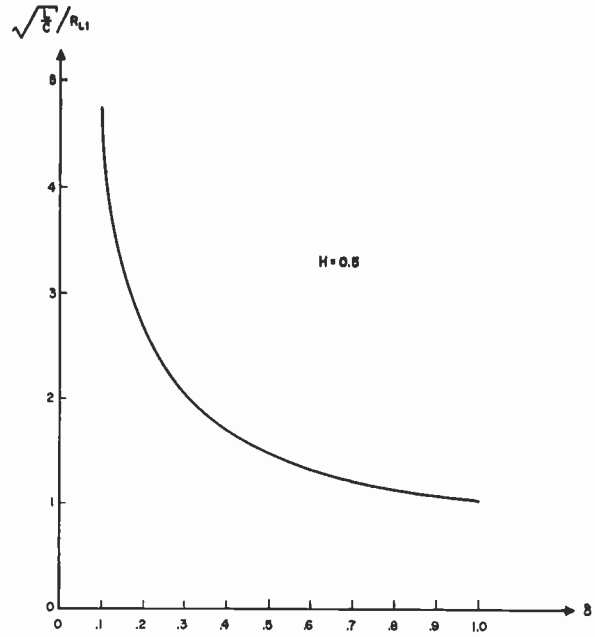


Fig. 6—Maximum error curve for RLC-circuit design.

A design procedure for the circuit of Fig. 5 may now be established. The criteria for selecting or calculating R_{L2}, R_k, R_{B1}, and C_k are the same as for the RC circuit and hence steps 1–7 in Table II follow the procedure derived in the previous section. Resistor R_{B2} may be selected within the constraints given by inequality (21). However, it is not practical to use 0 as the lower limit of H since this implies no turn-off action on T_2. Consequently, as an example, the lower limit of H will be selected arbitrarily as 0.2 (corresponding to $\omega t = 79°$). Since H is equal to the right-hand side of (15), the following constraint conditions are established.

$$0.2\, I_{L1} \leq \left[\frac{E_{cc} - V_{CE}}{\alpha_{E\min} R_{L2}} + \frac{E_B + V_{BE}}{R_{B2}}\right] \leq 0.5\, I_{L1}. \tag{25}$$

Since $I_{L1} = (E_{cc} - V_{CE})/R_{L1}$, step 8 in Table II may be derived from (25).

The next step in the design procedure is to select L within the proper constraints. For validity of the analysis, the most important constraint is given by (16a). Thus

$$L \gg CR_{L1}^2/4. \tag{26}$$

Now from the error curve of Fig. 6, it is apparent that if the approximation given by (17) is to be valid to within a maximum error of +50 per cent,

$$\sqrt{L/C} \geq 1.4\, R_{L1}$$

DESIGN TABLE II

Circuit Configuration: Fig. 5
Given: E_{cc}, E_B, ΔI_{L2}, ΔV_{c2}, T_D, T_{pmin}
Find: R_{L1}, R_{L2}, R_{B1}, R_{B2}, R_D, R_k, C_k, C, L

Procedure:

1) Calculate $R_{L2}=\Delta V_{c2}/\Delta I_{L2}$
2) Select $R_{L1}=R_{L2}$
3) Calculate Z:

$$Z=\frac{1-A-V_{CE}/E_{cc}}{1-(V_s+V_{BE})/E_{cc}}\Bigg\}A=\Delta V_{c2}/E_{cc}$$

4) Calculate R_k: $R_k=(1-Z)R_{L2}/Z$
5) Select σ:

$$0<\sigma<\left[\frac{\alpha_{E\min}Z}{A}\left(1-\frac{V_{BE}+V_s}{E_{cc}}\right)-1\right]$$

6) Calculate R_{B1}:

$$R_{B1}=\frac{\alpha_{E\min}E_B}{\Delta I_{L2}\left[\frac{\alpha_{E\min}Z}{A}\left(1-\frac{V_{BE}+V_s}{E_{cc}}\right)-1-\sigma\right]}$$

7) Calculate C_k:*

$$C_k\le\frac{(R_k+R_{L2}+R_{B1})(T_D)}{5R_k(R_{L2}+R_{B1})}$$

8) Select R_{B2}:

$$\frac{\alpha_{E\min}(E_B+V_{BE})R_{L1}}{(E_{cc}-V_{CE})(0.2\alpha_{E\min}-1)}>R_{B2}\ge\frac{\alpha_{E\min}(E_B+V_{BE})R_{L1}}{(E_{cc}-V_{CE})(0.5\alpha_{E\min}-1)}$$

9) Select L:

$$\frac{(T_{pmin}-T_D)R_{L1}}{2}\ge L\ge\frac{1.5R_{L1}T_D}{\cos^{-1}\left[\frac{E_{cc}-V_{CE}+\alpha_{E\min}(E_B+V_{BE})R_{L1}/R_{B2}}{\alpha_{E\min}(E_{cc}-V_{CE})}\right]}$$

10) Calculate C:

$$C=\frac{1}{L}\left\{\frac{T_D}{\cos^{-1}\left[\frac{E_{cc}-V_{CE}+\alpha_{E\min}(E_B+V_{BE})R_{L1}/R_{B2}}{\alpha_{E\min}(E_{cc}-V_{CE})}\right]}\right\}^2$$

11) Select R_D:† If $2.5R_{B2}C>(T_{pmin}-T_D)$

$$R_D\le\frac{(T_{pmin}-T_D)R_{B2}}{2.5CR_{B2}-(T_{pmin}-T_D)}$$

* $(T_D/5)$ is selected as the period in order to insure that the discharge-time constant associated with C_k is much less than the required pulse duration.

† R_D in parallel with R_{B2} should be greater than r_{b2}' of transitor T_2 for proper switching action.

or

$$L\ge 2CR_{L1}^2. \tag{27}$$

If (27) is satisfied, inequality (26) is satisfied by a factor of at least 8 and hence constraint (27) precludes (26). Solving (17) for C and substituting into (27) gives

$$L\ge\frac{\sqrt{2}R_{L1}T_D}{\cos^{-1}\left[\frac{E_{cc}-V_{CE}+\alpha_{E\min}(E_B+V_{BE})R_{L1}/R_{B2}}{\alpha_{E\min}(E_{cc}-V_{CE})}\right]}. \tag{28}$$

A lower limit on the value of L may be derived by remembering that the circuit must recover to its initial condition between trigger pulses. Hence, when transistor T_1 switches into a conduction state, T_D seconds after the initial trigger pulse, full current must be established in R_{L1} and L *before* the arrival of the second trigger pulse or else the width of the second pulse will be less than the desired amount. When transistor T_1 is conducting, the time constant associated with the collector-load impedance is approximately L/R_{L1}. The time during which full current must be established in the collector circuit is $T_{pmin}-T_D$, where T_{pmin} is the minimum period between trigger pulses. Consequently, specifying at least 90 per cent recovery of the collector current between trigger pulses leads to the constraint:

$$2.3L/R_{L1}\le(T_{pmin}-T_D). \tag{29}$$

The conditions given by (28) and (29) serve as the basis for selecting L in step 9 of Table II.

Once R_{B2} and L have been selected, subject to the constraints described above, the required value of C may be calculated directly from (17). (See step 10 in Table II.) The design is then completed by selecting a suitable value for R_D. The criterion for the last step is the same as that used in Table I.

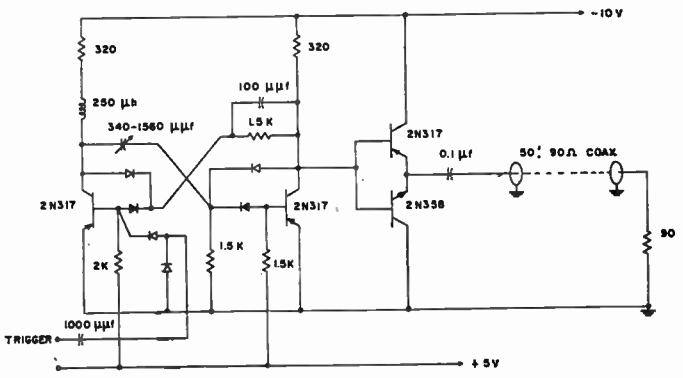


Fig. 7—Pulse generator for line driving application.

As an example of the use of Table II, consider the problem of the design of a pulse generator for driving long coaxial cables. A complementary-transistor emitter-follower stage will be used, as illustrated in Fig. 7, to buffer the pulse generator from the cable.

Example—Suppose that it is desired to drive 50 feet of 90-ohm coaxial cable with a pulse 8 volts in amplitude, 0.5 μsec wide, and recurring at a PRF of 500 kc. The cable is terminated in a 90-ohm resistive load, and battery supplies of −10 and +5 volts are available. In order to keep the power dissipation in the output transistors low, capacitor coupling between the driver stage and the cable will be used. (See Fig. 7.)

The first step in the design procedure is to determine the size of the output coupling capacitor which is required. If it is assumed that 0.5-volt (6 per cent) "droop" in the top of the output pulse is tolerable, the coupling capacitor should have a value equal to or greater than the total charge which will flow during the pulse duration, divided by the 0.5-voltage buildup which may be

allowed across the capacitor. The total charge flow is (8 volts)/(90 ohms) multiplied by 0.5 μsec, or approximately 45×10^{-9} coloumbs. Hence,

$$C_0 \geq 45 \times 10^{-9} \text{ coloumbs}/0.5 \text{ v, or}$$

$$C_0 \geq 0.09\ \mu\text{f}.$$

Thus, a coupling capacitor of 0.1 μf is selected. With the output circuit determined, the next step is to find a suitable driving level for the pulse-generator circuit. For maximum driving speed, it is desirable that $R_{L2} \ll \alpha_{E\min} R_{LO}$. If a minimum α_E of 25 is assumed (this is the minimum value at 90 ma), it is apparent that $R_{L2} \ll 2250$ ohms. A value of $R_{L2} = 320$ ohms is satisfactory. From this point, Table II is followed directly. Thus

1) $I_{L2} = (8 \text{ volts})/320 = 25$ ma.
2) $R_{L1} = R_{L2} = 320$ ohms.
3) $Z = (1 - 0.8 - 0.5)/(1 - 0.7) = 0.16$.
4) $R_k = (0.84)(320)/(0.16) = 1.7$ K ohms.
5) Select $\sigma = 1$ (limits $0 < \sigma < 3.4$).
6) $R_{B1} = (25)(5)/(25)(10^{-3})(2.4) \cong 2.0$ K ohms.
7) $C_k = (3.8)(10^3)(0.5)(10^{-6})/(5)(1.5)(10^3)(2.3)(10^3) \cong 100\ \mu\mu f$.
8) Select R_{B2}:

$$\left[\frac{(25)(5.2)(320)}{(9.5)(4)} = 1200\right]$$

$$\geq R_{B2} \geq \left[\frac{(25)(5.2)(320)}{(11.5)(9)} = 450\right]$$

Select $R_{B2} = 1200$ ohms.

9) Select L:

$$\left[\frac{(10^{-6})(2 - 0.5)(320)}{2} = 240\ \mu h\right]$$

$$\geq L \geq \left[\frac{(1.5)(320)(0.5)(10^{-6})}{\cos^{-1}\left(\frac{9.5 + 35}{240}\right)} = 170\ \mu h\right]$$

Thus, select $L = 240\ \mu h$.

10) $$C = \frac{1}{240 \times 10^{-6}}\left[\frac{0.5 \times 10^{-6}}{\cos^{-1}(44.5/240)}\right]^2$$

$$= 540\ \mu\mu\text{f}.$$

11) Select R_D: $R_D \leq 15$ K ohms. Select $R_D = 1500$ ohms.

The circuit design is now complete. An estimate of the maximum error which may be expected in the calculation of pulse width, due to approximations made in the analysis, may be determined from the error curve of Fig. 6. Since $\sqrt{L/C}/R_L = 2.1$, it is seen that the maximum expected error is in the order of 30 per cent, that is, the actual pulse width may be 30 per cent less than the calculated width due to the simplifications which were made in the derivation of (17).

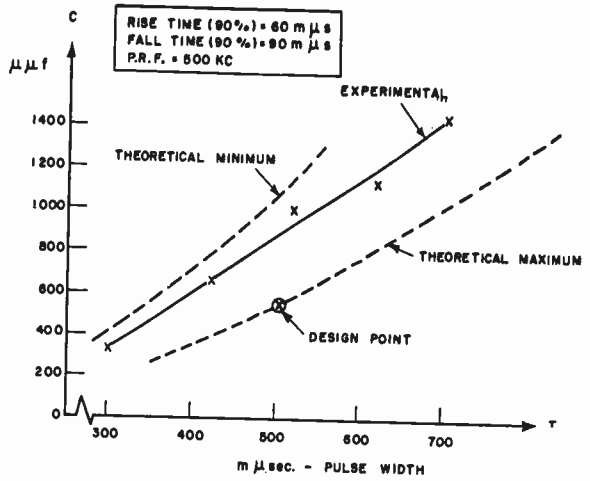


Fig. 8—Experimental vs theoretical results for circuits of Fig. 7.

A comparison of the theoretical and experimental pulse widths generated by the circuit of Fig. 7, when the cable is disconnected, is illustrated in Fig. 8. Reasonably close agreement between the calculated and measured results is indicated. The pulse width, as measured from the start of the leading edge to the start of the trailing edge, was found to be relatively unaffected by the cable load. A photograph of the generated pulse, at the input and output of the cable, for various pulse widths, adjusted by C, is illustrated in Fig. 9.

An estimate of the power dissipated in the driver transistors is difficult due to the complex transient conditions which exist. During the pulse duration, the *n-p-n* transistor is supplying roughly 100 ma with a base-collector voltage of approximately 0.5 volt, resulting in a peak-power dissipation of the order of 50 mw. During the switching transient the dissipation may be as high as 350 mw (8 volts × 100 ma/2). However, averaged out over the period of 2 μsec, the average dissipation is only on the order of 30 mw per transistor. Other performance characteristics of the circuit of Fig. 7 are:

Peak-pulse power efficiency = 65 per cent.
Average power efficiency (500 kc) = 35 per cent.
Power gain (peak-pulse) = 15 db.

Peak-pulse power efficiency is defined as the peak-pulse power delivered to the load divided by this power plus the total dissipation in the rest of the circuit during the pulse duration. The average power efficiency varies as the duty cycle and is defined as the average power delivered to the load divided by this power plus the average dissipation in the rest of the circuit. Peak-pulse power gain is defined as the peak pulse power delivered to the load divided by the peak trigger power.

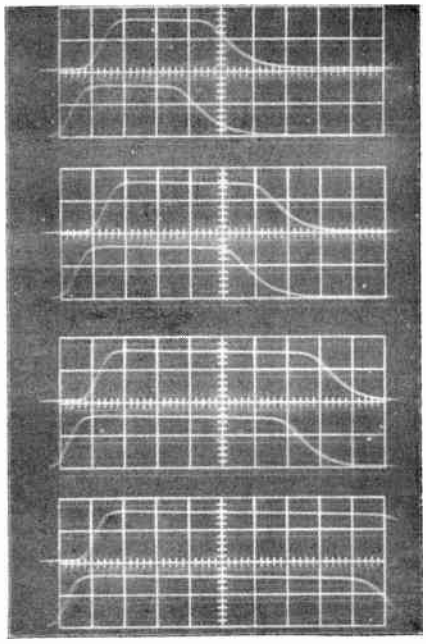


Fig. 9—Pulse waveforms for circuit of Fig. 7. Top picture in each set is pulse across 90-ohm load at end of 50 feet of cable and lower picture in each set is pulse at the collector of T_2 Sweep = 100 mμsec/cm, voltage scale 5 = v/cm, PRF = 500 kc. Each set illustrates pulse of different width as obtained by varying circuit capacitor C.

Transistor Selection

The selection of transistors for application to monostable multivibrator-pulse generators of the type discussed in the preceding sections is usually based upon considerations relating to maximum power, current, voltage, and thermal ratings and to gain-bandwidth capabilities. Maximum current and voltage ratings determine the driving capabilities of the circuit; maximum power and thermal ratings determine the allowable ambient conditions for given driving requirements; and the gain-bandwidth capabilities of the transistors determine the maximum repetition rate and minimum pulse width of the output waveform.

In considering the peak inverse rating of collector-to-emitter voltage when applications of the RLC design of Fig. 5 are made, it should be remembered that inverse voltages in excess of E_{cc} may be encountered between collector and emitter of T_1 during the generation of the pulse due to the underdamped response of the tuned circuit. Although collector breakdown under these conditions is not necessarily deleterious to the transistor if its power dissipation rating is not exceeded, collector breakdown may result in trailing-edge jitter and negate a major benefit of the LC circuit effect. However, voltage overshoot at the collector of T_1 is controllable and may in fact be eliminated without compromising the "clean" trailing edge of the generated pulse. For example, if (13) is approximated by

$$i_D(t) \approx (E_{cc}/R_{L1}) \cos \omega t \tag{30}$$

the voltage at the collector of T_1 is given by

$$v_{c1} \approx (1/C) \int_0^t i_D(t)\mathrm{d}t \cong (E_{cc}/\omega R_{L1}C) \sin \omega t. \tag{31}$$

Furthermore, since

$$\omega \approx \sqrt{1/LC}$$

the maximum voltage swing at the collector of T_1 is

$$v_{c1\max} \approx \frac{\sqrt{L/C}}{R_{L1}} E_{cc}. \tag{32}$$

Consequently, it is apparent from (32) that the ratio $\sqrt{L/C}/R_{L1}$ determines the overshoot of the voltage at the collector of T_1. For example, if the minimum condition for the inequality of (27) is satisfied, the maximum collector-voltage overshoot will be 40 per cent of the collector-supply voltage. Hence, if $E_{cc} = 10$ volts and $\sqrt{L/C}/R_{L1} = 1.4$, the peak-collector voltage of T_1 during pulse generation will be 14 volts. Furthermore, if design accuracy is sacrificed somewhat by making $\sqrt{L/C}/R_{L1} = 1$, no collector-voltage overshoot will be encountered and the resolution of the trailing edge will not be affected since the RLC circuit is still underdamped.

The gain-bandwidth requirement of the transistors is determined primarily by the desired width, T_D, of the generated pulse. Since internal reactive effects in the transistor were neglected in the analysis, which is a practical assumption if individual transistor circuit adjustments are to be avoided, the natural rise time of the transistor should not exceed approximately 25 per cent of the generated pulse width. The bandwidth, f_B, of a linear system compatible with a rise time, t_r, is given approximately by $BW \approx 1/2t_r$. If $t_r \leq 0.25\ T_D$, the required transistor bandwidth is at least $4/2T_D$. However, if the gain of the transistor is to be equal to at least two in order to make up for losses in the circuit, the required gain-bandwidth product of the transistor should be approximately $4/T_D$. For example, if a pulse width of 250 mμsec is desired, a transistor having a gain bandwidth of at least 16 mc is required. In diffusion-type transistors the gain-bandwidth product corresponds approximately to the α_B-cutoff frequency of the transistor. Hence, a "rule of thumb" which may be used for the selection of transistors having adequate frequency response to generate a pulse T_D seconds wide is:

$$f_{\alpha B} \geq 4/T_D. \tag{33}$$

In applying (33), however, it should be remembered that many modern high-frequency transistors rely strongly on drift effects for their high-frequency characteristics and that the equating of α_B-cutoff frequency with common-emitter current gain-bandwidth product

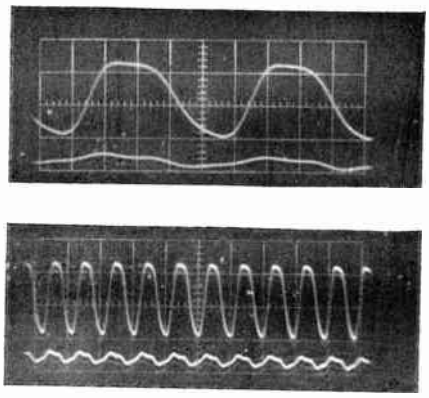


Fig. 10—Pulse waveform for 10 mc mmv. Top: Generated pulse and trigger signal, 20 mμsec/cm horizontal, 5v/cm vertical. Bottom: Generated pulse and trigger signal, 100 mμsec/cm horizontal, 5 v/cm vertical.

may lead to gross overestimates of the transistor's switching-speed capability. For example, laboratory measurements of a drift transistor nominally rated to have an α_B-cutoff of 100 mc revealed a peak gain-bandwidth product of only 40 mc over an operating-point switching locus of 2–15 volts (collector voltage) and 0–10 ma (emitter current). Experience has shown that gain-bandwidth is the pervasive factor governing speed and hence inequality (33) should more generally be interpreted as

$$\text{G-BW} \geq 4/T_D \tag{34}$$

Inequality (34) was tested in the design of high-speed pulse generators using the configuration of Fig. 5. In general, it was found to be somewhat conservative but the experimental results clearly showed that gain bandwidth rather than α_B-cutoff frequency governed the speed at which the circuits could operate. Table III summarizes these results.

TABLE III

Transistor Type	Nominal $f_{\alpha B}$	Measured G-BW	T_D	Max PRF
1	100 mc	35 mc	80 mμsec	5 mc
2	50 mc	40 mc	80 mμsec	6 mc
3	—	110 mc	30 mμsec	12 mc

Fig. 10 illustrates transistor Type 3 operating in a monostable multivibrator configuration at a pulse-repetition frequency of 10 mc and generating a pulse waveform approximately 40 mμsec wide.

Conclusions

Transistor-pulse generators using RC or RLC timing circuits may be designed with relatively straight-forward analytical techniques. Some of the advantages and disadvantages of the various circuits treated in this report are given below.

RC-type pulse generating circuits may be designed to give very narrow pulses (*e.g.*, 200 mμsec wide) but exhibit considerable jitter at the trailing edge of the pulse due to the decreasing derivative of the exponential timing waveform.

RCL-type pulse generating circuits may be designed with little additional complexity to give superior performance to the RC circuits and with significantly less jitter at the trailing edge of the pulse.

Line-driving circuits consisting of complementary transistor emitter-follower stages may be directly coupled to the pulse generating circuits without significantly affecting the pulse waveform.

Monostable multivibrator-pulse generators have the following advantages over blocking-oscillator types:

1) eliminate critical video transformers in high-speed circuits;
2) require very simple trigger circuits;
3) eliminate transformer flyback effects;
4) easily designed with readily-available components;
5) variable pulse amplitudes and widths may readily be obtained without affecting the pulse waveform.

Appendix

Derivation of Loading-Factor Relationships

In deriving the significant relations involving the loading factor, σ, reference is made to the normally conducting transistor in the mmv circuits illustrated in Fig. 2(a) or 5(a). The current conventions shown in Fig. 1 are used in the ensuing analysis; thus

$$I_F = I_c - I_L \tag{35}$$

where

$$I_c = \alpha_E I_B \tag{36}$$

$$I_L = (E_{cc} - V_{CE})/R_L. \tag{37}$$

In addition,

$$I_F = I_D - (I_B + I_b) \tag{38}$$

where

$$I_D = [E_{cc} - (V_S + V_{BE})]/(R_L + R_K) \tag{39}$$

$$I_b = (E_B + V_{BE})/R_B. \tag{40}$$

Eliminating I_F in (35) and (38),

$$I_c - I_L = I_D - (I_b + I_c/\alpha_E).$$

Hence

$$I_c = \alpha_B(I_D + I_L - I_b) \tag{41}$$

since

$$\alpha_B = \alpha_E/(1 + \alpha_E).$$

The loading factor is defined as

$$\sigma = \alpha_E(I_c - I_L)/I_L. \tag{42}$$

Now if $\alpha_B \approx 1.0$, substituting (41) into (42) results in

$$\sigma \cong \alpha_E(I_D - I_b)/I_L. \tag{43}$$

Consequently, from (37), (39), and (40):

$$\approx \alpha_E \frac{[E_{cc}-V_S-V_{BE})/(R_k+R_L)]-(E_B+V_{BE})/R_B}{(E_{cc}-V_{CE})/R_L}. \quad (44)$$

olving (44) for R_B results in

$$R_B = \frac{E_B + V_{BE}}{[(E_{cc} - V_S - V_{BE})/(R_k + R_L)] - \sigma(E_{cc} - V_{CE})/\alpha_E R_L}. \quad (45)$$

If physically realizable values of R_B are to be obained, it is apparent from (45) that the maximum alue of the loading factor is

$$\sigma_{max} = \left(\frac{\alpha_E R_L}{R_k + R_L}\right)\left[\frac{E_{cc} - (V_S + V_{BE})}{E_{cc} - V_{CE}}\right]. \quad (46)$$

Substituting (46) into (45) results in the form given by 9).

ACKNOWLEDGEMENT

The author wishes to thank B. K. Eriksen and J. Schmidt for their valuable assistance.

BIBLIOGRAPHY

[1] Shea, R. F., *et al.*, *Transistor Circuit Engineering*, New York: John Wiley and Sons, Inc., 1957.
[2] Linvill, J. G., and Mattson, R. H. "Junction Transistor Blocking Oscillators," PROCEEDINGS OF THE IRE, vol. 43 (November, 1955), pp. 1632–1639.
[3] Baker, R. H. *Maximum Efficiency Switching Circuits.* Cambridge: Massachusetts Institute of Technology, Lincoln Laboratory, Report No. TR-110, 1956.
[4] Linvill, J. G. "Nonsaturating Pulse Circuits Using Two Junction Transistors," PROCEEDINGS OF THE IRE, vol. 43 (July, 1955), pp. 1826–1834.

A Design Basis for Junction Transistor Oscillator Circuits*

D. F. PAGE†, ASSOCIATE MEMBER, IRE

Summary—**A unified approach to the design of junction transistor oscillators is presented, based on a study of the dynamic admittance presented to the load at the oscillation frequency by a bilateral active four-pole network with passive feedback. The method, applicable to practical oscillator arrangements, leads to simple design expressions when appropriately simplified transistor parameters are used. The equivalent circuit used is suitable for transistors operating by drift and/or diffusion of minority carriers in the base region.**

Oscillation to the highest possible frequencies, or with the highest possible load, is achieved with optimized feedback networks; for practical transistor oscillators, these are found to be easily designable in terms of the transistor parameters h_{11} and α. An expression for the maximum frequency of oscillation, f_m, valid for transistors operating by drift and/or diffusion, leads to new methods for the measurement of certain high-frequency transistor parameters.

INTRODUCTION

THE basic sinusoidal oscillator circuit of Fig. 1 is considered, where Tr is the transistor, F is a passive feedback network providing impedance transformation and phase shift, and G_L is a load consuming the output power $i_L v_L$.

Because of its wide use in practical oscillator circuits, open-circuit instability, as indicated in Fig. 1, is assumed for this presentation. In this form of instability the feedback is proportional to the load voltage and the load can be shunted with a parallel-resonant circuit. The following four transistor configurations are therefore of interest,

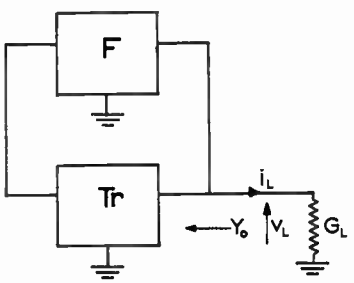


Fig. 1—Basic circuit diagram for open-circuit unstable transistor oscillator.

A: e-c—common-base, emitter input
B: b-c—common-emitter, base input
C: b-e—common-collector, base input
D: e-b—common-collector, emitter input. (1)

The two remaining possible configurations are useful for short-circuit unstable circuits in which the feedback is proportional to the load current, and where a series-resonant circuit can be put in series with the load. The various possible feedback arrangements have been dis-

* Original mansucript received by the IRE, February 27, 1958; revised manuscript received, April 18, 1958. The research program reported in this paper was carried out under the support of the Dept. of Scientific and Industrial Research of the United Kingdom.
† Elec. Eng. Dept., Imperial College, London, England.

cussed;[1] it is sufficient to point out that the design techniques to be presented here for open-circuit instability are equally applicable, in dual form, to the design of short-circuit instability.

If the conductive and susceptive components of Y_0, the oscillator admittance presented to the load, are written as

$$Y_0 = G_0 + jB_0, \tag{2}$$

continuous oscillations exist if

$$G_0 + G_L < 0 \tag{3}$$

at a frequency, for practical sinusoidal oscillators,[2] such that

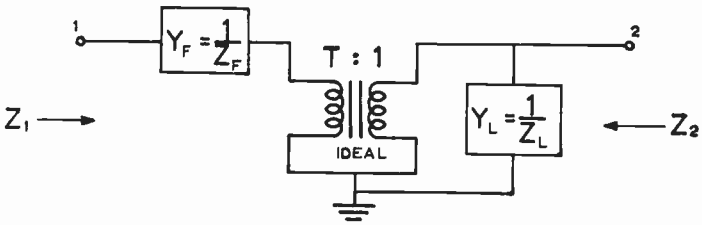
$$B_0 \doteq 0. \tag{4}$$

In general, F is a four-terminal, passive network, providing an impedance transformation and phase shift from output to input. Such a network, at a single frequency and in the steady state, can be most simply represented by the equivalent forms of Fig. 2.[1] The transformer-feedback form of Fig. 2(a), used in many practical oscillators[3] including the Hartley circuit, is the basic form for which the analysis in this paper is carried out. The pi impedance-coupled feedback form of Fig. 2(b), simply related to the transformer form as shown, is also widely used: for example, in the Colpitts circuit. It is important to note, however, that when a design for F is obtained in the forms of Fig. 2, other equivalent feedback networks can be calculated readily. For example, equivalent transmission line feedback (distributed or lumped) can be realized in terms of the characteristic impedances Z_1 and Z_2, and the image transfer function θ_{21}.

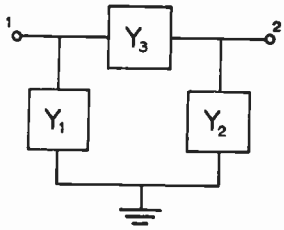
$$\left.\begin{aligned} Z_1 &= \sqrt{Z_F(Z_F + T^2 Z_L)} \\ Z_2 &= \sqrt{\frac{Z_L^2 Z_F}{Z_F + T^2 Z_L}} \\ \theta_{21} &= \tanh^{-1}\sqrt{\frac{Z_F}{Z_F + T^2 Z_L}} \end{aligned}\right\}. \tag{5}$$

If the equivalent transmission line is of the form of a lumped ladder-type resistance-capacitance network, the familiar phase-shift oscillator circuits result.

The difference between these various feedback forms arises when the transient build-up characteristics of the oscillation are considered. If nonlinearities in the transistor are negligible, the oscillation amplitude increases exponentially with time, at a rate governed by two

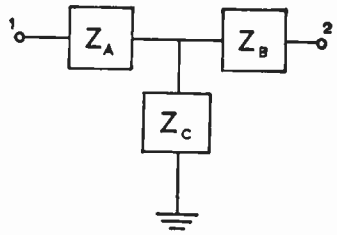


(a) Transformer-coupled form



(b) Pi impedance-coupled form

$Y_1 = Y_F(1 - T)$
$Y_2 = Y_L - TY_F(1 - T)$
$Y_3 = TY_F$



(c) Tee impedance-coupled form

$Z_A = Z_F - TZ_L(1 - T)$
$Z_B = Z_L(1 - T)$
$Z_C = TZ_L$

Fig. 2—Simplest circuit forms for F providing arbitrary impedance transformation and phase shift.

factors: 1) the negative value of (G_0+G_L), and 2) the energy storage in Tr and F at the oscillation frequency; the second factor, governed by the over-all distribution of reactances, very much depends on the network form chosen for F. As the oscillation amplitude becomes large, however, transistor nonlinearities reduce the dynamic value of $|G_0|$, averaged over one cycle; oscillations limit at an amplitude such that $|G_0|$ is reduced to the value of the load G_L. The details of these large-signal nonlearities are different for each transistor configuration, and are also dependent on the particular network chosen for F.

This paper considers the design for F, in the network forms of Fig. 2, for oscillations to start: that is, such that conditions (3) and (4) are fulfilled. To obtain oscillation to the highest possible frequency with a given load, or conversely, so that oscillation is possible at a given frequency with maximum load conductance G_L, it is necessary to optimize F for maximum potential instability: that is, for maximum negative output conductance G_0. A study has been made[1] of Y_0 for an active four-pole with general passive feedback, and with the feedback optimized for maximum potential instability. The results of this study are summarized in the next section.

The Active Four-Pole with Feedback

Fig. 3 shows an active four-pole with general feedback in the form of Fig. 2(a). With Y_G neglected for convenience, the output admittance Y_0' can be written in terms of the feedback elements and the four-pole h parameters as

[1] D. F. Page and A. R. Boothroyd, "Instability in two-port active networks," IRE Trans. on Circuit Theory, vol. CT-5, No. 3; June, 1958.

[2] This point is clarified later in this paper.

[3] It is shown in Appendix I that a practical transformer is readily represented in this form.

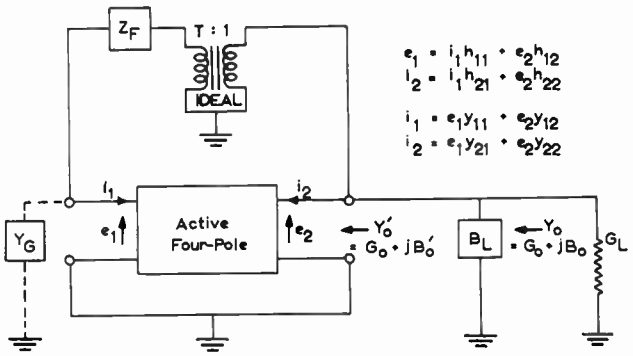


Fig. 3—Active four-pole with feedback applied proportional to output voltage.

$$Y_0' = G_0 + jB_0' = h_{22} + \frac{(T - h_{12})(h_{21} + T)}{Z_F + h_{11}}. \quad (6)$$

The last term of (6), which shall be symbolized as

$$\frac{(T - h_{12})(h_{21} + T)}{Z_F + h_{11}} = - Y_n = - (G_n + jB_n), \quad (7)$$

expresses the modification on the four-pole open-circuit admittance due to the applied feedback. Y_n may be expanded in real and imaginary parts as

$$G_n = \frac{-T^2 + T\{r_1 + r_2 + k(x_1 + x_2)\} - k(r_1x_2 + r_2x_1) - r_1r_2 + x_1x_2}{R'(1 + k^2)} \quad (8)$$

$$B_n = \frac{kT^2 + T\{x_1 + x_2 - k(r_1 + r_2)\} + k(r_1r_2 - x_1x_2) - r_1x_2 - r_2x_1}{R'(1 + k^2)} \quad (9)$$

where

$$\left.\begin{aligned} h_{12} &= r_1 + jx_1 \\ h_{21} &= -(r_2 + jx_2) \\ k &= \frac{X_F + \mathrm{Im}\{h_{11}\}}{R'} \\ R' &= R_F + \mathrm{Re}\{h_{11}\} \\ Z_F &= R_F + jX_F \end{aligned}\right\}. \quad (10)$$

The circuit of Fig. 3 is assumed in this analysis to be potentially open-circuit unstable. It therefore experiences maximum potential instability, at a given frequency, when the feedback is optimized so that instability occurs at that frequency with the highest possible value for the load G_L. This requires that the feedback be optimized so that G_0 has a maximum negative value, that is, so that G_n is maximum. A maximum value for G_n occurs for the following optimum values for T and k:

$$T^0 = \frac{r_1 + r_2 + \left\{\frac{x_1 - x_2}{r_1 - r_2}\right\}(x_1 + x_2)}{2} \quad (11)$$

$$k^0 = \frac{x_1 - x_2}{r_1 - r_2}. \quad (12)$$

The resulting maximum negative G_0 and the corresponding B_0' are

$$G_0^0 = \mathrm{Re}\left\{\frac{1}{h_{22}}\right\} - \frac{1}{4R'}\{(r_1 - r_2)^2 + (x_1 + x_2)^2\} \quad (13)$$

$$B_0'^0 = \mathrm{Im}\left\{\frac{1}{h_{22}}\right\} + k^0\left[G_0^0 - \mathrm{Re}\left\{\frac{1}{h_{22}}\right\}\right]. \quad (14)$$

This optimization has assumed that R' is positive, that is, that the active network is stable with short-circuit output. R' can, however, be negative if Re (h_{11}) is negative;[4] in this case, not considered here, the active network is short-circuit unstable due to an internal feedback mechanism proportional to output current, and external feedback should be applied accordingly.[1] Such short-circuit instability can be optimized in a manner analogous to that discussed here.

In the above expressions in terms of the h parameters, Z_F can have a loss (resistive) component. Alternatively, if Z_F is purely reactive $(=jX_F)$, an input termination admittance Y_G may be included by expressing Y_0' as follows, in terms of the real and imaginary parts of the four-pole y parameters:

$$G_0^0 = g_{22} - \frac{1}{4(g_{11} + G_G)}\{(g_{21} + g_{12})^2 + (b_{21} - b_{12})^2\} \quad (15)$$

$$B_0'^0 = b_{22} - \left\{\frac{g_{21} - g_{12}}{b_{21} - b_{12}}\right\}(G_0^0 - g_{22}). \quad (16)$$

Eqs. (13) and (15) are identical if the feedback is lossless $(R_F = 0)$, and if the input termination is lossless $(G_G = 0)$; the resulting G_0^0 is the maximum negative conductance available from the active network, and can be related to Mason's invariant U function for the active network[5] as

$$G_0^0 = \left\{g_{22} - \frac{g_{12}g_{21}}{g_{11}}\right\}\{1 - U\}. \quad (17)$$

That Y_G is not required in the feedback mechanism is evident from (15) and (16); therefore, in the following

[4] It should be noted that for junction transistors, Re (h_{11}) can be negative if the current gain α incorporates a current multiplication mechanism.

[5] S. J. Mason, "Power gain in feedback amplifiers," IRE TRANS. ON CIRCUIT THEORY, vol. CT-1, pp. 20–25; June, 1954.

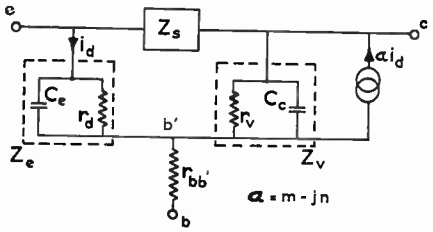


Fig. 4—Junction transistor equivalent circuit, excluding possible series emitter and collector impedance or interelectrode capacitance.

discussions Y_G is considered zero, and the above expressions in terms of h parameters are found to be the most convenient.

Y_0 for the Junction Transistor with Feedback

The foregoing results for the general active four-pole may be applied for the transistor oscillator if the h and y parameters for the transistor are substituted into the expressions. These parameters may be obtained from an equivalent circuit representation for the transistor, the form of which, used in this paper, is now discussed.

The junction transistor operating by drift and/or diffusion of minority carriers through a base region is represented by the equivalent circuit of Fig. 4.[6] Not included in this equivalent circuit are possible series emitter and collector impedance or interelectrode capacitances; these, if important, may be included in the external network.

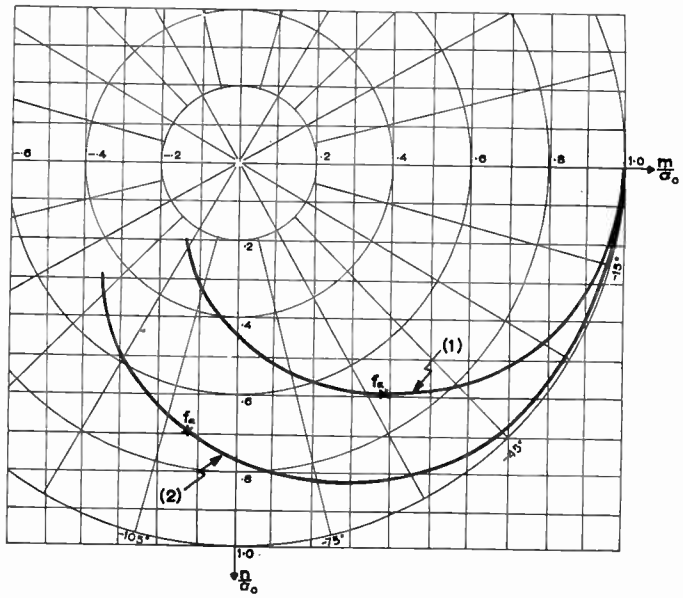
The emitter capacitance C_e includes the emitter depletion region capacitance. The current gain α relates the current crossing the collector junction to the *total* current crossing the emitter junction, including the displacement current flowing in the emitter depletion region capacitance. Z_s represents the effects of base width modulation; it has a magnitude much greater than that of Z_e and, for the present discussion is usually high enough to be neglected. $r_{bb'}$ is the effective base-spreading resistance between the effective intrinsic base b' and the external base terminal b.

The analytic form for α depends on the relative importance of drift and diffusion in the mechanism of carrier transport through the base, and on the magnitude of the emitter depletion region capacitance. For transistors operating by pure diffusion and having small emitter depletion region capacitance, α can be approximated satisfactorily at most frequencies for which oscillation is possible, by[7]

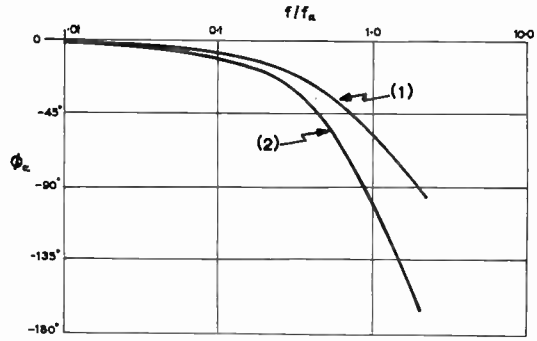
$$\alpha = \alpha_0 \left\{ \frac{1 - j0.2\frac{\omega}{\omega_\alpha}}{1 + j1.04\frac{\omega}{\omega_\alpha}} \right\} \tag{18}$$

[6] Excluded from this discussion are devices with a complex base impedance such as, for example, the rate grown transistor. Furthermore, in the case of a drift transistor, it is assumed that the collector voltage bias is high enough so that the collector depletion region extends far enough into the base region that loss in series with C_c can be neglected.

[7] R. D. Middlebrook and R. M. Scarlett, "An approximation to alpha of a junction transistor," IRE Trans. on Electron Devices, vol. ED-3, pp. 25–29; January, 1956.



(a) α/α_0



(b)

Fig. 5—Variations with frequency of the magnitude and phase of α for (1) diffusion transistors and (2) RCA 2N247 drift transistor with $f_\alpha = 38.5$ mc.

where α_0 is the low-frequency value for α, and ω_α is the frequency at which $|\alpha| = \alpha_0/\sqrt{2}$. When a drift mechanism is present, and/or if emitter depletion region capacitance is not small, the analytic expression for α is more complex, and is not in general known.

An alternative possibility is to measure α in magnitude and phase over the frequency range of interest.[8] A polar plot of a, obtained from direct measurements of the magnitude a and phase ϕ_a, is shown in Fig. 5(a) for a RCA 2N247 drift transistor.[9] Also shown is α for diffusion transistors, plotted from the one-dimensional solution for diffusion in the base;[7] for this plot, a value for a_0 of 0.99 is assumed. The phase curves in Fig. 5(b) relate the polar plots to frequency.

For circuit design at a single frequency, it is therefore convenient to express α simply in real and imaginary parts

$$\alpha = m - jn \tag{19}$$

each of which is a function of frequency. m and n can be

[8] F. J. Hyde and R. W. Smith, "An investigation of the current gain of transistors at frequencies up to 105 mc/s," *Proc. IEE*, vol. 105B, pp. 221–228; May, 1958.

[9] M. B. Das, Elec. Eng. Dept., Imperial College, London, Eng., made these measurements.

TABLE I

Y_n WITH ARBITRARY FEEDBACK

	Configuration A Common-Base Emitter Input	Configuration B Common-Emitter Base Input	Configuration C Common-Collector Base Input	Configuration D Common-Collector Emitter Input
Re $\{h_{22}\}$	$\frac{1}{r_v}+\frac{(1-m)}{r_s}$	$\frac{1}{r_s}+\frac{(1+\mu)}{r_v}+q\omega C_c$	$\frac{1}{r_s}+\frac{(1+\mu)}{r_v}+q\omega C_c$	$\frac{1}{r_v}+\frac{(1-m)}{r_s}$
Im $\{h_{22}\}$	$\omega C_c+\frac{n}{r_s}$	$(1+\mu)\omega C_c-\frac{q}{r_v}$	$(1+\mu)\omega C_c-\frac{q}{r_v}$	$\omega C_c+\frac{n}{r_s}$
G_n	$\frac{\{m-kn-T\}T}{R_b'(1+k^2)}$	$\frac{\{kq-\mu-T\}T}{R_e'(1+k^2)}$	$\frac{\{kq-\mu-(1-T)\}(1-T)}{R_e'(1+k^2)}$	$\frac{\{m-kn-(1-T)\}(1-T)}{R_b'(1+k^2)}$
B_n	$\frac{\{kT-n-km\}T}{R_b'(1+k^2)}$	$\frac{kT+q+k\mu\}T}{R_e'(1+k^2)}$	$\frac{\{k(1-T)+q+k\mu\}(1-T)}{R_e'(1+k^2)}$	$\frac{\{k(1-T)-n-km\}(1-T)}{R_b'(1+k^2)}$

read from the polar plot of Fig. 5(a), or, if the analytic form for α is known, may be calculated.

The h and y parameters for the transistor in the four relevant configurations (1) are listed in Appendix II in terms of the equivalent circuit parameters of Fig. 4. These were calculated with the following simplifying assumptions, valid at all oscillation frequencies and at normal bias and small-signal conditions:

$$|Z_s| \gg |Z_c|, r_{bb'}. \tag{20}$$

Further simplifying assumptions to be made are:

$$|Z_v(1-\alpha)| \gg Z_e \tag{21}$$

$$|Z_v| \gg r_{bb'}. \tag{22}$$

Except for poor transistors with a high $r_{bb'}C_c$ product, assumptions (21) and (22) break down only at the highest frequencies, immediately approaching the maximum oscillation frequency.

With these assumptions, the resulting h and y parameters may be substituted into (8) and (9), with the resulting G_n, B_n, and h_{22} values listed in Table I. Similar substitutions into (11) and (12) yield the optimized elements for F; these and the resulting G_n^0 and B_n^0 from (13) and (14) are listed in Table II. In these tables, the transformer turns ratio T and the quantity k are as defined in Fig. 3 and (10). The terms R_b' and R_e' are defined from (10) as

$$\left.\begin{aligned} R_b' &= R_F + r_b' \\ R_e' &= R_F + r_e' \end{aligned}\right\} \tag{23}$$

where r_b' and r_e' are the real parts of h_{11} for the common-base and common-emitter configurations respectively. The terms u and q are the real and imaginary parts of $\alpha/(1-\alpha)$,

$$\frac{\alpha}{1-\alpha} = u - jq \tag{24}$$

and are related to m and n, the corresponding components of α, as

$$\left.\begin{aligned} u &= \frac{m-(m^2+n^2)}{(1-m)^2+n^2} \\ q &= \frac{n}{(1-m)^2+n^2} \end{aligned}\right\} \tag{25}$$

TABLE II

FEEDBACK AND Y_n, OPTIMIZED FOR MAXIMUM G_n

	Configuration A	Configuration B	Configuration C	Configuration D
k^0	$-n/m$	$-q/\mu$	$-q/\mu$	$-n/m$
T^0	$\frac{m^2+n^2}{2m}$	$-\frac{\mu^2+q^2}{2\mu}$	$1+\frac{\mu^2+q^2}{2\mu}$	$1-\frac{m^2+n^2}{2m}$
G_n^0	$\frac{m^2+n^2}{4R_b'}$	$\frac{\mu^2+q^2}{4R_e'}$	$\frac{\mu^2+q^2}{4R_e'}$	$\frac{m^2+n^2}{4R_b'}$
B_n^0	$\{-n/m\}G_n^0$	$\{-q/\mu\}G_n^0$	$\{-q/\mu\}G_n^0$	$\{-n/m\}G_n^0$

A few general remarks may be made at this point. Expressions for the conductance and susceptance presented to the load are

$$G_0 = \text{Re}\left\{\frac{1}{h_{22}}\right\} - G_n \tag{26}$$

$$B_0 = \text{Im}\left\{\frac{1}{h_{22}}\right\} - B_n + B_L. \tag{27}$$

The term Re $\{1/h_{22}\}$ is seldom important in determining G_0 (except at the highest possible oscillation frequencies) and the following simplification can usually be made:

$$G_0 \doteq -G_n. \tag{28}$$

In practice, the load susceptance B_L is chosen to resonate with B_0' at the oscillation frequency; that is, B_0 is approximately zero.[2] Thus, in practice,

$$B_L = B_n - \text{Im}\left\{\frac{1}{h_{22}}\right\}. \tag{29}$$

Within the restrictions imposed by approximations (20) to (22), Tables I and II are useful as a basis for the design of feedback, with a given load G_L, such that oscillation conditions (3) and (4) are satisfied. This design can proceed from measurements (or calculation in the case of pure diffusion transistors) of the parameters h_{11} and α [or $\alpha/(1-\alpha)$]. Plots of these parameters in real and imaginary components [as in Fig. 5(a)] are convenient for this purpose. Some possibilities of this design technique are now discussed.

Design Considerations

So that the frequency will be less dependent upon B_n, which in turn is governed by transistor parameters, it is usual to include a resonant circuit in F. Thus, B_L in the transformer feedback or B_2 in the pi feedback are usually parallel resonant circuits. Oscillators with such tuned feedback arrangements include the well-known Hartley, Colpitts, and transformer-coupled circuits.

Alternatively, a series-resonant circuit may be inserted in an appropriate network branch giving short-circuit instability, for example, in the common lead of the active network in Fig. 1. Feedback may be applied in the same manner as before, and the resulting arrangement analyzed, as before by consideration of Y_0. If, however, the load is to be considered in series with a series-tuned circuit—that is, in a short-circuit unstable branch—feedback is applied proportional not to load voltage but to load current. In such short-circuit unstable circuits, feedback must be optimized for maximum negative resistance presented to the load. It must be stressed that the analyses and discussions in this paper are relevant, in dual form, to such short-circuit unstable oscillator design.

Design for maximum negative output conductance G_0^0 always produces oscillation, if oscillation be possible, for the highest possible load. At high frequencies, when the available G_0^0 drops (becoming zero at the maximum possible frequency of oscillation f_m), design for G_0^0 produces oscillation to the highest possible frequencies.

It is apparent from Table II that the four transistor configurations fall into two classes.

1) configurations A and D with

$$G_n^0 = \frac{m^2 + n^2}{4R_b'} . \tag{30}$$

2) configurations B and C with

$$G_n^0 = \frac{u^2 + q^2}{4R_e'} . \tag{31}$$

In Fig. 6, (30) is plotted, with $R_F = 0$, for the 2N247 drift transistor of Fig. 5 and for diffusion transistors types Raytheon CK760 and Philco surface barrier SB100; parameter values for these transistors are given in Table III. Also indicated for the CK760 and SB100 are the maximum oscillation frequency (f_m) values, calculated from (41). Now, since the value of G_0^0 is zero at f_m, the frequency region approaching f_m over which (30) does not accurately represent G_0^0 can be observed from Fig. 6. Factors contributing to this discrepancy are the neglect of Re $\{h_{22}\}$, and, more important, the breakdown of assumption (22). Thus the curve for the SB100 differs the most from G_0^0 at high frequencies, due to the high $r_{bb'}C_c$ value for this transistor. The value of f_m for the 2N247 transistor, given by the manufacturer as approximately 130 mc, was not calculated because measurements for α were not available beyond 60 mc. It

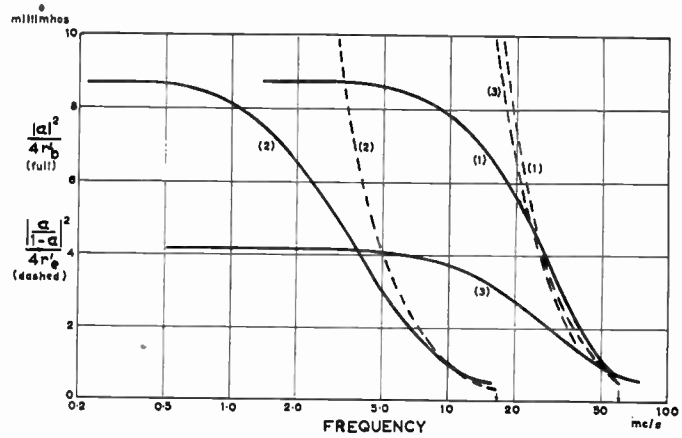


Fig. 6—Maximum possible negative value of G_n for the transistors of Table III: (1) 2N247, (2) CK760, and (3) SB100.

TABLE III
Example Transistor Parameters

Transistor	f_α mc	$r_{bb'}$ ohms	C_c $\mu\mu$f	α_0	r_d ohms	C_e $\mu\mu$f
2N247 drift*	38.5	21	1.6	0.966	26	100
CK760 alloy junction	5.4	47.5	16	0.977	26	
SB100 surface barrier	73.6	318	2.3	0.981	52	

* Same sample as used for Fig. 5.

appears however, from Fig. 6, that the available negative G_0^0 for this transistor drops to a low value for a considerable range of frequencies below f_m.

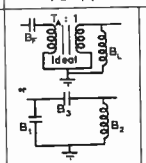
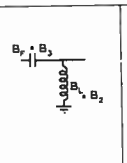
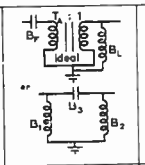
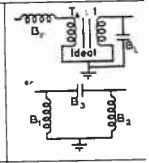
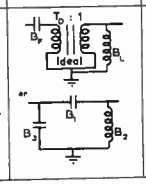
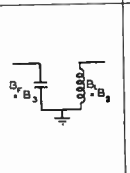
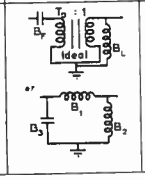
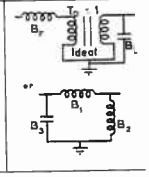
The dashed curves in Fig. 6 represent the high-frequency values for (31); lower frequency values, though very high, are associated with impractical high values for output susceptance.

The elements of F, optimized to give G_0^0, are given in Table II. The value for Z_F^0 is found from the value for k^0 and from (10); this, together with the optimum transformer turns ratio T^0 and the load susceptance B_L^0 (approximately equal to B_n^0), gives the optimized F in the form of Fig. 2(a). Straightforward transformations give the corresponding optimum pi and tee forms. For oscillation frequencies lower than that at which $-\phi_\alpha = 180°$, Table IV lists the optimized transformer and pi feedback networks for transistor configurations A and D. This Table IV can be extended to higher frequencies, if necessary, from the expressions given for the network elements. An identical table can be drawn up for configurations B and C.

Limitations on this design for G_0^0 arise from large-signal, nonlinear considerations. It is desirable to keep the rate of build-up of oscillations (or regeneration) sufficiently small to prevent grossly nonlinear circuit operation (and perhaps blocking); otherwise a poor output waveform may result. The regeneration may be reduced by using large energy storage in a tuned circuit (a high C/L ratio) and keeping the load conductance high. This last point requires that the allowable transistor dissipation be borne in mind at low frequencies where the available G_0^0 is high. If the regeneration is

TABLE IV*

SOME LOSSLESS FEEDBACK NETWORKS GIVING MAXIMUM NEGATIVE OUTPUT CONDUCTANCE

CIRCUIT	F^* for frequencies such that $-\phi_a < \phi_1$	$-\phi_a = \phi_1$	$\phi_1 < -\phi_a < 90°$	$90° < -\phi_a < 180°$
Configuration A				
Configuration D				

where

$$T_A = \frac{m^2+n^2}{2m} \qquad B_1 = \frac{2m-(m^2+n^2)}{2(nr_b'+mx_b')}$$

$$T_D = 1-\frac{m^2+n^2}{2m} \qquad B_2 = \frac{-(m^2+n^2)}{2(nr_b'+mx_b')}\left\{1-\frac{m}{2}+\frac{n}{2}\left[\frac{x_b'}{r_b'}\right]\right\}$$

$$B_F = \frac{m}{nr_b'+mx_b'} \qquad B_3 = \frac{m^2+n^2}{2(nr_b'+mx_b')}$$

$$B_L = \frac{-n}{m}\left\{\frac{m^2+n^2}{4r_b'}\right\} \qquad m-jn=\alpha$$

$$r_b'+jx_b' = h_{11_b} \qquad \phi_1 = \cos^{-1}\left\{\frac{|a|}{2}\right\} \qquad 0<\phi_1<90°$$

* This table assumes a positive value for the term $(nr_b'+mx_b')$. If this term is negative, B_F, B_1, B_2, and B_3 are capacitive where shown inductive and are inductive where shown capacitive. Furthermore, if x_b' is negative, the nature of B_2 must be examined more closely.

thus suitably restricted, the change of oscillation frequency, as oscillations rise to the limiting amplitude, should be small, provided the frequency is determined largely by an external tuned circuit with susceptances large compared with B_0'.

Thus at relatively low frequencies, where a large degree of regeneration is available, optimization of the circuit in regard to G_0^0 is possible and worthwhile if the oscillator must operate into a large load conductance. On the other hand, at high frequencies, where the possible regeneration is limited, it is essential to optimize G_0^0 if oscillation to the maximum frequencies is to be realized.

HIGH-FREQUENCY CONSIDERATIONS

It is of interest to observe the change with frequency of the optimum feedback transformer turns ratio. Consider, for example, configurations A and D, for which T^0 is given by

$$T_A^0 = 1 - T_D^0 = \frac{m^2+n^2}{2m}. \tag{32}$$

The respective low-frequency values of T_A^0 and T_D^0 are therefore $\alpha_0/2$ and $1-\alpha_0/2$ for the diffusion and/or drift transistors under consideration. In Fig. 7, T_A^0 has

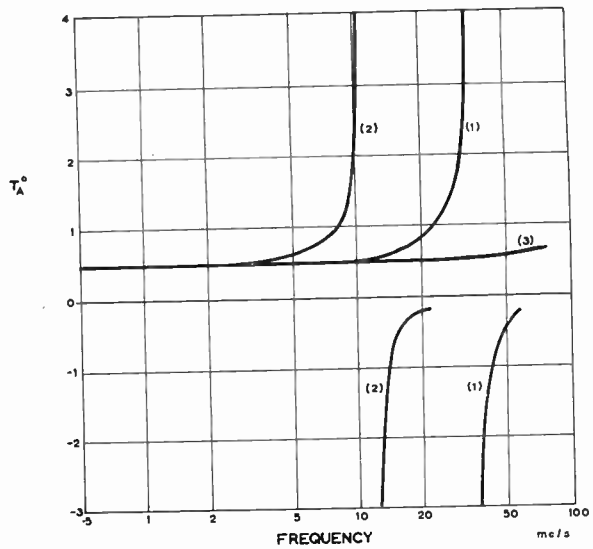


Fig. 7—Optimum common-base oscillator turns ratio for the transistors of Table III: (1) 2N247, (2) CK760, and (3) SB100.

been plotted with frequency for the three example transistors of Table III. Certain frequencies are of obvious interest.

At the frequency f', where

$$2m = m^2+n^2, \qquad -\phi_a = \cos^{-1}\left\{\frac{|a|}{2}\right\}, \tag{33}$$

the optimum values for transformer turns ratio are

$$\left.\begin{aligned} T_A^0 &= 1 \\ T_D^0 &= 0 \end{aligned}\right\}. \tag{34}$$

This is the frequency at which internal feedback within the transistor, in the common-collector configuration, is just sufficient when correct reactive terminations are applied to achieve maximum potential instability; that is, any external feedback, if applied, would degrade the instability. In the common-base configuration, maximum instability is achieved at f' by connecting Z_F^0 directly from collector to emitter. If feedback is expressed in the pi form, a further point is evident from Table IV. Only at frequencies below f' can the oscillator be optimized in the familiar Colpitts tapped-capacitor form. For diffusion transistors, f' is approximately $1.5f_\alpha$.

Another frequency of interest is f'', at which

$$\left.\begin{aligned} m &= 0 \\ \phi_\alpha &= -90° \end{aligned}\right\}. \tag{35}$$

Here, the values for T^0, k^0, and B_n^0, given by Table II, become very high. The expression for configuration A

$$G_n = \frac{\{m-kn-T\}T}{R_b'(1+k^2)} = -\frac{\{kn-T\}T}{R_b'(1+k^2)}, \tag{36}$$

however, may be reduced approximately to its maximum value

$$G_n = \frac{n^2}{4R_b'} \tag{37}$$

if

$$\left.\begin{aligned} k^2 &\gg 1 \\ T &= \frac{-nk}{2} \end{aligned}\right\}. \tag{38}$$

Thus, practical k and T values can give very nearly the maximum negative output conductance $G_0{}^0$. The corresponding B_n, however, is

$$B_n = \frac{n^2}{4R_{b'}}\left\{k + \frac{2}{k}\right\} \tag{39}$$

and thus is very high.

Now it may be required to reduce dependence of the oscillation frequency on B_n, by using external tuned circuit susceptances much larger than B_n; such a requirement arises, for example, if the oscillation frequency is to be stabilized against changes in bias levels or temperature. At f'', however, B_n can be greatly reduced only at the expense of greatly reducing G_n; B_n is zero only when G_n is zero.

A further point arises here. As oscillations increase in amplitude and limit, the effective value of G_n is reduced in magnitude; in the same way, the effective value of B_n changes from its small-signal value. If B_n is significant in determining frequency, as is likely at f'' where its value is high, the large-signal frequency differs from the frequency at start of oscillation. This difference may be somewhat more than that for the case where external tuned circuit susceptances are used which are large compared with B_n.

The critical frequencies f' and f'', described above, are relevant to configurations A and D. Similar frequencies occur for configurations B and C when $\alpha/(1-\alpha)$ has the same phase angles.

The expressions of Table I may be used to determine the deviations from optimum performance when a spread in transistor parameters is to be accommodated in the design. This technique must be used also when the oscillator is to be tuned over a range of frequencies, where it is usually desirable to design the feedback for optimum performance at the high frequency end of the tuning range.

Maximum Oscillation Frequency

The maximum oscillation frequency f_m is calculated assuming lossless terminations and feedback, and is the same for all transistor configurations.[5] It is that frequency for which the maximum available negative $G_0{}^0$ becomes equal to zero;[10] therefore, from (15) f_m occurs for

$$4g_{11}g_{22} = (g_{21} + g_{12})^2 + (b_{21} - b_{12})^2. \tag{40}$$

If the y parameters for the circuit of Fig. 4 are substituted in (40), then with the simplifications that r_v and Z_e are high and r_d is low, the following expression for frequency is obtained.

$$\omega_m = 2\pi f_m = \frac{m^2 + n^2}{4nr_{bb'}C_c}. \tag{41}$$

It is interesting to note that though C_e does not appear explicitly in this expression, its effect is included in α. The neglect of r_d requires the assumptions

$$\left.\begin{aligned} r_d &\ll \frac{nr_{bb'}}{\omega\tau_e}\left\{1 + (\omega\tau_e)^2\right\} \\ r_d &\ll (1-m)r_{bb'}\left\{1 + (\omega\tau_e)^2\right\} \\ r_d &\ll \frac{nr_{bb'}}{n + (1+m)\omega\tau_e}\left\{1 + (\omega\tau_e)^2\right\} \end{aligned}\right\}. \tag{42}$$

where

$$\tau_e = r_dC_e.$$

Expression (41) is a useful transistor criterion, indicating the absolute maximum possible oscillation frequency. Furthermore, assumptions (42) are often valid at f_m, especially for diffusion transistors. If α is given the oversimplified analytic form

$$\alpha = \frac{\alpha_0}{1 + j\dfrac{\omega}{\omega_\alpha}}, \tag{43}$$

(41) reduces to the form, widely used for diffusion transistors,

$$f_m{}^2 = \frac{\alpha_0 f_\alpha}{8\pi r_{bb'}C_c}. \tag{44}$$

The value for f_m from (41) may be observed from the polar plot for α, with a knowledge of the product $r_{bb'}C_c$; the validity of assumptions (42) also may be observed from a knowledge of $\omega\tau_e$. For diffusion transistors, f_m occurs at an angle somewhat greater than 60 degrees on the polar α plot, except for poor units having high $r_{bb'}C_c$ products. For drift transistors, f_m can occur at angles approaching, or greater than, 180 degrees.

The assumptions (21) and (22), upon which are based the design expressions of Table I and II, may be examined now in the light of (41). Assumption (21)

$$|Z_v(1-\alpha)| \gg |Z_e|$$

may, for the transistors under discussion, be considered valid to f_m, if at f_m

$$|Z_v| \gg |Z_e|. \tag{45}$$

This is a reasonable assumption. Assumption (22)

$$|Z_v| \gg r_{bb'}$$

is, at f_m,

$$r_{bb'}\omega_m C_C \ll 1. \tag{46}$$

From (41), this assumption becomes

[10] Or, Mason's U function for the transistor may be set equal to unity.

$$\frac{m^2+n^2}{4n} \ll 1. \tag{47}$$

It is clear from Fig. 5 that this is a valid assumption for drift transistors for which the phase of α is of the order of 180 degrees (or greater) at ω_m. For diffusion transistors this assumption is reasonably valid provided the $r_{bb'}C_c$ product is low enough so that f_m is not much smaller than f_α. The fact that the true maximum oscillation frequency is lower than that given by (41), because of the presence of r_d, increases the validity of these assumptions to the highest oscillation frequencies.

ω_m, given by (41), may be reduced to a lower frequency ω_m' if $r_{bb'}$ is artificially increased by adding a series base resistance R_B; such an artifice increases the validity of assumptions (42), so that (41) can be used to express ω_m' accurately. A method thus arises for the measurement of C_c, from a measurement of ω_m' and a knowledge of $(r_{bb'}+R_B)$, m, and n at ω_m'; R_B may be appreciably greater than $r_{bb'}$. Such a measurement gives the true operating value of C_c. Or, if C_c is known from other measurements, this technique may be used to deduce the phase of α at ω_m' from a measurement of the amplitude $|\alpha|$ at ω_m'. In these measurements, ω_m' can be deduced from bridge measurements of G_0 shunted by an external load G_L such that G_0+G_L is positive. Such methods have been used, giving substantial agreement with direct measurements.

Conclusions

A unified approach to the design of transistor oscillators has been presented. Design of initial oscillation conditions is conveniently carried out for open-circuit unstable circuits by consideration of the dynamic admittance presented to the load, at the oscillation frequency, by the transistor with general feedback. The technique leads to simple design expressions when appropriately simplified transistor parameters are used. An exactly similar approach will apply for circuits which are short-circuit unstable.

Oscillation to the highest possible frequencies can be achieved, with a given load, if optimized feedback is used. Such optimum feedback networks are easily designed in terms of a general feedback network form from which any other feedback form can be derived. At high frequencies, such design must, of course, include the effects of stray capacitances and lead inductances.

A general approach to the maximum oscillation frequency has yielded a simple expression, useful for transistors with arbitrary base transfer characteristics. The expression also leads to new techniques for the measurement of certain high frequency transistor parameters.

It is suggested that the approach to oscillator design presented in this paper could be extended to studies of 1) frequency stabilization, from a study of B_0', 2) frequency modulation, 3) large-signal limiting conditions and available output power, and 4) design of oscillators for special applications. A brief discussion of these possibilities follows, with the common-base transistor configuration chosen for example.

Stabilization of frequency is most easily achieved when a crystal replaces one of the reactances in the feedback network. If a crystal is not used, frequency stabilization against transistor parameter change requires that B_0' be made small compared with the resonant circuit susceptances. From Table I, for configuration A, B_n is zero for the feedback element values

$$T' = \frac{m}{2}; \qquad k' = \frac{-2n}{m}.$$

The resulting G_n is

$$G_n' = \frac{m^2}{4R_b'}.$$

From these expressions, it is apparent that this type of frequency stabilization becomes impossible as ϕ_α approaches -90 degrees and m approaches zero. Stabilization could be carried further by a study of the rate of change of B_0 with various transistor parameters.

A corollary to the above discussion suggests that the oscillation frequency could be modulated by suitably modulating B_0'; for example, in configuration A the emitter current could be modulated, thus modulating the value of R_b'.

This design technique can be extended to the study of large-signal oscillation conditions, if suitable large-signal transistor parameters are used. Since the design relationships arise from a set of linear equations for the active network with feedback, the use of these relationships at large-signal levels requires that the circuit can be appropriately "linearized." In a practical oscillator this can be done because the loop gain for harmonics of the oscillation frequency is low.

Other oscillator requirements may be studied using this approach. In a linear superregenerator, for example, it is required that G_n and B_n stay reasonably constant to high oscillation levels. This requires a high value for k; for example, in the common-base configuration and when $-\phi_\alpha < 90$ degrees, this requires a low feedback capacitance. The resulting load restrictions are seen in Table I. Another example concerns the design of Class-A oscillators with low harmonic content. Insertion of a large feedback resistance R_F will give "linear Class-A" oscillation; the load restrictions are calculated from Table II. Further examples might include oscillators using more than one transistor.

Finally, the concept of maximum potential instability for a transistor with feedback, and its realization by optimization of the feedback, can be applied, at frequencies approaching f_m, to the design of tuned amplifiers having the highest power gain with a given degree of stability.[11]

[11] This problem has been discussed for the transistor without feedback by A. P. Stern, "Stability and power gain of tuned transistor amplifiers," Proc. IRE, vol. 45, pp. 335–343; March, 1957.

TABLE V

	Configuration *A* Common-Base Emitter Input	Configuration *B* Common-Emitter Base Input	Configuration *C* Common-Collector Base Input	Configuration *D* Common-Collector Emitter Input
h_{11}	$Z_e+\dfrac{1-a}{\dfrac{1}{r_{bb'}}+\dfrac{1}{Z_v}}$	$\left\{\dfrac{r_{bb'}+Z_v}{Z_e+Z_v(1-a)}\right\} h_{11A}$	h_{11B}	h_{11A}
h_{22}	$\dfrac{1}{r_{bb'}+Z_v}\left\{1+\dfrac{Z_v(1-a)}{Z_s}\right\}$	$\dfrac{1}{Z_e+Z_v(1-a)}\left\{1+\dfrac{Z_v(1-a)}{Z_s}\right\}$	h_{22B}	h_{22A}
h_{12}	$\dfrac{r_{bb'}}{r_{bb'}+Z_v}+\dfrac{h_{11A}}{Z_s}$	$\dfrac{Z_e}{Z_e+Z_v(1-a)}$	$\dfrac{Z_v(1-a)}{Z_e+Z_v(1-a)}$	$\dfrac{Z_v}{r_{bb'}+Z_v}$
h_{21}	$-\dfrac{r_{bb'}+aZ_v}{r_{bb'}+Z_v}$	$\dfrac{aZ_v-Z_e}{Z_e+Z_v(1-a)}$	$\dfrac{-Z_v}{Z_e+Z_v(1-a)}$	$\dfrac{-Z_v(1-a)}{r_{bb'}+Z_v}$
y_{11}	$1/h_{11A}$	$\left\{\dfrac{Z_e+Z_v(1-a)}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	y_{11B}	y_{11A}
y_{22}	$\dfrac{1}{Z_s}+\left\{\dfrac{r_{bb'}+Z_e}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	y_{22A}	y_{11A}	y_{11B}
y_{12}	$-\dfrac{1}{Z_s}-\left\{\dfrac{r_{bb'}}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	$\left\{\dfrac{-Z_e}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	$\left\{\dfrac{-Z_v(1-a)}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	y_{21C}
y_{21}	$-\left\{\dfrac{r_{bb'}+aZ_v}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	$\left\{\dfrac{aZ_v-Z_e}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	$\left\{\dfrac{-Z_v}{r_{bb'}+Z_v}\right\}\dfrac{1}{h_{11A}}$	y_{12C}

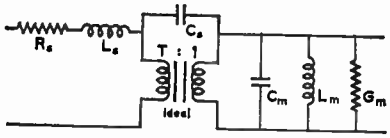


Fig. 8—Equivalent circuit for a practical transformer.

Appendix I

Many useful oscillators have a transformer in the feedback network; the equivalent circuit for a practical transformer is shown in Fig. 8. If the effects of interwinding capacitance C_s are neglected, the feedback network of Fig. 2(a) is realized by lumping the transformer parameters with the feedback elements.

L_s = leakage inductance. Lump with Z_F.
L_m = magnetizing inductance. Lump with Y_L.
C_m = stray capacitance. Lump with Y_L.
G_m = transformer loss. Lump with Y_L.
R_s = winding resistance. Lump with Z_F.

If C_s is not negligible, two feedback mechanisms are present: one is due to the transformer [and of the form represented by Fig. 2(a)], and the other to impedance-coupled feedback [of the form represented by Fig. 2(b)].

Appendix II

For the general transistor equivalent circuit of Fig. 4, the h and y parameters are listed in Table V. The assumptions

$$|Z_s| \gg |Z_e|, \quad r_{bb'}$$

have been made in the calculation of these parameters.

Acknowledgment

This work resulted from a program of research under the supervision of Dr. A. R. Boothroyd, to whom the author is indebted for many helpful discussions, and whose advice in preparing this manuscript was invaluable.

Properties of Silicon and Germanium: II*

E. M. CONWELL†

Summary—This paper attempts to bring up to date the information on fundamental properties of silicon and germanium. Much the same topics are covered as in the author's earlier article "Properties of Silicon and Germanium" (henceforth referred to as I), which appeared in the 1952 Transistor Issue of PROCEEDINGS. Also included is some of the detailed knowledge on the band structure which has been obtained since 1952. This is essential to the understanding of many of the properties of these materials.

INTRODUCTION

THE amount of new information on silicon and germanium that has been obtained since 1952, the date of the first Transistor Issue of PROCEEDINGS, is overwhelming, as most readers no doubt know. It has been said, in fact, that germanium is now one of the best known of all materials. Anything like complete coverage of all that is interesting, useful, or even important is not possible. This paper attempts to do three things: 1) present the physical background of the important advances made since 1952; 2) give a sufficient (not complete!) bibliography to make it easy for the reader to go further in the areas of his particular interest, and 3) give the values for important physical quantities considered best at the time of writing.

A major advance since 1952, from the basic point of view, has been the acquisition of detailed knowledge concerning the band structure of silicon and germanium. This knowledge has made it possible to understand many properties previously considered anomalous. Since it is basic, we begin with a short description of the band structure, designed (hopefully) to give some physical insight into what is involved. The treatment starts from fundamentals and includes a minimum of mathematical machinery. This is followed by a section dealing with properties of conduction electrons and holes under the headings of mobility, optical and infrared absorption, some properties of intrinsic material, and Zener current. The last section is concerned with effects more directly related to impurities and defects in the materials. It contains discussions of energy levels due to impurities and defects, and the effects of heat treatment. The subject of the lifetime of excess carriers is not taken up; it is covered in papers by Shockley[1] and Bemski.[2] Great advances also have been made in the study of surfaces of germanium and silicon. These, too, are beyond the scope of this article. The interested reader is referred to Kingston.[3]

* Original manuscript received by the IRE, May 1, 1958.
† Sylvania Electric Products, Inc., Bayside, N. Y.
[1] W. Shockley, "Electrons, holes, and tapes," this issue, p. 974.
[2] G. Bemski, "Recombination in semiconductors," this issue, p. 990.
[3] R. H. Kingston, ed., "Semiconductor Surface Physics," Universilty of Pennsylvania, Philadephia, Pa.; 1957.

BAND STRUCTURE OF SILICON AND GERMANIUM

According to quantum theory, a wave, described mathematically by a wave function, is associated with every particle. For a free electron the wave function can be taken as a plane wave exp $(iP\cdot r/\hbar)$, where P is the momentum, r the spatial coordinate, and $\hbar$ Planck's constant divided by 2π. The energy is then $P^2/2m_0$ where m_0 is the free electron mass. A plot of energy E vs P is a parabola through the origin. For an electron inside a crystal a basically plane-wave function exp $(iP\cdot r/\hbar)$ can still be used, but this is modulated now by a function with the periodicity of the lattice. P no longer really represents the electron momentum, because this is not constant inside the crystal on account of the potential variations. Nevertheless, it is still convenient to characterize the wave function and energy of the electron as a function of P, which is now called the crystal momentum. Actually, P still retains some of the important properties of a momentum; one of these is that it changes under an applied force with a time rate of change equal to the force. Also in distinction to the free electron case, for the electron in the crystal there is a finite, although very closely spaced, set of allowed values of P. This set is determined by the boundary conditions imposed on the wave functions at the edge of the crystal. Due to the periodicity of the crystal, only a certain number of these P's lead to physically different behavior, and the others can be neglected. The volume of P-space, centered at $P=0$, containing this reduced group of P's is called the Brillouin Zone [1].

To determine the relation between energy and P for the electrons of a particular crystal is a matter for detailed and difficult calculation. Some of the possible results of such a calculation are shown in Fig. 1. Only the regions around the band edges (*i.e.*, minimum energy for the conduction band and maximum energy for the valence band) have been shown, because these are where the characteristically few carriers of a semiconductor will be. In the first column is the so-called simple model, where the energy measured from the edge of the band is given by a constant times P^2, so that E vs P is a parabola as in the free electron case. For this model the same parabola would be obtained whatever direction of P had been chosen for the plot. The similarity between this and the free electron case can be emphasized by writing the constant of proportionality between E and P^2 as $\frac{1}{2}m$, *i.e.*, $E=P^2/2m$. Another way of representing this result is shown at the bottom of the first column. If all P's belonging to the same E are connected, a single sphere is obtained just as for a free electron. As a result of this simple relation between

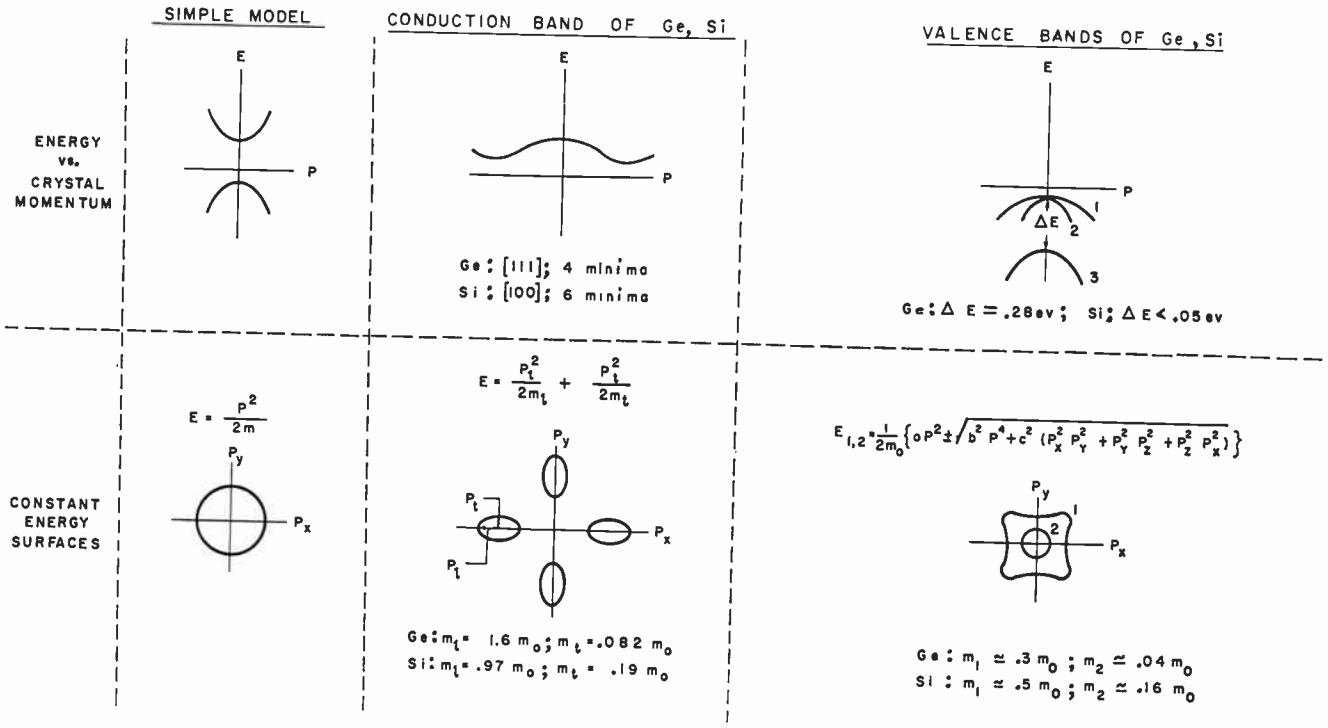


Fig. 1—Band structure of silicon and germanium.

E and P, it is possible for most purposes to treat the electron inside the crystal like a free electron, using, however, a different number for the mass. It is obvious why, in the absence of detailed information on the band structure of a material, one would prefer to calculate with this model. It is mainly the results of these calculations which make up the usual semiconductor theory, such as that covered by the author [2].

It is now known from cyclotron resonance experiments and many other types of evidence that for the conduction bands of silicon and germanium the relation between E and P is as shown, rather schematically, in the second column of Fig. 1. The plot is for P in the (100) direction in the case of silicon, (111) in the case of germanium; it would not be the same for other directions [3]. The major difference between this and the simple model is that the minimum energy is not at $P=0$ but at some other P well out in the Brillouin Zone. The minima in the directions chosen for the E vs P plots actually represent the minimum energy in the conduction bands. The fact that minimum energy is not at $P=0$ gives rise to further complication. Symmetry requires that all equivalent P's, such as two equally spaced on opposite sides of $P=0$ (*i.e.*, all P's reachable from each other by crystal symmetry operations), have the same energy. Thus the band edge or position of minimum energy must be multiple, and the electron population must be divided among the various minima, now commonly called valleys. In silicon there are six such valleys, one for each of the six cube-edge directions. In germanium, eight would be expected because the minimum is along a body diagonal direction. Actually, there are only four. The point of minimum energy in germanium turns out to be at the edge of the Brillouin Zone, so that each minimum can contribute only half a valley. In silicon the point of minimum energy is 0.8 of the distance from the center to the edge of the Brillouin Zone.

Within each valley, not far from the minimum energy, E can be represented as a quadratic function of the components of P measured from the minimum energy point. Thus the points in P-space belonging to a particular energy are on a set of ellipsoids, one for each minimum. Actually, these are ellipsoids of revolution here, so the energy can be written as a constant times P_l^2 plus another constant times P_t^2, where P_l and P_t are longitudinal and transverse components. Fig. 1 shows this situation for silicon in the lower half of the second column. Because the longitudinal and transverse components are unequal, we must deal with two effective masses, m_l and m_t. The values of these masses, obtained by cyclotron resonance [4], are at the bottom of the second column in Fig. 1. Other sections of this paper discuss the way this type of band structure complicates transport and other properties of conduction electrons in silicon and germanium [5].

Consider now the structure found for the valence band. In the simple model for the valence band, electron energy measured from the top of the band decreases as a constant times P^2, giving the inverted parabola in the first column of Fig. 1. Hole energy then increases as P^2, just as electron energy does in the conduction band. What is actually found for germanium and silicon is seen in the third column of Fig. 1. It is a set of three bands, one split off by an energy ΔE from the other two. If the spin of the electron is taken into

account each of these bands is doubly degenerate. The occurrence of this type of band structure can be made understandable in the following way. Calculations show that for the top of the valence band the wave function (which consists only of the periodic modulating part since $P=0$ and $\exp(iP\cdot r/\hbar)=1$ for this case) has within each cell the character of an atomic p state. Without spin these states are triply degenerate. This would give rise to three bands stuck together at the origin of P space. In the atomic case, the interaction of spin and orbital motion breaks up the degeneracy of the p states, resulting in final states of different energies characterized by $j=3/2$ and $j=1/2$. The $j=3/2$ state is quadruply degenerate, and the $j=1/2$ state doubly degenerate. It has been shown by Elliott that spin-orbit interaction will operate in a similar way in the solid also, splitting one band off from the other two [6]. The two top bands have symmetry properties related to $j=3/2$, while the bottom one is of the $j=1/2$ type.

As is the case for the conduction band, E vs P for the valence band is different in different directions. This results in the constant energy surfaces being warped rather than spherical. The surfaces from the two upper bands belonging to a given energy are shown, again rather schematically, in the lower part of the third column of Fig. 1; their equation is given also. The values of a, b, and c have been determined by the cyclotron resonance experiments [4]. Because of the warping of the surfaces, the effective masses are a function of direction. The anisotropy is not very great, however, particularly for germanium, and to a good approximation we can describe the holes in bands 1 and 2 by the masses presented in Fig. 1. The band with the greater curvature has the smaller mass. At temperatures low enough so that $kT \ll \Delta E$, only these two bands will be populated. At these temperatures the lighter holes should constitute about 5 per cent of the total hole population in germanium and about 20 per cent in silicon. In the case of silicon, if the splitting between bands is really as small as shown, there will be holes in all three bands even below room temperature. This should complicate considerably the analysis of experimental data.

Properties of Conduction Electrons and Holes

Mobility

Mobility determinations in germanium and silicon now are sufficiently precise to make it worthwhile to distinguish between values obtained from different types of measurements. The basic measurements are listed below.

1) Drift mobility, μ_D, in which the drift velocity of a pulse of injected carriers is measured in a known electric field.
2) Conductivity mobility, μ_C, which is the mobility calculated from the conductivity and the number of carriers, the latter obtained by a method other than Hall measurement. (Methods that have been used involve determination of impurity concentration by use of radioactive impurities, or neutron activation analysis.)
3) Hall mobility, μ_H, which is the product of Hall constant, R, and conductivity, σ.

The relations among these three mobilities are well known for the simple model of the band structure [7]. It is not necessarily true, however, that the same relations hold for the actual band structures of germanium and silicon, and re-examination of the entire problem is necessary to determine this. We shall consider first drift and conductivity mobility, and show that under the experimental conditions usually employed these are still equal. Hall mobility is taken up later.

To illustrate what is involved, take the case of holes in germanium. The conductivity for this case is

$$\sigma = n_1 e\mu_1 + n_2 e\mu_2 \tag{1}$$

where again the subscript 1 is taken to refer to the heavy holes, 2 to the light holes. The latter are expected to have higher mobility because of their smaller inertia. The conductivity mobility is then

$$\mu_C = \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}, \tag{2}$$

which is simply the average mobility of the two holes. A pulse of holes injected into a specimen for a drift measurement will consist of the two kinds of holes, with two different drift velocities. It is then conceivable that two separate pulses of holes will be observed at the collector. In fact, this has been looked for very carefully but not found [8]. The explanation lies in the fact that upon being scattered the two kinds of holes can make transitions between the bands, *i.e.*, change from being light to heavy holes and vice versa. The time for this switching apparently is very short compared with the average drift time. Consequently one can expect that, under the conditions of the usual drift experiment, all holes will have spent $n_1/(n_1+n_2)$ of their time in the heavy-hole band, $n_2/(n_1+n_2)$ in the light-hole band. This leads to observation of a single pulse at the collector [9] with a drift mobility $\mu_D = \{n_1/(n_1+n_2)\}\mu_1 + \{n_2/(n_1+n_2)\}\mu_2$, which is equal to μ_C. Of course, we have assumed that μ_D is measured with appropriate precautions, *e.g.*, by using suitably extrinsic samples with negligible trapping. It is presumed that equality of drift and conductivity mobility will hold for p-silicon also.

In the case of n-type germanium and silicon the expression for σ must also involve a sum, as in (1), in this case over all of the valleys. There is the additional complication here that the mobility of the carriers within any one valley is quite anisotropic due to the anisotropy of the effective mass. The over-all conductivity and conductivity mobility are not anisotropic, however,

TABLE I

LATTICE MOBILITIES IN CM²/VOLT-SEC AT 300°K

Carrier	Germanium		Silicon	
	Drift Mobility	Hall Mobility	Drift Mobility	Hall Mobility
Electron	3900 ± 100 [11]	3950 ± 250* [12]	1350 ± 100 [14]	ca. 1900 [15]
Hole	1900 ± 50 [11]	ca. 3400*[13] at 3000 oersteds	480 ± 15 [14]	ca. 425 [15]

* These are not actual values of $R\sigma$, which for high resistivity samples at 300°K represent a combination of electron and hole mobilities. They are obtained by extrapolation of the respective one-carrier Hall mobilities from lower temperatures.

TABLE II

TEMPERATURE DEPENDENCE OF LATTICE MOBILITY IN CM²/VOLT-SEC

Carrier	Germanium		Silicon	
	Mobility	Temperature Range	Mobility	Temperature Range
Electron	$4.9 \times 10^7\ T^{-1.66}$	100–300K [18]	$(2.1 \pm 0.2) \times 10^9\ T^{-2.5 \pm 0.1}$	160–400K [14]
Hole	$1.05 \times 10^9\ T^{-2.33}$	125–300K [18]	$(2.3 \pm 0.1) \times 10^9\ T^{-2.7 \pm 0.1}$	150–400K [14]

because of the adding of the contributions of all of the valleys [10]. Here again it is reasonable to conclude that under the usual conditions of measurement, μ_D and μ_C reflect the same type of average. In the discussion that follows, therefore, we shall generally not distinguish between μ_D and μ_C.

Experimental values of drift mobility in high-resistivity germanium and silicon at 300°K are presented in Table I. These values have been obtained on samples pure enough so that impurity scattering is negligible, as evidenced by the fact that mobility does not increase beyond these values with further increase in ρ [16]. They therefore represent lattice mobility. It is noteworthy that with the exception of p-silicon the values are not much higher than the ones given in I. Even in the case of silicon, the values seem to have settled down; other and more recent determinations of both drift and Hall mobility are in substantial agreement with the above. It should be remarked, however, that fluctuations in mobility among silicon samples in the same resistivity range remain relatively large [17]. Still, there appears no reason to expect a significant increase in mobility as better silicon becomes available.

The real change observed (or at any rate acknowledged) in the mobility since 1952 is in the matter of the temperature dependence. This is summarized for the conductivity and drift mobility in Table II. In all four cases the temperature dependence is steeper than the $T^{-3/2}$ predicted for lattice mobility. This theoretical prediction is based on the simple model of the band structure and on a number of assumptions about the interaction of current carriers with the lattice vibrations. To understand these assumptions, and the proposals that have been made to account for the steeper temperature dependence, it is necessary to go somewhat more deeply into the theory of lattice mobility. This is especially worthwhile because interaction of current carriers with lattice vibrations has been found to play an important role in, for example, infrared absorption and thermoelectric power.

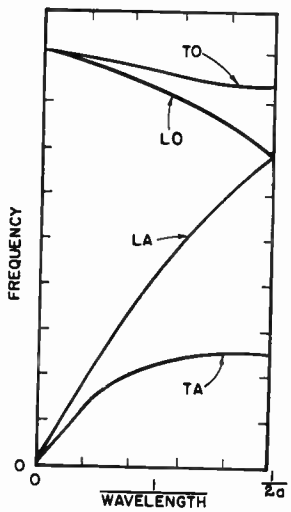


Fig. 2—Schematic plot of frequency vs reciprocal of wavelength for lattice vibrations of germanium or silicon. This has been drawn for the (100) direction. A and O stand for acoustical and optical and T and L for transverse and longitudinal, respectively. In this direction the two sets of transverse vibrations of either mode lie on the same curve.

The vibrations of the lattice atoms can be resolved into a set of normal modes that are running waves. Frequency vs reciprocal of wavelength for these vibrations is illustrated in Fig. 2. They range in wavelength from a minimum of twice the lattice constant a to a maximum of the order of the size of the crystal, which is effectively infinite for our purposes. For the lattice mobility, as will be shown, it is the long wavelength modes that are of greatest importance. Some of these long-wavelength modes are familiar as sound waves—specifically the ones of the two lower branches. The vibrations of these branches are therefore referred to as acoustical. In germanium and silicon there are two atoms in a unit cell. Under the acoustical vibrations,

the two atoms in the unit cell move almost together, *i.e.*, with only a small phase difference between them. Another type of long wavelength vibration is possible: the two atoms in the unit cell move opposite to each other, with the same thing being repeated in the next cell in slightly different phase. These vibrations can also be both longitudinal and transverse. They are called optical modes because, in a lattice with ions of opposite sign, they produce a polarization that can interact with light waves [19].

For the long wavelength acoustical waves the relation between frequency of oscillation ν and wavelength λ is $\nu\lambda = \text{constant}$. This constant is identified as the phase velocity. Thus ν varies inversely with λ and is very small in the limit of large λ for the acoustical branches of the spectrum. In the case of the optical modes the opposing motion of the two atoms in the unit cell gives rise to a large restoring force and therefore a high frequency of vibration even for very large wavelengths. This has the further consequence that the frequency is not strongly dependent on λ, since λ only determines the relative phase of adjacent unit cells.

The vibrational energies of each normal mode, of course, can have only certain discrete values differing by amounts $h\nu$. The energy steps or quanta are called phonons. The energy of optical phonons in germanium has been deduced recently from observations of the scattering of slow neutrons by lattice vibrations [20]. This energy is 0.034 ev, corresponding to a characteristic temperature of 400°K. The value for silicon has been estimated as lying within 10% of 1300°K [20a].

For a crystal in thermal equilibrium the average number of phonons in a normal mode having frequency ν is given by Planck's law

$$n = 1/(e^{h\nu/kT} - 1). \tag{3}$$

Thus around room temperature, for example, the average number of optical phonons will be small, while acoustical phonons—those of long wavelength—will be numerous. For the latter case $h\nu$ is sufficiently less than kT that the right-hand side of (3) can be approximated by $kT/h\nu$ at all but extremely low temperatures.

A conduction electron or hole can interact with, *i.e.*, be scattered by, any of the normal modes described. In doing so, the carrier exchanges energy by emitting or absorbing one or perhaps several phonons. It generally is assumed that the one-phonon process is by far the most probable, and we shall consider only this case. The size of phonons with which the electron can interact is determined by two conditions: 1) the difference between the final and initial energy of the electron must equal the phonon energy and 2) the difference between the final and initial P's of the electron must equal the phonon crystal momentum [20b]. The first condition, of course, is conservation of energy. The second is referred to as "conservation of momentum." The momentum of a phonon is h/λ. Consider the result of applying these two conditions to the case of acoustical modes interacting with carriers of thermal energy around room temperature. For the simple model and valence band structures such carriers have small P. For this case, the energy of a phonon with a given momentum is much less than the energy of a carrier having this momentum would be. As a result, the conservation conditions dictate that the magnitude of P and the carrier energy cannot change a great deal in a transition. Thus the momentum of the phonons with which a carrier of small momentum interacts must also be small and their wavelength large. In the conduction bands of germanium and silicon, as discussed in the last section, thermal electrons have large P. Conservation of energy and momentum permit the long-wavelength acoustical phonons to interact also with electrons of large P. It is apparent that here too they will not cause large changes in electron energy or momentum. After scattering by such a phonon, the electron will remain within the same valley.

The more numerous the phonons of a given mode, the more frequently electrons are scattered by it. In more exact terms the mean free time τ between scattering processes is proportional to $1/n$. Since $\mu = e/m$ times a suitable weighted average of τ over the electrons [7], this leads to a $1/T$ dependence of μ for the acoustical modes and an $(e^{h\nu/kT} - 1)$ dependence for the optical modes. There is an additional factor $T^{-1/2}$ in the mobility for acoustical modes, and approximately that for optical modes, which comes from the density of final states into which the electron can be scattered. Thus the temperature dependence of mobility due to optical-mode scattering is much steeper than that due to acoustical-mode scattering. It has been shown that combination of acoustical-mode scattering and optical-mode scattering in what seems to be reasonable proportions can lead to temperature dependences of the sort listed in Table II. In fact, this is considered now to be the most likely explanation for the temperature dependence of lattice mobility for p-germanium [21]. It may also be the explanation for the cases of n-germanium and p-silicon. This is not, however, the explanation for the case of n-silicon because the interaction of electrons in silicon with optical modes has been shown to be too small for them to affect the mobility appreciably [22]. The mechanism that has been suggested to account for the temperature dependence observed in n silicon is absorption of phonons to give scattering from one valley to another. Because of the conservation of momentum condition, the phonon required to do this would have to possess relatively large energy, comparable to that of an optical phonon, and the same type of argument as that just presented for the optical modes would then be applicable. The same process may also be operative in n-germanium.

One consequence of these considerations worth noting is that the temperature dependence in Table II does not necessarily persist above or below the range specified. For p-germanium, for example, if the steep slope of μ_L vs T is in fact due to optical modes, it must

start changing shortly beyond the range of the table. As the temperature goes below this range, the degree to which the optical modes are excited becomes so small that they cannot affect the mobility very greatly. As the temperature goes above this range, kT becomes comparable to $h\nu$ of the optical modes and n no longer varies so steeply with temperature. Thus, below about 60°K and above about 500°K, lattice mobility would again be proportional to $T^{-3/2}$. Change in slope of lattice mobility vs temperature is expected on the basis of any of the mechanisms proposed to explain the steep slope.

Hall mobility differs from the other types in that it reflects the details of band structure and scattering processes as well as the actual mobilities of the carriers. Since the measurements are relatively simple and convenient to perform, it is worthwhile to go into some detail as to what a Hall experiment actually measures. For the simple model of the band structure the Hall coefficient, R, is proportional to $1/ne$. Since $\sigma = ne\mu$, it is clear that the constant of proportionality is $R\sigma/\mu$, or μ_H/μ, and it is referred to as such in the following discussion. For low magnetic fields R is independent of magnetic field strength, and therefore μ_H/μ is also. Its value depends, however, on the scattering mechanism, specifically on the velocity dependence of the mean free time between scattering processes [7]. For scattering by acoustical modes, on the simple model, $\mu_H/\mu = 3\pi/8$ or 1.18. The value should be about the same, specifically between 1.18 and 1, for scattering by optical modes or a combination of acoustical and optical modes because the velocity dependence of the mean free time is not too different. If impurity scattering is added to the lattice scattering in increasing amounts, μ_H/μ first decreases to approximately unity, then increases. The maximum value is attained in the limit of all impurity scattering, and it depends on temperature. From about liquid nitrogen temperature on up the value of this maximum has been estimated as 1.7. At lower temperatures it is less [23].

If the band structure is such that the constant-energy surfaces are not spherical, the values of μ_H/μ are different from those given above. It is plausible that the shape of the constant-energy surfaces would affect the values of μ_H/μ because a magnetic field causes a rotation of the carriers around these surfaces. Apparently, departure from spherical shape tends to decrease μ_H/μ, and might even result in a value below unity [7]. For ellipsoidal surfaces Herring has shown that μ_H/μ is obtained by multiplying the value for the simple model by a factor that involves the shape of the ellipsoids, this factor having a maximum value of unity. For the values of m_l and m_t determined by cyclotron resonance for electrons in germanium this factor is 0.784 [5]. This leads to a μ_H/μ of 0.93 for acoustical mode scattering in n-type germanium, which is within the range obtained using the figures of Table I.

For holes in germanium, μ_H/μ in Table I is about 1.8 when μ_H is measured at 3000 oersteds. This is much higher than the value expected for lattice mobility on the simple model. Here, however, we must take into account the fact that there are two kinds of holes, heavy and light. As an approximation to the low-field Hall constant, it is reasonable to use an expression analogous to that for intrinsic material based on the simple model. This is

$$R = \frac{1}{e} \frac{n_1\mu_{1H}\mu_1 + n_2\mu_{2H}\mu_2}{(n_1\mu_1 + n_2\mu_2)^2} . \tag{4}$$

Thus in this case the Hall constant reflects the mobilities as well as the numbers of heavy and light holes, respectively. Corresponding to this,

$$\frac{\mu_H}{\mu} = \left(\frac{n_1\mu_{1H}\mu_1 + n_2\mu_{2H}\mu_2}{n_1 + n_2}\right) \Big/ \left(\frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}\right)^2 . \tag{5}$$

If we insert in this expression the values obtained in the cyclotron resonance experiment, $n_2/n_1 = 1/20$, $\mu_1/\mu_2 = 1/8$ [24] and also assume $\mu_{H_1}/\mu_1 = \mu_{H_2}/\mu_2 = 1.18$, we obtain $\mu_H/\mu = 2.7$. Thus a relatively small number of high-mobility holes can boost the μ_H/μ ratio considerably [25]. The difference between 2.7 and the experimental value 1.8 may arise from the fact that warping or non-spherical shape of the surfaces has been neglected in this treatment. It also is possible that values of n_2/n_1 and μ_1/μ_2 are somewhat different at 300 degrees from the cyclotron resonance values which are, of course, obtained at 4°K. A small difference in these values would have a considerable effect on μ_H/μ.

According to Table I the μ_H/μ ratio of 1.8 is obtained with a magnetic field of 3000 oersteds. If a high-resistivity p-germanium sample is measured at a constant field strength of 3000 oersteds, R is found to go down by a factor of about 1.6 as the temperature goes from 300°K down to 78. Certainly this does not reflect any change in carrier concentration. (It would imply an increase in carrier concentration with decreasing temperature!) As a result of this decrease in R, μ_H is proportional to $T^{-1.8}$ in this range, while μ is proportional to $T^{-2.33}$. It therefore does not appear possible to explain this decrease as resulting from a change in n_2/n_1, μ_1/μ_2, or the scattering mechanism.

The explanation of this decrease was provided by Harman, Willardson, and Beer [25]. The formulas for R and μ_H/μ used so far in this section are all for the case of low magnetic field; they are no longer valid at such fields that R becomes a function of H. The magnetic field strength at which this occurs depends on the mobility of the carrier involved. The higher the mobility, the lower the field for which the low-field approximation breaks down. Quantitatively, the criterion for the low-field approximation to be valid within a few per cent is that the quantity $(9\pi/16)(\mu H)^2 \leq 0.01$ [25]. In high-resistivity p-germanium at 78°K the Hall constant is already changing with magnetic field intensity at 100 oersteds [25]. This change would not be expected for a

case of one hole with a mobility equal to the measured drift or conductivity mobility. It is due to the very high mobility of the light holes.

In the limiting case of very high magnetic fields the Hall constant again attains a steady value. This turns out to be a much simpler situation than that for low magnetic fields; the steady value is just $1/ne$, where n represents the total carrier concentration, independent both of scattering mechanism and details of band structure [26]. Thus in the high-field limit $\mu_H/\mu = 1$. The field strength required for the high-field formulation to be valid within a few per cent is determined by the condition $(9\pi/16)(\mu H)^2 \geq 25$.

It is possible now to explain the decrease in μ_H/μ observed with decreasing temperature. At room temperature low-field conditions should still be valid, at least approximately, for both heavy and light holes. This already has been assumed in the treatment of μ_H/μ for this case. At 78°K in a high-resistivity sample, mobilities of both holes are much higher, and at 3000 oersteds the high-field approximation is quite good for the light holes, and not bad for the heavy ones. In agreement with this, the experimental value of μ_H/μ at 78°K is 1.1. The decrease in μ_H/μ with decreasing temperature at 3000 oersteds then is due to the transition from low-field to high-field conditions, which transition essentially knocks out the effect of the light holes. This points up the necessity of study of the magnetic field dependence of the Hall constant.

It might be noted that for electrons in germanium, whose mobility at 78°K is about equal to that of the holes, 3000 oersteds also must be in or close to the high-field region. However, in this case the low-field μ_H/μ is already so close to unity that transition from low to high field conditions makes little difference.

In p-silicon, where probably there are three types of holes at room temperature, μ_H/μ in Table I is less than unity. This is the low-field value because no change is observed as the field is increased from a very low value to 13,000 oersteds [27]. The lack of change is consistent with low hole mobilities. The fact that μ_H/μ is less than unity may be due to the warping of the constant-energy surfaces. Below room temperature the low-field μ_H/μ increases slowly, approaching unity around 78°K [27].

It may be helpful to the reader to have the foregoing information concerning μ_H/μ summarized briefly. We shall take the case of weak fields first. For the simple model, *i.e.*, spherical constant-energy surfaces and only one type of carrier, and acoustical lattice scattering, $\mu_H/\mu = 1.18$. The presence of impurity scattering can decrease μ_H/μ to approximately unity, or increase it to as much as 1.7. The effect of there being two types of carriers (of the same sign) is to increase μ_H/μ. In the limit of strong magnetic fields $\mu_H/\mu = 1$, independent of details of band structure or scattering mechanism. The higher the mobility of the carrier, the smaller the magnetic field strength for which this limit is realized.

Before leaving the discussion of mobility we shall consider briefly its variation with ionized impurity content. Figs. 3 and 4 plot drift and conductivity mobility data from several sources, and a theoretical curve. The drift mobility data were taken directly from Ludwig and Watters [14] and Prince [28]. Conductivity mobility data were taken from Backenstoss [29] and Horn [30] (as quoted by Carlson [30]), and corrected by Smits as will be described. With the exception of the purest samples, the data are all subject to some degree of uncertainty, which in some cases is large. For the conductivity mobility the uncertainty arises from the fact that the data obtained experimentally are resistivity and total concentration of the dominant impurity. (The samples under discussion here are sufficiently impure that minority impurity concentration can be neglected.) In very pure silicon, and in very impure (above about 10^{19} impurities per cm^3) silicon doped with column III or V impurities other than indium, ionization will be complete at 300°K. For the very impure samples this is, in fact, true at all temperatures because the activation energy is reduced to zero. At intermediate concentrations the degree of ionization of the

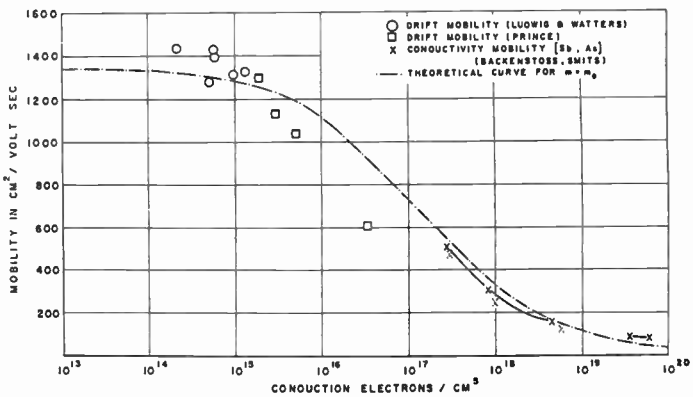


Fig. 3—Mobility vs carrier concentration for electrons in silicon at 300°K. Continuous and dashed lines, as well as points, represent experimental data. Data indicated by broken lines have not been corrected for ionization being incomplete, and require such correction. The corrected points are indicated by solid lines, and are displaced a small amount to the left and upward.

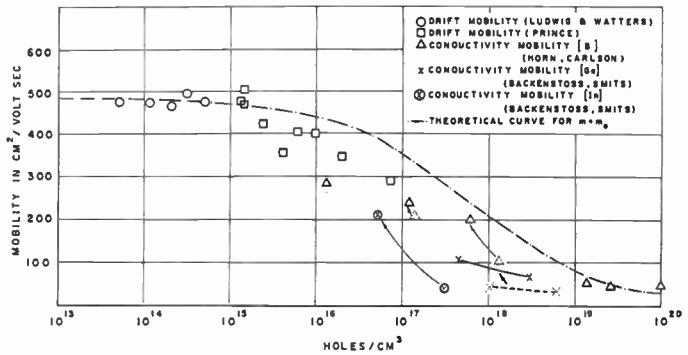


Fig. 4—Mobility vs carrier concentration for holes in silicon at 300°K. Continuous and dashed lines as well as points represent experimental data. Data indicated by broken lines have not been corrected for ionization being incomplete, and require such correction. The shift due to correction is indicated by an arrow.

impurities must be calculated theoretically in order to obtain carrier concentration. The theoretical formulas that relate carrier concentration to total impurity content, temperature, etc., have been revised to take into account the peculiarities of the band structure [31], and appear to work quite well for lightly doped material [31]. Unfortunately, however, there are many uncertainties in applying this theory to more heavily doped material. Still, there is as yet no better theory; consequently the one we have was used in calculating Figs. 3 and 4 [32]. The points that would be obtained if carrier concentration were equal to total impurity concentration are shown in broken lines, the corrected points in solid lines. For n-type material the points are shifted only a small amount, so the uncertainty and the error are undoubtedly small. For p-type material, however, this is obviously not the case. The one point included for indium-doped material, in particular, must be considered quite speculative. Over-all, the inaccuracy of this correction may be an important source of the scatter in the p-type conductivity mobility data.

The drift mobility data, with the exception of those for the purest samples, also contain uncertainties. What is measured in these experiments is, of course, the drift mobility of the minority carriers. To convert this into drift mobility of majority carriers for samples where impurity scattering cannot be neglected requires knowledge of both total impurity content and impurity mobility. Here again existing theory is not quite satisfactory. The theoretical impurity scattering formula, given in I, and the formula for combining mobilities plotted in Fig. 2 of I, are based on the simple model of the band structure and have not been revised for the actual band structures. More serious than this, however, is the fact that the theory is not expected to be valid for high impurity concentrations. Here again, however, this is the only theory available and it was used also to calculate the theoretical curves plotted in Figs. 3 and 4. In detail, these were obtained by combining the lattice mobilities in Table I with the theoretical impurity scattering formula as described in I. For both holes and electrons an effective mass of m_0 was used. As is seen, this value gives a theoretical curve that fits the experimental data well. It might be noted that, in the case of electrons, the Ludwig and Watters data lie above the theoretical curve, indicating that a small correction for impurity scattering should have been made for their n-type samples.

A word of caution might also be inserted here against extrapolation of the usual formulas to low temperature. It has been found, for example, that the dependence of mobility on temperature in fairly impure samples can become steeper than the $T^{3/2}$ predicted for impurity scattering. Also, at low temperatures where impurity scattering would be expected to dominate, the mobility is found in some instances to be larger in more impure samples. These effects have been shown to be related to the onset of impurity-band conduction [33]. As this name implies, electrons in the donor states and holes in the acceptor states can still contribute to the conductivity without being excited to the conduction or valence bands. As the temperature goes down and carriers drop out of the conduction or valence bands, decreasing the contribution of these bands to the conductivity, a temperature is finally reached at which the conduction in the impurity states becomes comparable in magnitude. Conductivity and Hall effect must then be represented by formulas similar to (1) and (4), where 1 and 2 now represent carriers in the conduction band (or valence band) and carriers in the impurity band, respectively. Care must be taken in the interpretation of experimental data in this temperature range [33]. At still lower temperatures, where there are even fewer carriers in the conduction band, conduction in impurity states dominates completely. For very low impurity concentrations this conduction has quite different characteristics from those of the usual conduction process [33].

A discussion of magnetoresistance is beyond the scope of this paper. It should be mentioned, however, that knowledge of the band structure has made it possible to explain, at least qualitatively, the features found anomalous before. The existence of large magnetoresistance in n-type material for a magnetic field parallel to the current is one of these features [34]. The relatively large size of the transverse magnetoresistance and anomalous temperature dependence in p-type material has been explained by taking into account the light holes with their larger mobility [25].

Infrared Absorption and Emission

Infrared or optical absorption in semiconductors is usefully divided into four types: 1) "fundamental" absorption, associated with transitions of the electrons across the energy gap; 2) absorption of free carriers; 3) absorption associated with transitions to and from energy levels due to impurities or defects, and 4) absorption due to excitation of lattice vibrations. Considerable work has been done on silicon and germanium in all four categories. This discussion is confined mainly, however, to the first two [35].

For a transition across the energy gap, an electron absorbs a photon. Compared to the electron's energy and momentum, the photon's energy is large and its momentum is small. Conservation of energy and momentum applied to the transition predict that P of the electron can scarcely change at all in the process. In other words, the transitions, depicted in a band structure diagram such as Fig. 1, must be vertical. Thus for germanium and silicon the smallest photon or the lowest energy absorbed should correspond to a transition from $P=0$ in the valence band to $P=0$ in the conduction band. When an electron in the valence band makes a transition to the conduction band under thermal excitation, no such selection rule operates, and it can go directly from the top of the valence band to the minimum energy in the conduction band. The energy required to do this is determined by measurements of intrinsic conductivity, for example. Thus, it is expected

that the energy gaps of germanium and silicon determined by the lowest energy for which optical absorption is observed be larger than the so-called thermal energy gaps. Experimentally, this is found not to be the case. Apparently, the selection rules for the optical transitions are not operative.

The explanation that was proposed for this is the following. Momentum can be conserved in the transition if the electron also emits or absorbs a phonon with momentum equal to the momentum difference between $P=0$ and the band edge. Since the phonon energy is small compared to the photon energy, the minimum photon energy absorbed would still be closely equal to the thermal gap. The probability of such a transition, called an indirect transition, is less than that for the vertical or direct transition which involves the photon only. It is still sufficient, however, to cause easily observable absorption as was shown by a calculation of Bardeen, Blatt and Hall [36].

These ideas concerning direct and indirect transitions are borne out by the shape of the experimental curves of absorption vs photon energy shown in Fig. 5. For germanium the absorption starts a little above 0.6 ev, rises to a plateau or knee just above 0.7 ev, and then starts another abrupt rise at about 0.8 ev. For silicon the curve shows the same type of behavior with the second rise starting just above 2 ev. The first and second rises have been attributed to the onset of indirect and direct transitions, respectively. From this type of data it is concluded that the energy at $P=0$ is higher than that at the band edge by 0.18 ev in germanium, 1.5 ev in silicon.

The region of very small absorption was studied under high resolution for germanium by MacFarlane *et al.* [37] over a wide temperature range. They found a fine structure, not shown in Fig. 5, with one or more knees in the absorption. Analysis of this structure led them to conclude that the indirect transitions take place with the cooperation of either of two phonons of different energy, both, of course, having momentum equal to the momentum difference between $P=0$ and the band edge. The two phonons are identified as belonging to the longitudinal and transverse acoustical branches of the phonon spectrum, respectively. The two possible transverse vibrations have the same energy in the direction of P concerned. The energies of these two phonons correspond to characteristic temperatures of 90°K and 320°K. From this analysis MacFarlane *et al.* also obtain values for the energy gap (from $P=0$ in the valence band to the conduction band minimum) in the range 4°K to 300°K. The latter value is listed in Table III. The value of the gap at 0°K obtained by extrapolating their data also is set forth in Table III. It is noteworthy that they found the variation of the gap with temperature to be linear from 150°K to 300°K, but not below this.

The inverse process to fundamental absorption, recombination of an electron and hole with emission of a photon, has been observed also. To do this requires a high density of excess electrons and holes, since recombination via a level in the forbidden gap is a very effective competing process. In the case of germanium two peaks have been observed in the recombination radiation, with maxima at 0.81 and 0.70 ev [38]. These then correspond to direct and indirect transitions, *i.e.*, recombination without and with phonon cooperation, respectively. For silicon, by studying the fine structure of the intrinsic recombination radiation, Haynes *et al.* [39] have obtained results of the same nature as those obtained by MacFarlane *et al.* in studying the absorption properties of germanium. They found that four different energy phonons can cooperate in the transitions, one from each of the branches of the phonon spectrum: transverse optical, longitudinal optical, transverse acoustical, and longitudinal acoustical. From the energies obtained for these phonons their momenta have been estimated. It is this estimate which leads to the result that the conduction band minima in silicon are located at a P about 0.8 of the distance from the center to the edge of the Brillouin Zone. The values of the energy gap in silicon deduced from Haynes' data are in Table III.

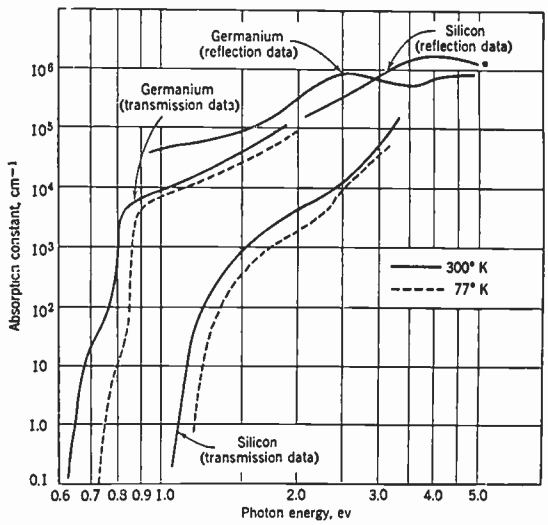


Fig. 5—Intrinsic absorption spectra of silicon and germanium. (After Burstein, Picus, and Sclar [35]. The transmission curves are due to Fan, Shepard, and Spitzer [105], and to Dash, Newman, and Taft, [106]. The reflection curve for germanium is based on the data of Avery and Clegg [107] and the reflection curve for silicon on Pfestorf [108].

By absorbing a photon, free electrons and holes can make a transition from one level to another in the conduction band and valence band, respectively. This absorption can take place for photon energies much less than the energy gap, in fact, for almost zero photon energy because of the close spacing of the energy levels. Theoretically, the absorption can be expected to increase uniformly as the square of the photon wavelength, and show no structure [40]. Behavior fitting this theoretical prediction is found beyond the fundamental absorption edge in silicon, and in n-type germanium. Something different is found, however, in p-type germanium.

Infrared absorption for a heavily doped p-type sample is seen in Fig. 6. At wavelengths longer than the

TABLE III
SOME PROPERTIES OF INTRINSIC MATERIAL

Quantity	Value	
	Germanium	Silicon
Energy gap at 0°K	0.75 ev [37]	1.153 ev [39]
Energy gap at 300°K	0.67 ev [37]	1.106 ev [39]
np	$3.1\times10^{32}\exp(-0.785/kT)$ [45]	$1.5\times10^{33}T^3\exp(-1.21/kT)$ [46]
n_i at 300°K	2.4×10^{13}/cm³	1.5×10^{10}/cm³
ρ_i at 300°K	46 ohm-cm	2.3×10^{5} ohm-cm
$m_n^{(N)}$ at 4°K	$0.55\ m_0$	$1.1\ m_0$
$m_p^{(N)}$ at 4°K	$0.37\ m_0$ [44]	$0.59\ m_0$ [44]

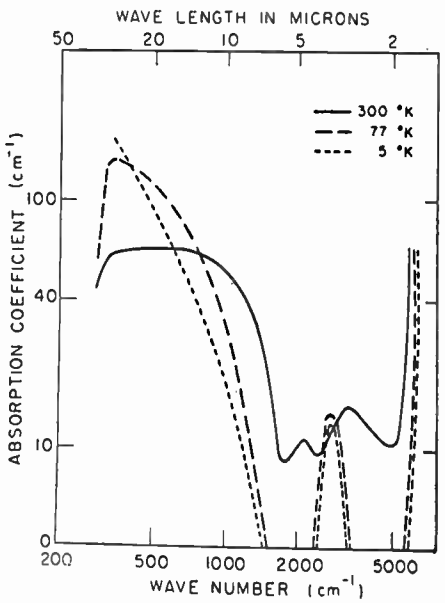


Fig. 6—Absorption coefficient for a p-type germanium sample with room temperature resistivity 0.07 ohm-cm. After Kaiser, Collins, and Fan [41].

fundamental absorption edge, which is at about 1.8 microns, there are three peaks for the 300-degree data. At the lower temperatures two of these have merged. It was established experimentally that this absorption is due to the free holes rather than to any impurity levels [41]. This structure can be explained as being due to transitions among the three valence bands shown in Fig. 1 [42]. The longest wavelength peak is due to transitions from band 1 to band 2, the next longest to transitions from 2 to 3, the shortest from 1 to 3. This provides strong evidence for the correctness of the valence band structure in Fig. 1. In fact, it is the position of these absorption peaks rather than a direct calculation of the spin-orbit coupling that has determined the ΔE values in Fig. 1. In the case of silicon, the value given there reflects the fact that the hunt for similar structure has been carried out to 25 microns, or 0.05 ev, without success.

Some Properties of Intrinsic Material

The number of electrons in a given small energy range in the conduction band at a given temperature depends on two factors: 1) the Fermi-Dirac distribution function, which gives the probability of the electron having this energy, and 2) the number of available states in the conduction band in this energy range. The total number of electrons in the conduction band is then given by an integration over energy of the product of these two factors. For the simple model of the band structure and a small enough number of electrons so that Maxwell-Boltzmann statistics can be used, this yields the familiar expression

$$n = 2\left(\frac{2\pi m_n kT}{h^2}\right)^{3/2} e^{-(E_C-E_F)/kT} \tag{6}$$

where E_C-E_F is the energy interval between the edge of the conduction band and the Fermi level [7], and m_n is the effective mass of the electron. This expression can be generalized simply for arbitrary band structure by replacing m_n by $m_n^{(N)}$, the appropriate value for the particular band structure. The general expression is then

$$n = 2\left(\frac{2\pi m_n^{(N)} kT}{h^2}\right)^{3/2} e^{-(E_C-E_F)/kT} \tag{7}$$

The quantity $m_n^{(N)}$ defined in this way is called the density-of-states mass for the conduction band [43]. If there were a single ellipsoidal constant-energy surface, $m_n^{(N)}$ would equal the geometric mean of the three principal effective masses. This type of dependence arises because the number of available states depends on the volume of momentum space available, *i.e.*, on the volume of the constant-energy ellipsoids. For a structure like that of the conduction band of germanium and silicon, where there are g_c ellipsoids for any one energy, each ellipsoid contributes equally to the number of states, and

$$(m_n^{(N)})^{3/2} = g_c(m_l m_t^2)^{1/2}. \tag{8}$$

The values of $m_n^{(N)}$ deduced from cyclotron resonance results [4] are in Table III. The number of valleys, g_c, has been taken as 4 for germanium, 6 for silicon.

An expression similar to (7) can be written for the valence band.

$$p = 2\left(\frac{2\pi m_p^{(N)} kT}{h^2}\right)^{3/2} e^{(E_V-E_F)/kT} \tag{9}$$

where E_V-E_F is the energy interval between the edge of the valence band and the Fermi level, and $m_p^{(N)}$ the density-of-states mass for the valence band. In the temperature range where holes occupy only the two top bands, the contributions to p of the two bands can

simply be added. This gives

$$(m_p^{(N)})^{3/2} = (m_{p_1}^{(N)})^{3/2} + (m_{p_2}^{(N)})^{3/2}. \quad (10)$$

Taking into account by a perturbation method the warping of the bands, Lax and Mavroides [44] have calculated from cyclotron resonance data for germanium $m_p^{(N)} = 0.37\ m_0$, for silicon $m_p^{(N)} = 0.59\ m_0$.

From (7) and (9) the theoretical product of n and p is given by

$$np = n_i^2 = 2.33 \times 10^{31} \left(\frac{m_n^{(N)} m_p^{(N)}}{m_0^2} \right)^{3/2} T^3 e^{-E_G/kT} \quad (11)$$

where E_G is the value of the energy gap at the temperature T. The experimental values that have been obtained for n_i^2 in germanium and silicon are in Table III. These are valid in the range 300–500°K for germanium and 400°K–700°K for silicon. Actually, the quantity measured was the intrinsic conductivity, and n_i was computed from this by using lattice mobility values extrapolated from the lower temperature range. As noted previously, this probably gives some error, but it should not be large. In the case of silicon, unfortunately, the samples used were sufficiently impure so that a correction had to be made for impurity scattering, and the lattice mobilities deduced and used for the calculation of n_i do not agree with those in Table II. This would introduce additional error.

It has been customary to compare the empirical np product with (11) in order to obtain $(m_n^{(N)} m_p^{(N)}/m_0^2)$, or information about the energy gap. To deduce the former it is necessary to know how E_G varies with T. It has been usual to assume that $E_G(T) = E_G(0) + \beta T$. If this were correct, 0.785 and 1.21 (see Table III) would represent the gaps at 0°K in germanium and silicon, respectively. As mentioned in the last section, however, although the gap does vary linearly with temperature over a range, this does not hold to 0°K. The best values of the gap at 0°K are believed to be those obtained by optical methods, as described in the last section, and these are listed in Table III. Values of the energy gap as a function of temperature, deduced from optical data, have been used with the empirical np product by MacFarlane *et al.* [37] to obtain information about the mass factor for germanium. It was their conclusion that the number of minima in the conduction band is four [47], and that $(m_n^{(N)} m_p^{(N)}/m_0^2)^{1/2}$ at 300°K has increased about 20 per cent from the cyclotron resonance (4°K) value.

Values of the intrinsic resistivity at room temperature calculated from n_i and the μ values in Table I are also included in Table III.

Zener Current

A typical current-voltage characteristic of a p-n junction biased in the reverse direction (for example, see Fig. 8 of I) can be divided into two regions. At the low voltages there is an approximately constant current called the saturation current. This can originate from minority carriers created by thermal generation either at traps within the space-charge region [48], or within the n- and p-regions, or possibly at the surface. In the two latter cases the carriers must be generated close enough to the space-charge region that they can get there by diffusion within a lifetime, since the fields outside the space-charge region are small. Once at or in the junction the holes are swept over by the field to the p-region and the electrons to the n-region, making up the saturation current. This is also called the thermal generation current. At sufficiently high voltage a sort of breakdown occurs and the current rises steeply, almost vertically with voltage. These large currents have been called "Zener currents" [49] from the idea that they are due to a mechanism proposed by Zener, actually in another connection. According to this idea the steep increase is due to internal field emission, or direct excitation of electrons from valence band to conduction band by the high field [2]. Another possible mechanism to explain the steep increase is impact ionization. On this mechanism conduction electrons or holes gain sufficient energy in the high-field region of the junction to free valence electrons, producing electron-hole pairs. These in turn can create additional electron-hole pairs, Breakdown occurs when the charge multiplication becomes very large. The situation is analogous to that of avalanche breakdown in gases and lends itself to similar theoretical treatment.

A practical way of distinguishing between the two mechanisms experimentally is the following. In the avalanche case, multiplication of charge injected into the junction would take place at voltages below the breakdown voltage, while in the Zener case it would not. Multiplication of injected charge carriers before breakdown was not found in the early investigations [49] on germanium junctions, and it was for this reason that the avalanche mechanism was rejected. Subsequent investigations by McKay and McAfee [50] did reveal multiplication in p-n junctions in both germanium and silicon. Small changes in minority carrier current, produced either by illumination or by α-particle bombardment, were shown to be multiplied in passing through the junction. These authors pointed out the possibility of both mechanisms occurring in germanium and silicon, the Zener mechanism accounting for breakdown in the cases where the junction was too narrow for an electron or hole to make a number of ionizing collisions, the avalanche mechanism accounting for breakdown in wide junctions. There is now considerable evidence that this is the actual situation. The avalanche process has been found and studied in a wide range of junctions in silicon [51] and germanium [52]. The junctions investigated included grown, diffused, and alloyed types, step junctions, and ones with linear impurity gradient. The breakdown voltage (*i.e.*, the voltage at which the steep increase in current occurs) ranged from about 8 to 250 volts for the silicon junctions, 6 to 150 volts for those in germanium. Associated with the avalanche breakdown there was found to be a characteristic type of electrical noise [51], and in silicon emission of visible light was

observed [53], [54]. It has been shown that the light is emitted from highly localized regions where avalanche breakdown is taking place. A study of the spectral distribution indicates that the light arises both from radiative recombination of electrons and holes freed by the avalanche, and intraband transitions of highly energetic carriers. The fact that breakdown takes place in spots or localized regions suggests that nonuniformities of some kind are important. By comparison of the light emission patterns and etch pit patterns that reveal dislocations, it has been shown that the avalanche breakdown takes place preferentially where dislocations pass through a junction [55]. The noise that accompanies the onset of breakdown has been attributed to instability of the breakdown [51]. At a value of voltage across the junction that just causes breakdown in some particular region the region carries current intermittently, *i.e.*, it is continuously switching in and out of the breakdown condition. At higher voltages the breakdown in this particular region becomes stable, but regions for which the breakdown voltage has just been reached are in the unstable condition.

Very narrow junctions formed in highly doped silicon by diffusion were studied by Chynoweth and McKay [56]. They estimated that the junction widths were as little as 400 Angstroms. The reverse currents in such junctions are not sharply divided into an approximately constant saturation-current region at low voltage and a steep rise at the breakdown voltage; typically they start curving upward at almost zero voltage and make a smooth transition into a more rapid rise. A current-voltage characteristic of this type is called "soft," while one like that of Fig. 8 in I is called "hard." The effects associated with breakdown in these junctions are rather different from those previously described. Light emission is observed, but it is more uniform over the area of the junction. No noise was found to be associated with the onset of breakdown. Also, the change with temperature of the current-voltage characteristic is different from what is observed for the avalanche breakdown. From such facts and others, it was concluded that the Zener effect is the cause of breakdown in these very narrow junctions [56]. It was also demonstrated that Zener emission is taking place in these junctions at voltages below the breakdown voltage, in fact from zero voltage on up. This is indicated by the fact that a large decrease in temperature resulted in a relatively small change in the magnitude of the low-voltage reverse currents. This indicates that the carriers responsible for the current flow at voltages below breakdown are not thermally generated, but rather are due to Zener emission. It was concluded that the built-in field in these junctions is sufficient to cause Zener emission even at zero applied bias. The softness of the current-voltage characteristic of these junctions has been attributed to the nonuniformity of the field in the junction [56], and to space-charge effects, among other things. The latter, of course, are less important in avalanche breakdown where both electrons and holes are present. Nonuniformity of the field in the junction makes it impossible to determine with any certainty the minimum field necessary for Zener emission. Chynoweth and McKay speculated that it may be as low as 6×10^5 volts/cm.

Even in the very narrow junctions where Zener emission is dominant, some multiplication of injected carriers is observed at reverse potentials of a couple of volts. Application of reverse voltage widens the junction, of course, and increases the probable number of ionizing collisions within the junction. It therefore would be expected that multiplication of the Zener-emitted electrons would occur at high enough reverse voltage across the junction, and evidence for this has been found [56].

From the work of Chynoweth and McKay, the following picture emerges for silicon. For junctions with a reverse breakdown voltage below about 7 volts the process that mainly determines the reverse characteristic is Zener emission. For junctions with a reverse breakdown voltage above about 40 volts, it is the avalanche process that causes breakdown, specifically by multiplication of the thermally excited carriers which constitute, at lower voltages, the saturation current. For junctions with breakdown voltages in between 7 and 40 volts it is common to find that both processes are important in determining the reverse characteristic. Typically, the Zener emission may dominate at low reverse voltage, the current rising slowly with voltage. Then at some value of voltage the current increases rapidly with voltage, leading into the sharp vertical rise characteristic of the avalanche process. In this case, multiplication of the Zener-emitted electrons as well as of the thermally generated carriers is occurring.

The occurrence of Zener emission in very narrow germanium junctions has been suggested by Esaki to explain an anomalous current-voltage characteristic found there [57].

Observation of avalanche breakdown has been reported also for homogeneous material, in particular, an n-type germanium specimen [58]. The electric field intensity at which it occurred was 6×10^4 volts/cm. Evidence that the avalanche occurs in the high field of a point contact on germanium also has been presented [58]. A related effect, ionization of neutral donors and acceptors by impact of fast electrons or holes, has been observed [59]. This, of course, can take place only at temperatures low enough so that the impurities are not thermally ionized.

Effects Arising from Impurities and Defects

A wealth of information has been accumulated on effects related to the addition of impurities to germanium and silicon. This includes distribution or segregation coefficients, solubilities, diffusion constants, energy levels provided for electrons in the forbidden gap, capture cross section of these energy levels, etc. The following discussion of effects of impurities is con-

fined mainly to the topic of the energy levels, since these have the most direct influence on the electrical properties of the materials. Information on the first three topics can be obtained in comprehensive review articles by Burton [60], Dunlap [61], and Hannay [62], and frequently in the references given for the energy levels of the different impurities.

It is well known that germanium and silicon specimens can undergo large changes in resistivity and lifetime when they are heated and cooled. These changes present a complex problem because both impurities and lattice defects are influencing factors. Nevertheless, considerable progress has been made in understanding them, and this is summarized in the subsection on heat treatment effects.

Energy Levels in the Forbidden Gap

The current information on energy levels provided by impurities in germanium and silicon is summarized in Figs. 7 and 8, respectively, where the levels are shown in their relative position in the energy gap. The arrangement of elements from left to right more or less follows that in the periodic chart. Donor or acceptor nature, where determined, is indicated in parentheses. Below the center of the gap the numbers indicate energy relative to the valence band; above the center, energy relative to the conduction band. The energies given are characteristic of small concentrations of the impurities. At high enough concentrations these levels broaden, and the bands that arise from the shallow levels may overlap the conduction band, in which case the activation energy is zero.

We shall first consider impurity levels in germanium. Here, despite the wide variety of doping elements on which data are available, the number and type of levels provided by a particular element are given almost entirely by a few simple rules. Basic to these is the type of position the impurity occupies in the lattice. There are two ways in which an impurity atom may go into an otherwise perfect germanium or silicon lattice. It can substitute for a lattice atom, or it can occupy a site not normally occupied by a lattice atom, *i.e.*, an interstitial site. All of the levels shown in Fig. 7 except that of Li and the donor level of Au are explainable on the hypotheses that the impurity goes into the lattice substitutionally, and that there is a tendency for the substitutional atom to form covalent or electron-pair bonds with the four nearest-neighbor lattice atoms just as a germanium (or silicon) atom would do in this position. Germanium (or silicon) has four valence electrons, each going into one of the four bonds. It then follows from the above hypotheses that elements of the third column of the periodic table with three valence electrons—B, Al, Ga, In, Tl—can accept one electron, Zn and Cd in column II provide two acceptor levels, while Cu, Ag, and Au in column I are triple acceptors. On the other side of the periodic chart P, As, Sb, and Bi, with one electron more than is required to complete the bonds, can donate one

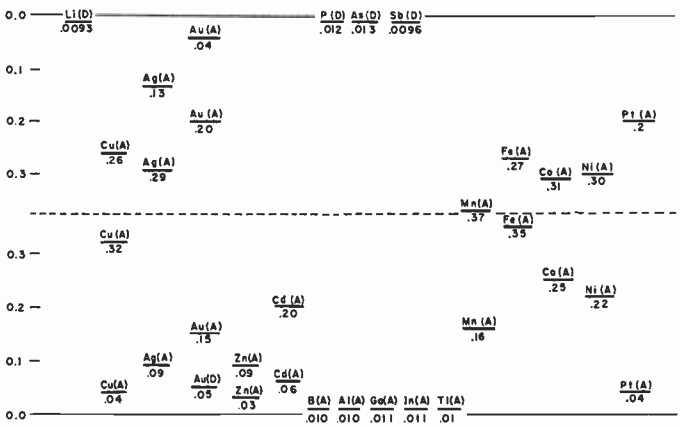


Fig. 7—Energy levels of impurities in germanium. For levels above the dashed line (gap center), energy is measured from the conduction band; for levels below, from the valence band. The sources for this information are as follows: Li [65]; Cu, Au [67] (this paper contains references to many other authors who contributed to this work); Ag [109]; Zn, Pt [110], [111]; Cd [112]; B, Al, Ga, In, P, As, Sb [113]; Ti [61]; and Mn, Fe, Co, Ni [114] (this paper contains references to the work on the fourth row elements other than Mn).

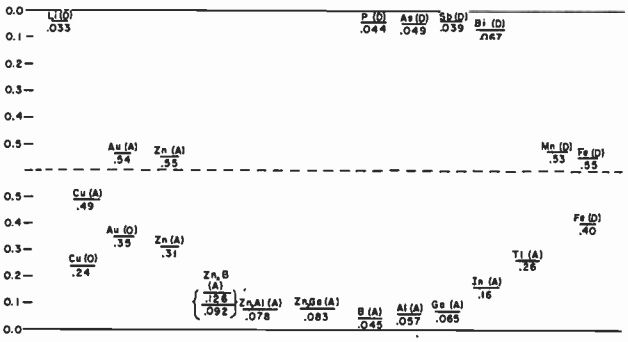


Fig. 8—Energy levels of impurities in silicon. For levels above the dashed line (gap center), energy is measured from the conduction band; for levels below, from the valence band. In cases where the value obtained by optical measurements (long wavelength photoconductivity or optical absorption) differs from that determined by Hall measurements, the latter has been quoted. The sources for this information are as follows: Li, B, Al, Ga, In, P, As, Sb [115]; Cu, Fe [71]; Au [70]; Zn; Zn,B; Zn,Al; Zn,Ga [73]; Zn [116]; Bi—F. J. Morin, private communication quoted in [62]; Tl [77]; and Mn [117].

electron. The transition elements Mn, Fe, Co, Ni, with two 4*s* electrons in the outer shell, act as double acceptors. The incomplete *d* shell apparently plays no part. In the case of platinum, with one 6*s* electron, only two acceptor levels have been reported. On the basis of the hypotheses presented above, a third should be found. The one case involving a substitutional impurity that does not fit these hypotheses is that of the donor level of Au. When the Fermi level is low enough, the Au atom can lose its single valence electron to assume a noble gas configuration.

It may be worth emphasizing a point concerning the meaning of the energy levels for the cases where several acceptor levels belong to a given impurity. This can be done conveniently by using the case of zinc, a double acceptor, for illustration. The lower zinc level is defined by placing an electron on a neutral zinc atom, the upper

zinc level by placing an electron on a zinc ion with a single negative charge. Thus, if a neutral zinc atom were put into the lattice only the lower level would exist and an electron from the conduction band or another impurity atom would have to fall into the lower zinc level. Once this level was occupied, a second electron could not fall into it, but would have to fall into the upper level.

Some of the other column IV elements, *e.g.*, C [61], Si [63], Sn [64], have been found in or deliberately added to germanium in very large concentrations with apparently little electrical effect. The absence of donor or acceptor activity of these elements could be explained by the two hypotheses presented above. In the cases of Si and Sn there is experimental evidence for a substitutional position in the lattice. Experimental evidence on this point however, is lacking for *C*, and this case might be more complicated.

The donor behavior of Li is explainable on the basis of an interstitial position in the lattice. By releasing its one electron, Li attains the noble gas configuration. There is some additional evidence for an interstitial position in the large diffusion rate of Li in germanium and silicon, which suggests interstitial diffusion [65]. It should be pointed out, however, that a given impurity can occupy more than one type of position in the lattice. Copper and nickel, for example, have large diffusion rates but, as has been seen, provide energy levels consistent with a substitutional position in the lattice. To explain this and other features of their diffusion, it has been postulated that these impurities are present at both types of positions and that the relative number in each is a function of temperature [66]. This is discussed further in the next subsection. It might be noted that donor behavior of interstitial copper has been looked for but not found [67].

Hydrogen and oxygen can be present in single crystals of germanium at concentrations as high as 10^{18}/cc [68]. They also do not appear to affect the electrical properties, including lifetime, in any way. Hydrogen, like lithium, may occupy an interstitial position in the lattice. A possible explanation for its inactivity in such a position is taken up later. It is also possible that the hydrogen and oxygen are present in the crystal as molecular H_2O [68]. In the case of oxygen, however, there is evidence that at least some of it is present in the germanium lattice in an interstitial solid solution of the sort discussed below in detail for silicon [69]. This model also predicts no effect of oxygen on the electrical properties.

We consider now impurity levels in silicon. The simple considerations applied previously explain the single acceptor action of the column III elements, and single donor action of the column V elements and lithium. Beyond this, however, they do not do quite as well for silicon as they did for germanium. Zinc is the only element that has been found, up to the present, to show double acceptor action. In gold-doped samples the lowest level is a donor as in germanium, but a search for acceptor levels has revealed only one. These two levels are relatively much higher in the gap than the two lowest gold levels in germanium, and it has been postulated that the other two acceptor levels are shifted so high in silicon as to be in the conduction band, where they could not be detected [70]. Copper is similar to gold in showing one donor and one acceptor level. In this case, however, no search has been reported for additional acceptor levels, presumably because of the low and uncontrollable solubility of copper in silicon. Manganese and iron are donors, rather than acceptors as in germanium. This might result from an interstitial position in the lattice, and there is evidence for this in the rapid diffusion observed for these elements. Diffusion experiments with iron indicate that it may exist in the lattice in two forms with different diffusion rates, reminiscent of the case of copper in germanium. An additional peculiarity of iron-doped silicon is the fact that the 0.40 donor level changes after standing at room temperature for about a month to 0.55 ev [71]. This conversion is not understood.

In the foregoing discussion we have assumed explicitly that the impurity atom went into an otherwise perfect lattice, and implicitly that the individual impurity atoms acted independently of each other. If these assumptions are dropped, the different ways an impurity atom can go into the lattice increase tremendously beyond the two simple possibilities envisaged before. One case of interest here is that of oxygen in silicon, where the position of an oxygen atom may be correlated with those of other oxygen atoms in the lattice. This case is detailed in the next subsection. Another case of interest here is that of a second type of impurity put into a sample already doped with one type of impurity. A system in which these effects can be studied in a straightforward manner is one in which a rapidly diffusing ion is added to a specimen already containing relatively immobile impurity ions [72]. One such system is that of lithium diffused into gallium-doped germanium. Except at the very lowest temperatures, lithium is present as a positive ion while gallium is a negative ion. Coulomb attraction can then cause the more mobile lithium ion to take up an interstitial position adjacent to a gallium ion. Formation of such ion pairs will affect many of the properties of the impurity and the crystal, notably the diffusion rate, mobility of free carriers and energy levels provided for electrons and holes. These and related effects have been studied by Reiss, Fuller, and Morin [72]. The ion pair formed of a donor and acceptor, being a dipole, will have very little attraction for either an electron or hole, and will therefore provide levels very close to the conduction band or valence band, respectively. The original acceptor levels, *e.g.* the gallium levels in the system referred to, should be simply removed by the pairing, and this has been experimentally verified by analysis of Hall data. Another set of systems studied involves zinc, which is also a rapidly

diffusing impurity, combined with various of the column III acceptors in silicon [73]. It is found that introduction of zinc into these materials causes the level of the original acceptor to disappear, while the new shallow levels in Fig. 8 appear. The latter have about the same concentration as the original acceptor levels. These features, combined with determination of donor and acceptor concentrations, have been taken as evidence that the new levels are due to a complex formed between zinc and the original acceptor. In a more detailed model suggested for this complex, zinc is in a lattice site adjacent to that of the negatively-charged acceptor, and bonded to it, as well as to one of the neighboring silicon atoms, covalently [73]. This has a net negative charge and can bind a hole. The formation of ion pairs and triplets between lithium and zinc in germanium has also been studied [74]. In this case pairing of Li^+ and Zn^- appears to result in a single new acceptor level with an activation energy of 0.01 ev. It is clear that there are many more interesting effects arising from interaction of impurities to be explored. An intriguing application suggested for these effects is the use of doping to eliminate undesirable levels, such as trapping levels, from the forbidden gap [72].

We shall now take a more detailed look at the nature of the impurity levels in order to understand in a qualitative way the differences in depth. When a column III or V atom substitutes for a germanium or silicon atom, it must acquire one negative or positive charge to properly complete the valence bonds around it. With this one net charge it can bind a hole or electron, respectively. Because of the large dielectric constant of germanium or silicon, however, the electrostatic attraction is greatly weakened and the hole or electron is on the average several atoms away. It is then for many purposes a fair approximation to think of the system of impurity ion plus bound charge as a hydrogen atom in a homogeneous medium characterized by the dielectric constant K of germanium or silicon and appropriate effective mass for electrons and holes [75]. This model assumes a single, isotropic mass, and so is clearly based on the simple model of the band structure. It is clear, for example, that with an anisotropic effective mass of the type shown for the conduction band in germanium and silicon, the wave function of an electron bound to a donor ion will be anisotropic. In the directions for which the mass is smaller, the average distance from the donor ion will be larger. The hydrogen model predicts an activation or ionization energy equal to that of hydrogen, 13.6 ev, divided by K^2 and multiplied by m/m_0. The factor K^2 arises because the electrostatic attraction is decreased by a factor K, which in turn increases the radius of the orbit by a factor K. What to put in for m/m_0 on this model is not clear because, of course, the model is not correct. Nevertheless, if average masses of some sort are put in, this model predicts an activation energy fairly close to what is observed for the column III and V impurities in germanium and silicon. It cannot be relied on, however, to give correctly such information as intervals between ground state and excited states or degeneracies of states [76].

Differences in activation energy between the different elements of a column arise from the fact that the bound charge, of course, does spend some time in the immediate neighborhood of the impurity ion and therefore is affected to some extent by the properties of the particular ion. These differences are smaller in germanium than in silicon because the larger dielectric constant and smaller effective masses result in larger orbits. In the case of acceptors there is an additional source of differences in that ionization involves substituting a Ge- or Si-acceptor bond for a Ge-Ge or Si-Si bond. The sharp increase in binding energy observed in going from gallium to indium in silicon has been correlated with a similar increase in the bonding energies involved [77].

It is a reasonable guess that a hydrogen model would work well also for lithium in germanium and silicon. Here the valence bonds are not involved, since lithium is interstitial, but of course the lithium atom already has the structure of one electron attracted by a core with a positive charge. Thus it is understandable that lithium should have approximately the same ionization energy as the column V donors. It should be remembered, however, that an interstitial position is presumed to be correlated with electrical inactivity for hydrogen, which of course has similar structure. To explain this difference it has been suggested that the small size of hydrogen enables it to fit into an interstitial cavity without experiencing the effect of the high dielectric constant [78].

Elements from other columns of the periodic chart in general provide more than one energy level, and thus acquire more than one charge. This greater net charge means that a bound hole or electron spends much more time near the parent ion with two consequences: deeper-lying levels or greater activation energies, and greater differences between different impurities, even when they are in the same column of the periodic chart. It is satisfying, however, that silver, which lies between copper and gold in column I of the periodic chart, provides levels intermediate between those of copper and gold. Unlike the case of the hydrogen-like impurities, almost no theoretical work has been reported that would explain in detail the positions of any of these deeper-lying levels.

Effects of Heat Treatment

Changes in the electrical properties of a sample on heat treatment can result from changes in impurity concentration or state of dispersion of impurities. They also can result from changes in concentration or state of dispersion of defects in the lattice other than impurities. The basic defects are missing lattice atoms or vacancies, interstitial lattice atoms, and dislocations. Changes in the number and distribution of dislocations can take place during heat treatment, particularly if plastic de-

formation is allowed to occur. There is no reason to believe, however, that dislocations are of primary importance in most of the phenomena to be discussed in this section. Vacancies and interstitials occur naturally in a crystal, with equilibrium concentrations that increase exponentially with increasing temperature. If a crystal is cooled rapidly, or quenched, from high temperature, excess vacancies and interstitials may be frozen in. Theory indicates that the vacancies will act as acceptors and interstitials as donors [79]. Bombardment experiments tend to bear this out. A somewhat similar situation, as regards dependence of solubility on temperature, exists for many of the impurities in germanium and silicon, for example, copper, lithium, and gold. Their solubility has a maximum at high temperature and decreases as the temperature goes down [60]. Thus again, if the impurity is dissolved, *i.e.*, in atomic dispersion, in the crystal at high temperature, quenching may freeze this state in. In both cases, heating at a temperature lower than that from which the sample was quenched but high enough for the defect or impurity to be mobile will cause a change in the state of dispersion. This in turn may cause a change in the energy levels provided in the forbidden gap. How rapidly these changes occur, and just what the new state of dispersion is, depend again on the detailed nature of the defects present in the sample. For example, the presence of a high dislocation density in germanium samples has been found to greatly speed the precipitation of copper [80]. Also, excess copper has been observed to precipitate along dislocation lines in silicon [81]. This brief general discussion should serve to indicate the complexity of the problem of changes on heat treatment. Despite this, some features are now well understood, and we shall dwell on these in the remainder of this section.

It was found early that acceptor centers are introduced into germanium when it is heated at temperatures greater than 500°C and quenched. The concentration of these centers was usually in the range 10^{14} to 10^{17} per cm^3, becoming higher as the heating temperature increased. Lifetime was also found to deteriorate greatly as a result of quenching. The effect on resistivity and part of the effect on lifetime could be wiped out by prolonged heating at about 500°C [82]. This thermal conversion could be accounted for, in principle, by the introduction of lattice defects or impurities. It now is generally accepted that the effect is mainly produced by copper [83]. Apparently, unless special precautions are taken, copper is usually present as a surface contaminant, and the heating causes it to diffuse into the bulk. In one set of experiments it was established that the diffusion rate [84], solid solubility, and energy levels [85] of copper are, within experimental error, those observed for the thermal acceptors. More conclusive, however, is a later set of experiments in which copper was substantially eliminated from samples by prolonged heating in high vacuum [86] or chemical treatment [87]. Upon heating and quenching after such elimination it was found that greatly reduced numbers of acceptors were introduced. The concentration was considerably higher, however, than could be accounted for by the copper remaining in the sample. By quenching rapidly from almost the melting temperature a maximum of about 10^{15} per cm^3 could be introduced [87], [88]. This number falls off rapidly with the temperature from which the sample is quenched because the activation energy associated with the formation process is 2 ev. A rapid quench seems necessary to freeze in an appreciable number. It has been estimated that a concentration of less than 6×10^{12} per cm^3 of such acceptors is to be expected in the usual pulled or zone-grown crystal [89].

Some controversy still remains as to the nature of these residual acceptors, and indeed as to whether those seen by different observers are the same [90]. The view that has most support is that they are lattice defects. A study of their annealing rates as a function of temperature reveals some unusual properties [88]. At temperatures above 600°C the acceptor concentration drops very rapidly at first, and then much more slowly with annealing time. Below about 500°C, the acceptor concentration is found to increase before it decreases. On the basis of these properties, Mayburg suggested that the defects are vacancies and interstitials present in equal number [88]. Letaw, however, presented arguments for identifying the defects as vacancies and ascribed the complications on annealing to the formation of divacancies and vacancy clusters of higher order [90]. In this picture, the increase in acceptor concentration with annealing is attributed to additional acceptor levels arising from formation of a divacancy.

Heat treatment of silicon was found also to affect profoundly resistivity, lifetime, and trap concentrations. Here, too, there are indications of different effects, *i.e.*, effects of different origin. Most of the following discussion is based on one very commonly found set of effects that has had considerable study. These differ from those in germanium in that quenching from elevated temperatures tends to make the material n-type rather than p-type. Also, they appear more dependent on the parameters of crystal growth [91]. The samples showing the largest effects are from pulled crystals rotated during growth. Unrotated pulled crystals are much less affected, and samples produced by the floating zone technique [92] apparently are not affected at all. In the pulled, rotated samples heating in the temperature range 350°–500°C can result in the introduction of as many as 5×10^{16} donors per cm^3 [91]. These donor states can be removed by heating at temperatures above 500°C, and in only a few minutes by heating at 1100°C. Heating the sample for a longer period (*e.g.*, 20 hours) at 1100°C "stabilizes" it in the sense that the rate of change of resistivity on subsequent heating at 450°C is markedly reduced [91]. This stabilization can be erased by heating for a time at 1300–1400°C and cooling rapidly.

As mentioned previously, it now has been shown that these effects are related to the presence of oxygen in the crystals. The first development along this line was the discovery that the commonly observed infrared absorption peak at 9 microns in silicon [93] is correlated with oxygen content of the specimens [69]. This absorption was found to be very much stronger in crystals pulled from a crucible than in those produced by the floating zone technique. Analysis by vacuum fusion revealed concentrations of oxygen up to $10^{18}/cm^3$ in pulled crystals, at least one hundred times less in the floating zone crystals. The source of this oxygen was presumed to be the quartz crucible in which silicon crystals are commonly pulled. It was suggested that the oxygen is dispersed in the lattice interstitially, each oxygen having formed a covalent bond with two neighboring silicon atoms. The 9-micron absorption is attributed to the stretching vibrations of the Si-O bonds [69], [94]. This model has been confirmed by further study of the absorption, including isotopic shifts on substitution of O^{18} for O^{16} [95].

The presence of oxygen in the form just described does not result in any free electrons or holes. Most of the oxygen is apparently in this form in a crystal that has undergone no heat treatment beyond what is involved in crystal growth. It was discovered subsequently, however, that the number of donors produced during heat treatment around 450°C depends very strongly on oxygen concentration. Specifically, the initial rate of donor production on heating at 450°C has been found to be proportional to the fourth power of the oxygen concentration in the sample [96]. To explain this dependence, Kaiser has speculated that the heating produces clustering of the initially dispersed oxygen into SiO_4 complexes, each complex somehow acting as or giving rise to a singly or multiply charged donor.

Prolonged heat treatment at 1000°C was found to decrease greatly the 9-micron absorption without decreasing the oxygen content of the sample. The rate of decrease of the 9-micron absorption was shown to be markedly affected by the dislocation density in the sample [97]. Such heat treatment was also found to result in Rayleigh scattering and observation of the Tyndall Effect [96]. These effects have been interpreted as due to the further clustering of oxygen to form particles of silicon oxide within the silicon matrix. Initial insensitivity of a sample to a subsequent 450°C heat treatment has been attributed to insufficient oxygen remaining in solution to produce an appreciable number of donor states [96]. On this model, prolonged heating at 450°C must bring the oxygen back into solution, for eventually about the same number of donors is formed [91].

Some objections have been raised to the model suggested by Kaiser [91]. For one thing, it appears to require a greater mobility of oxygen in the lattice around 500°C than would be expected by extrapolation of diffusion data for oxygen from higher temperature. There seems to be no question, however, about the importance of oxygen in these phenomena. The greater effects of heat treatment on rotated, as compared with unrotated, pulled samples have been correlated with higher oxygen contamination of the former [98]. The greater effect of heat treatment in the center of the rotated crystals than in the skin can be similarly explained. An investigation of the way in which oxygen enters and leaves a silicon melt, carried out by Kaiser and Keck [99], leads to an understanding of these facts and many others. These authors have demonstrated that the high oxygen content of pulled silicon crystals results from the reduction of the quartz crucible by the silicon melt. Oxygen also leaves the liquid silicon quite readily, however, as gaseous SiO. The combination of constant introduction of oxygen into the melt at the walls of the crucible and the loss of oxygen at the surface can obviously result in an inhomogeneous oxygen content of the pulled crystals. It is then reasonable, for example, to expect that the skin of a pulled crystal will contain less oxygen than the center. These authors also studied the effect of varying some of the crystal-pulling parameters, such as temperature, growth rate, and atmosphere. Changes in temperature during growing were found to produce distinct fluctuations in oxygen content. The larger oxygen content of rotated crystals, they suggest, is due to the stirring action of the rotating crystal, which supplies oxygen-rich melt to the interface. Variations in oxygen content also have been correlated with the growth ring pattern and strain pattern, as observed in the snooperscope [100] of pulled silicon crystals.

Oxygen impurities may have an important effect on many other properties of silicon crystals. The density of storage traps is typically about 10^{13} per cm^3 in pulled rotated crystals, 10^{11} per cm^3 in nonrotated crystals, and much less in floating zone specimens [101]. This trend parallels that of the oxygen concentration in the samples. The density of these traps is decreased by heating at 1000°C. It has also been suggested that mechanical properties of silicon are affected by the presence of oxygen [102]. The rate of chemical etching of silicon samples also has been shown to be affected by the presence of oxygen, dissolved oxygen decreasing it, precipitated oxygen increasing it [103].

Despite its obvious importance, oxygen is not responsible for all changes which take place on heat treatment of silicon samples. In a number of samples studied by Bemski and Struthers [104], it was found that heating to temperatures above 900°C frequently degraded the lifetime, this being accompanied sometimes by an increase in resistivity of both n- and p-type samples. These effects have been shown to be due to gold, which performs here much as copper does in germanium. Apparently, gold is very commonly present as a surface contaminant on silicon samples. Having a high diffusion constant, it diffuses in rapidly on heating at high temperatures. Like copper in germanium, it can be removed by heating in vacuum. Undoubtedly, other impurities

and defects play a role in the heat treatment changes also, but this has not as yet been established.

Acknowledgment

Special thanks are due to F. M. Smits of Bell Telephone Laboratories for calculating and plotting Figs. 3 and 4. It is a pleasure to acknowledge the constructive comments of A. Many and G. D. O'Neill.

Bibliography

[1] For a more detailed and rigorous discussion of wave functions, crystal symmetry, Brillouin Zones, band structure, etc., starting from fundamentals, see the excellent review article by Herman, F. "The Electronic Energy Band Structure of Silicon and Germanium," PROCEEDINGS OF THE IRE, Vol. 43 (December, 1955), pp. 1703–1732. This paper also contains a detailed discussion of theoretical and experimental results for the band structure of germanium and silicon, and a bibliography.

[2] Conwell, E. M. "Properties of Silicon and Germanium," PROCEEDINGS OF THE IRE, Vol. 40 (November, 1952), pp. 1327–1337. This paper is referred to henceforth as I.

[3] For more detailed plots, and plots in some other directions in P space, see Herman [1].

[4] Dresselhaus, G., Kip, A. F., and Kittel, C. "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals," *Physical Review*, Vol. 98 (April 15, 1955), pp. 368–384.
Dexter, R. N., Zeiger, H. J., and Lax, B. "Cyclotron Resonance Experiments in Silicon and Germanium," *Physical Review*, Vol. 104 (November 1, 1956), pp. 637–644.

[5] For a much more extensive treatment than can be given here see Herring, C. "Transport Properties of a Many-Valley Semiconductor," *Bell System Technical Journal*, Vol. 34 (March, 1955), pp. 237–290.

[6] Elliott, R. J. "Theory of the Effect of Spin-Orbit Coupling on Magnetic Resonance in Some Semiconductors," *Physical Review*, Vol. 96 (October 15, 1954), pp. 266–279.

[7] See, for example, Shockley, W. *Electrons and Holes in Semiconductors*. New York: D. Van Nostrand Co., Inc., 1950.

[8] Harrick, N. J. "Attempt to Detect High Mobility Holes in Germanium Using the Drift Mobility Technique," *Physical Review*, Vol. 98 (May 15, 1955), pp. 1131–1133.

[9] For a more detailed investigation of this situation see Rittner, E. S. "Simultaneous Transport of Heavy and Light Holes in Semiconductors with a Degenerate Valence Band," *Physical Review*, Vol. 101 (February 15, 1956), pp. 1291–1294.

[10] It has been speculated, however, that in a drift experiment with hot electrons it might be possible to observe separate pulses from the different valleys. See Gold, L. "Anisotropy of the Hot-Electron Problem in Semiconductors with Spheroidal Energy Surfaces," *Physical Review*, Vol. 104 (December 15, 1956), pp. 1580–1584.

[11] Prince, M. B. "Drift Mobilities in Semiconductors, I. Germanium," *Physical Review*, Vol. 92 (November 1, 1953), pp. 681–687.

[12] Debye, P. P., and Conwell, E. M., "Electrical Properties of N-Type Germanium," *Physical Review*, Vol. 93 (February 15, 1954), pp. 693–706.

[13] Private communication from Debye, Raytheon Manufacturing Co., Waltham, Mass., and Conwell, Sylvania Electric Products, Inc., Bayside, N. Y.

[14] Ludwig G. W., and Watters, R. L. "Drift and Conductivity Mobility in Silicon," *Physical Review*, Vol. 101 (March 15, 1956), pp. 1699–1701.

[15] Debye, P. P., and Kohane, T. "Hall Mobility of Electrons and Holes in Silicon," *Physical Review*, Vol. 94 (May 1, 1954), pp. 724–725.

[16] A small correction for impurity scattering may still be necessary for n-type silicon. See discussion later on in this section.

[17] See, for example, Cronemeyer, D. C. "Hall and Drift Mobility in High-Resistivity Single-Crystal Silicon," *Physical Review*, Vol. 105 (January 15, 1957), pp. 522–523. In connection with this reference it might be worthwhile to clarify another point. The author compares his drift mobility values with those obtained by Prince, and Ludwig and Watters, and with $8/3\pi$ $\times$Hall mobility values of Debye and Kohane. He remarks that the latter values disagree widely with all the others. This is indeed the case, but it does not imply, as the author seems to think, that anyone's values are incorrect. Because of the complicated band structure the relation $\mu_H = (3\pi/8)\mu_D$ need not hold. See discussion later on in this section.

[18] Morin, F. J. "Lattice Scattering Mobility in Germanium," *Physical Review*, Vol. 93 (January 1, 1954), pp. 62–63.

[19] For a more detailed discussion on this level see Herman [1].

[20] Pelah, I., Eisenhauer, C. M., Hughes, D. J., and Palevsky, H. "Detection of Optical Lattice Vibrations in Ge and ZrH by Scattering of Cold Neutrons," *Physical Review*, Vol. 108 (November 15, 1957), pp. 1091–1092.

[20a] Lax, M. Bell Telephone Laboratories, N. J. Private communication.

[20b] The situation with regard to momentum of phonons is similar to that described earlier in the article for electrons, in that it is convenient to ascribe to the phonon a crystal momentum that does not really represent the momentum of the phonon. In what follows since it is only crystal momentum of carriers and phonons in which we are interested, we shall refer to it simply as "momentum."

[21] Ehrenreich, H., and Overhauser, A. W. "Lattice Scattering Mobility of Holes in Germanium," *Physical Review*, Vol. 104 (November 1, 1956), pp. 649–659.

[22] Harrison, W. A. "Scattering of Electrons by Lattice Vibrations in Non-Polar Crystals," *Physical Review*, Vol. 104 (December 1, 1956), pp. 1281–1290.

[23] Mansfield, R. "The Hall Effect in Semiconductors," *Proceedings of the Physical Society B, London*, Vol. 69 (August, 1956), pp. 862–865.
Blatt, F. J. "Hall and Drift Mobilities; Their Ratio and Temperature Dependence in Semiconductors," *Physical Review*, Vol. 105 (February 15, 1957), pp. 1203–1205.

[24] It is deduced from the cyclotron resonance experiments (see Dexter, Zeiger, and Lax [4]) that the mean free time between scattering processes is approximately the same for the two holes. Since $\mu = (e/m)\tau$, the ratio of mobilities is the inverse ratio of the masses, approximately 1/8. There is evidence that this ratio holds around room temperature also. See reference [25].

[25] Willardson, R. K., Harman, T. C. and Beer, A. C. "Transverse Hall and Magnetoresistance Effects in p-Type Germanium," *Physical Review*, Vol. 96 (December 15, 1954), pp. 1512–1518.

[26] Swanson, J. A. "Saturation Hall Constant of Semiconductors," *Physical Review*, Vol. 99 (September 15, 1955), pp. 1799–1807.

[27] Long, D. "Galvanomagnetic Effects in p-Type Silicon," *Physical Review*, Vol. 107 (August 1, 1957), pp. 672–677.

[28] Prince, M. "Drift Mobility in Semiconductors. II—Silicon," *Physical Review*, Vol. 93 (March 15, 1954), pp. 1204–1206.

[29] Backenstoss, G. "Conductivity Mobilities of Electrons and Holes in Heavily Doped Silicon," *Physical Review*, Vol. 108 (December 15, 1957), pp. 1416–1419.

[30] Horn, F. H. "Densitometric and Electrical Investigation of Boron in Silicon," *Physical Review*, Vol. 97 (March 15, 1955), pp. 1521–1525.
Carlson, R. O. "Electrical Properties of Near Degenerate Boron-Doped Silicon," *Physical Review*, Vol. 100 (November 15, 1955), pp. 1075–1078.

[31] Conwell, E. M. "Hall Effect and Density of States in Germanium," *Physical Review*, Vol. 99 (August 15, 1955), pp. 1195–1198.

[32] In detail, the density-of-states masses of Table III were used, and the degeneracy of the ground state of the impurity was taken as 2 for donors, 4 for acceptors. Excited states were not considered, however. Activation energies appropriate to the particular concentrations were determined from Hall data vs temperature on similar samples.

[33] For further discussion and references see Conwell, E. M. "Impurity Band Conduction in Germanium and Silicon," *Physical Review*, Vol. 103 (July 1, 1956), pp. 51–61. For a somewhat more elementary discussion of the same topic, see Conwell, *Sylvania Technologist*, Vol. 9 (July, 1956), pp. 82–84.

[34] Abeles, B., and Meiboom, S. "Theory of the Galvanomagnetic Effects in Germanium," *Physical Review*, Vol. 95 (July 1, 1954), pp. 31–37.
Shibuya, M. "Magnetoresistance Effect in Cubic Semiconductors with Spheroidal Energy Surfaces," *Physical Review*, Vol. 95 (September 15, 1954), pp. 1385–1393.
Pearson, G. L., and Herring, C. "Magnetoresistance Effect and the Band Structure on Single Crystal Silicon" in "Proceedings of the International Conference on Semiconductors, Amsterdam, 29 June–3 July, 1954," *Physica*, Vol. 20 (1954), pp. 975–978.

[35] For a general review and discussion of the last two categories, see Burstein, E., Picus, G., and Sclar, N. "Optical and Photoconductive Properties of Silicon and Germanium" in *Photoconductivity Conference*, eds. R. G. Breckenridge, B. R. Russell, and E. E. Hahn. New York: John Wiley and Sons, Inc., 1956. For additional information on germanium see Fisher, P., and Fan, H. Y. "Absorption Spectra of III, V Impurities in Germanium," *Bulletin of the American Physical Society*, Ser. II, Vol. 3 (March, 1958), p. 128.

[36] Bardeen, J., Blatt, F. J., and Hall, L. H. "Indirect Transitions from the Valence to the Conduction Bands," in *Photoconductivity Conference*.

[37] MacFarlane, G. G., McLean, T. P., Quarrington, J. E., and Roberts, V. "Fine Structure in the Absorption-Edge Spectrum of Ge," *Physical Review*, Vol. 108 (December 15, 1957), pp. 1377–1383.
[38] Haynes, J. R. "New Radiation Resulting from Recombination of Holes and Electrons in Germanium," *Physical Review*, Vol. 98 (June 15, 1955), pp. 1866–1868.
[39] Haynes, J. R., Lax, M., and Flood, W. "Fine Structure of Intrinsic Recombination Radiation in Silicon," *Bulletin of the American Physical Society*, Ser. II, Vol. 3 (January, 1958), p. 30.
[40] For a discussion of the theory see, for example, Fan, H. Y., and Becker, M., "Infra-red Optical Properties of Silicon and Germanium" in *Semi-Conducting Materials*, ed. H. K. Henisch. London: Thornton Butterworth, Ltd., 1951.
[41] Kaiser, W., Collins, R. J., and Fan, H. Y. "Infrared Absorption in *P*-Type Germanium," *Physical Review*, Vol. 91 (September 15, 1953), pp. 1380–1381.
Briggs, H. B., and Fletcher, R. C. "Absorption of Infrared Light by Free Carriers," *Physical Review*, Vol. 91 (September 15, 1953), pp. 1342–1346.
[42] Kahn, A. H. "Theory of the Infrared Absorption of Carriers in Germanium and Silicon," *Physical Review*, Vol. 97 (March 15, 1955), pp. 1647–1652.
[43] Herring, C. "Theory of the Thermoelectric Power of Semiconductors," *Physical Review*, Vol. 96 (December 1, 1954), pp. 1163–1187.
[44] Lax, B., and Mavroides, J. G. "Statistics and Galvanomagnetic Effects in Germanium and Silicon with Warped Energy Surfaces," *Physical Review*, Vol. 100 (December 15, 1955), pp. 1650–1657.
[45] Morin, F. J., and Maita, J. P. "Conductivity and Hall Effect in the Intrinsic Range of Germanium," *Physical Review*, Vol. 94 (June 15, 1954), pp. 1525–1529.
[46] ——. "Electrical Properties of Silicon Containing Arsenic and Boron," *Physical Review*, Vol. 96 (October, 1954), pp. 28–35.
[47] Several determinations of the density-of-states mass from other types of data also have led to this conclusion. For references to them see MacFarlane *et al.* [37].
[48] Pell, E. M., and Roe, G. M. "Reverse Current and Carrier Lifetime as a Function of Temperature in Silicon Junction Diodes," *Journal of Applied Physics*, Vol. 27 (July, 1956), pp. 768–772.
[49] McAfee, K. B., Ryder, E. J., Shockley, W., and Sparks, M. "Observations of Zener Current in Germanium *p-n* Junctions," *Physical Review*, Vol. 83 (August 1, 1951), pp. 650–651.
[50] McKay, K. G., and McAfee, K. B. "Electron Multiplication in Silicon and Germanium," *Physical Review*, Vol. 91 (September 1, 1958), pp. 1079–1084.
[51] McKay, K. G. "Avalanche Breakdown in Silicon," *Physical Review*, Vol. 94 (May 15, 1954), pp. 877–884.
[52] Miller, S. L. "Avalanche Breakdown in Germanium," *Physical Review*, Vol. 99 (August 15, 1955), pp. 1234–1241.
[53] Newman, R. "Visible Light from a Silicon *p-n* Junction," *Physical Review*, Vol. 100 (October 15, 1955), pp. 700–703.
[54] Chynoweth, A. G., and McKay, K. G. "Photon Emission from Avalanche Breakdown in Silicon," *Physical Review*, Vol. 102 (April 15, 1956), pp. 369–376.
[55] Chynoweth, A. G., and Pearson, G. L. "Effect of Dislocations on Breakdown in Silicon *p-n* Junctions," *Bulletin of the American Physical Society*, Ser. II, Vol. 3 (March, 1958), p. 112.
[56] Chynoweth, A. G., and McKay, K. G. "Internal Field Emission in Silicon *p-n* Junctions," *Physical Review*, Vol. 106 (May 1, 1957), pp. 418–426.
[57] Esaki, L. "New Phenomenon in Narrow Germanium *p-n* Junctions," *Physical Review*, Vol. 109 (January 15, 1958), pp. 603–604.
[58] For discussion and references see Gunn, J. B. "High Electric Field Effects in Semiconductors" in *Progress in Semiconductors*, eds. A. F. Gibson, P. Aigrain, and R. E. Burgess. New York: John Wiley and Sons, Inc., 1957, Vol. 2, pp. 211–247.
[59] Sclar, N., and Burstein, E. "Impact Ionization of Impurities in Germanium," *Journal of Physical Chemistry and Solids*, Vol. 2 (March, 1957), pp. 1–23.
[60] Burton, J. A. "Impurity Centers in Ge and Si," *Physica*, Vol. 20 (1954), pp. 845–854.
[61] Dunlap, W. C., Jr. "Impurities in Germanium" in *Progress in Semiconductors*. Vol. 2, pp. 165–209.
[62] Hannay, N. B. "Recent Advances in Silicon" in *Progress in Semiconductors*. Vol. 1, pp. 1–35.
[63] Johnson, E. R., and Christian, S. M. "Some Properties of Germanium-Silicon Alloys," *Physical Review*, Vol. 95 (July 15, 1954), p. 560.
[64] Trumbore, F. A. "Solid Solubilities and Electrical Properties of Tin in Germanium Single Crystals," *Journal of the Electrochemical Society*, Vol. 103 (November, 1956), pp. 597–600.
[65] Fuller, C. S., and Ditzenburger, J. A. "Diffusion of Lithium into Germanium and Silicon," *Physical Review*, Vol. 91 (July 1, 1953), p. 193.
[66] van der Maesen, F., and Brenkman, J. A. "On the Behavior of Rapidly Diffusing Acceptors in Germanium," *Journal of the Electrochemical Society*, Vol. 102 (May, 1955), pp. 229–234.
[67] Woodbury, H. H., and Tyler, W. W. "Triple Acceptors in Germanium," *Physical Review*, Vol. 105 (January 1, 1957), pp. 84–92.
[68] Thurmond, C. D., Guldner, W. G., and Beach, A. L. "Hydrogen and Oxygen in Single-Crystal Germanium as Determined by Vacuum Fusion Gas Analysis," *Journal of the Electrochemical Society*, Vol. 103 (November, 1956), pp. 603–605.
[69] Kaiser, W., Keck, P. H., and Lange, C. F. "Infrared Absorption and Oxygen Content in Silicon and Germanium," *Physical Review*, Vol. 101 (February 15, 1956), pp. 1264–1268.
[70] Collins, C. B., Carlson, R. O., and Gallagher, C. J. "Properties of Gold-Doped Silicon," *Physical Review*, Vol. 105 (February 15, 1957), pp. 1168–1173.
[71] Collins, O. B., and Carlson, R. O. "Properties of Silicon Doped with Iron or Copper," *Physical Review*, Vol. 108 (December 15, 1957), pp. 1409–1414.
[72] Reiss, H., Fuller, C. S., and Morin, F. J. "Chemical Interactions Among Defects in Germanium and Silicon," *Bell System Technical Journal*, Vol. 35 (May, 1956), pp. 535–636.
[73] Fuller, C. S., and Morin, F. J. "Diffusion and Electrical Behavior of Zinc in Silicon," *Physical Review*, Vol. 105 (January 15, 1957), pp. 379–384.
[74] Morin, F. J., and Reiss, H. "Formation of Ion Pairs and Triplets between Lithium and Zinc in Germanium," *Physical Review*, Vol. 105 (January 15, 1957), pp. 384–389.
[75] For a review of the considerable work done on bound wave functions and energy levels for the actual band structure see Kohn, W. "Shallow Impurity States in Silicon and Germanium" in *Solid State Physics*, eds. F. Seitz and D. Turnbull. New York: Academic Press, Inc., 1957, Vol. 5, pp. 257–320.
[76] For such information see Kohn [75] and Conwell [31].
[77] Shulman, R. G. "Tight-Bonding Calculation of Acceptor Energies in Germanium and Silicon," *Journal of Physics and Chemistry of Solids*, vol. 2 (April 1, 1957), pp. 115–118.
[78] Reiss, H. "Theory of the Ionization of Hydrogen and Lithium in Silicon and Germanium," *Journal of Chemical Physics*, Vol. 25 (October, 1956), pp. 681–686.
[79] James, H. M. and Lark-Horovitz, K. "Localized Electronic States in Bombarded Semiconductors," *Zeitschrift für physikalische Chemie*, Vol. 198, Nos. 1–4 (1951), pp. 107–126.
[80] Logan, R. A. "Precipitation of Copper in Germanium," *Physical Review*, Vol. 100 (October 15, 1955), pp. 615–617.
[81] Dash, W. C. "Copper Precipitation on Dislocations in Silicon," *Journal of Applied Physics*, Vol. 27 (October, 1956), pp. 1193–1195.
[82] The residual effect on lifetime may have been due to plastic deformation during heat treatment. See Logan, R. A., and Schwartz, M. "Thermal Effects on Lifetime of Minority Carriers in Germanium," *Physical Review*, Vol. 96 (October 1, 1954), p. 46.
[83] For a critical review of the many experiments to date see Letaw, H. "Thermal Acceptors in Germanium," *Journal of Physics and Chemistry of Solids*, Vol. 1 (September/October, 1956), pp. 100–116.
[84] Fuller, C. S., and Struthers, J. D. "Copper as an Acceptor Element in Germanium," *Physical Review*, Vol. 87 (August 1, 1952), pp. 526–527.
[85] Morin, F. J., and Maita, J. F. "Comparison of Copper-Doped Germanium with Heat-Treated Germanium," *Physical Review*, Vol. 90 (April 15, 1953), p. 337.
[86] Finn, G. "Evaporation of Copper from Germanium," *Physical Review*, Vol. 91 (August 1, 1953), pp. 754–755.
Mayburg, S., and Rotondi, L. "Thermal Acceptors in Vacuum Heat-Treated Germanium," *Physical Review*, Vol. 91 (August 15, 1953), pp. 1015–1016.
[87] Logan, R. A. "Thermally Induced Acceptors in Single Crystal Germanium," *Physical Review*, Vol. 91 (August 1, 1953), pp. 757–758.
[88] Mayburg, S. "Vacancies and Interstitials in Heat Treated Germanium," *Physical Review*, Vol. 95 (July 1, 1954), pp. 38–43.
[89] ——. Sylvania Electric Products, Inc., private communication.
[90] Letaw [83].
[91] Fuller, C. S., and Logan, R. A. "Effect of Heat Treatment Upon the Electrical Properties of Silicon Crystals," *Journal of Applied Physics*, Vol. 28 (December, 1957), pp. 1427–1436.
[92] Keck, P. H. "Floating Zone Crystallization of Silicon," in "Proceedings of the International Conference on Semiconductors, Amsterdam, 29 June–3 July, 1954," *Physica*, Vol. 20 (1954), pp. 1059–1065.
[93] See, for example, Fig. 9 of I.

[94] These authors also propose that an absorption in germanium at 11.6 microns is due to similar Ge-O vibrations.
[95] Hrostowski, H. J., and Kaiser, R. H. "Infrared Absorption of Oxygen in Silicon," *Physical Review*, Vol. 107 (August 15, 1957), pp. 966–972.
[96] Kaiser, W. "Properties of Heat-Treated Silicon," *Physical Review*, Vol. 105 (March 15, 1957), pp. 1751–1756.
[97] Lederhandler, S., and Patel, J. R. "Behavior of Oxygen in Plastically Deformed Silicon," *Physical Review*, Vol. 108 (October 15, 1957), pp. 239–242.
[98] Fuller and Logan [91] make the point that the difference in oxygen content does not seem large enough, however, to explain the difference in the number of donors produced in heat treatment.
[99] Kaiser, W., and Keck, P. H. "Oxygen Content of Silicon Single Crystals," *Journal of Applied Physics*, Vol. 28 (August, 1957), pp. 882–887.
[100] Dash, W. C. "Birefringence in Silicon," *Physical Review*, Vol. 98 (June 1, 1955), p. 1536.
[101] Hannay, N. B., Haynes, J. R., and Shulman, R. G. "The Interaction of Traps and Heat Treatments in Silicon," *Physical Review*, Vol. 96 (November 1, 1954), p. 833.
[102] Pearson, G. L., Read, W. T., and Feldmann, W. L. "Deformation and Fracture of Small Silicon Crystals," *Bulletin of the American Physical Society*, Ser. II, Vol. 1 (June, 1956), p. 295.
[103] Logan, R. A., and Peters, A. J. "Effect of Oxygen on Etch-Pit Formation in Silicon," *Journal of Applied Physics*, Vol. 28 (December, 1957), pp. 1419–1423.
[104] Bemski, G., and Struthers, J. D. "Gold in Silicon," *Journal of the Electrochemical Society*, to be published. The author is indebted to Dr. Bemski for making the manuscript available prior to publication.
[105] Fan, H. Y., Shepard, M. L., and Spitzer, W. "Infrared Absorption and Energy-Band Structure of Germanium and Silicon" in *Photoconductivity Conference*.
[106] Dash, W. C., and Newman, R., "Intrinsic Optical Absorption in Single Crystal Germanium and Silicon at 77°K and 300°K," *Physical Review*, Vol. 99, August 15, 1955, pp. 1151–1155.
[107] Avery, D. G., and Clegg, P. L. "Optical Constants of Single Crystals of Germanium," *Proceedings of the Royal Society B, London*, Vol. 66 (June, 1953), pp. 512–513.
[108] Pfestorf, G. "Optical Constant of Metals in the Visible and Ultraviolet Region of the Spectrum," *Annalen der Physik*, Vol. 81 (December, 1926), pp. 906–928.
[109] Tyler, W. W. "Properties of Ag-Doped Germanium," *Bulletin of the American Physical Society*, Ser. II, Vol. 3 (March, 1958), p. 128.
[110] Dunlap, W. C., Jr. "Properties of Zn-, Cu- and Pt-Doped Germanium," *Physical Review*, Vol. 96 (October 1, 1954), pp. 40–45.
[111] Woodbury, H. H., and Tyler, W. W. "Properties of Zn-Doped Germanium," *Physical Review*, Vol. 100 (November 15, 1955), p. 1259.
[112] ——. "Recent Studies of Deep Impurity Levels in Ge," *Bulletin of the American Physical Society*, Ser. II, Vol. 1 (March, 1956), p. 127.
[113] Geballe, T. H., and Morin, F. J. "Ionization Energy of Groups III and V Elements in Germanium," *Physical Review*, Vol. 95 (August 16, 1954), pp. 1085–1086.
[114] Woodbury, H. H., and Tyler, W. W. "Properties of Germanium Doped with Manganese," *Physical Review*, Vol. 100 (October 15, 1955), pp. 659–662.
[115] Morin, F. J., Maita, J. P., Shulman, R. G., and Hannay, N. B. "Impurity Levels in Silicon," *Bulletin of the American Physical Society*, Vol. 29 (June, 1954), p. 22.
[116] Carlson, R. O. "Double-Acceptor Behavior of Zinc in Silicon," *Physical Review*, Vol. 108 (December 15, 1957), pp. 1390–1393.
[117] ——. "Properties of Silicon Doped with Manganese," *Physical Review*, Vol. 104 (November 15, 1956), pp. 937–941.

Correspondence

Parametric Amplification of the Fast Electron Wave*

The concept of space charge waves in electron streams is widely known. Restoring forces, generated in velocity modulation tubes by space charge fields and in transverse field tubes by the focusing field, produce a pair of such waves, commonly called the slow and the fast wave. Their phase velocities are

$$u_{12} = u_e \frac{\omega_1}{\omega_1 \pm \omega_r}$$

(u_e = velocity of stream; ω_1 = signal frequency; ω_r = resonant frequency corresponding to the restoring mechanism, such as plasma frequency or cyclotron frequency).

Conventional traveling-wave tubes employ interaction between a circuit and the slow electron wave to produce gain. It appears to be established[1] that the noise present in this wave cannot be removed.

On the other hand, noise cancellation is possible for the fast electron wave; interaction with this wave, unfortunately, does not produce gain in conventional devices.

It would be of great interest to construct a device which employs fast-wave interaction in its input and output sections and in which the amplitude of electron motion is increased between input and output by other means. Such an increase in amplitude might be obtained by utilizing an old principle which recently has become more widely known as parametric amplification. This method is based on nonlinear effects. It requires an external power source of large amplitude, operating at a pumping frequency ω_3 which must be higher than the signal frequency ω_1. An auxiliary resonant mesh must be provided at the beat frequency $\omega_2 = \omega_3 - \omega_1$, except in the special case where $\omega_3 = 2\omega_1$. A thorough discussion of this method was given by Manley and Rowe.[2]

As an example of a fast-wave device to which parametric amplification might be applied, consider Cuccia's electron coupler[3] (Fig. 1). This is a transverse field device, consisting of two lumped resonant cavities through which an electron stream flows; a homogeneous magnetic field confines this stream, its intensity so chosen that the cyclotron frequency equals the signal frequency. A signal applied to the first cavity

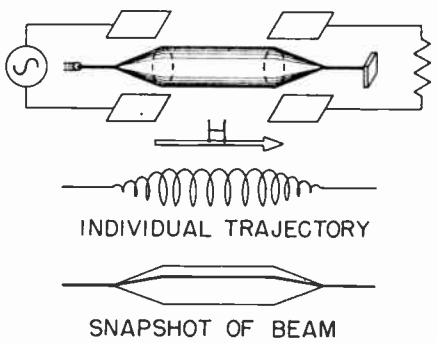


Fig. 1.

causes the electrons to spiral outward. They reach a maximum radius at the exit of the first cavity; at this point, all the signal power supplied appears in the form of kinetic energy.

In the second cavity the orbiting electrons induce a current and if the cavity is properly loaded, the resulting field causes them to spiral inward and give up all their energy.

There is no gain in this device and no interchange of energy occurs between the transverse signal motion and the longitudinal forward motion. All electrons throughout the device orbit in phase; the phase velocity is infinite in accordance with the condition $\omega_1 = \omega_e$.

* Received by the IRE, January 6, 1958. Presented at the Conference on Electron Tube Research, Berkeley, Calif., June, 1957.
[1] H. A. Haus and F. N. H. Robinson, "The minimum noise figure of microwave beam amplifiers," PROC. IRE, vol. 43, pp. 981–991; August, 1955.
[2] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements," PROC. IRE, vol. 44, pp. 904–913; July, 1956.
[3] C. L. Cuccia, "The electron coupler," *RCA Rev.*, vol. 10, pp. 270–303; June, 1949.

The second cavity absorbs from the stream all the available fast-wave signal power. The first cavity, properly loaded by a resistive source, should similarly absorb all the fast-wave noise power, leaving the stream quiet. This has been demonstrated (Fig. 2).

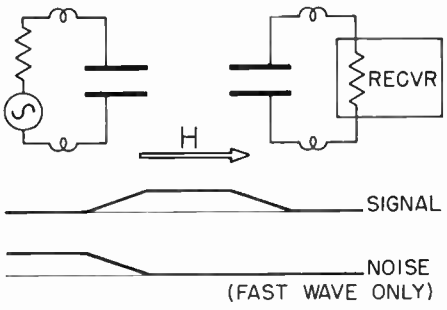


Fig. 2.

A small tube was built, using a strip beam of 200 μa at 6 volts in a magnetic field of about 180 Gauss. The two cavities were formed by two pairs of deflection plates, each pair having 1-cm length in the direction of beam travel, 1.5-cm width and 0.075-cm spacing. The two pairs were shielded from each other and tuned externally. Measurements were made at 528 mc.

A noise minimum was observed when the input circuit was carefully tuned and properly loaded and the magnetic field was correctly adjusted. Under these conditions signal transmission was also best.

Parametric amplification of the transverse electron motion could be produced by applying the pumping signal to an electrode system which produces an inhomogeneous transverse field across the stream. A fence-like array of electrodes of alternating polarity meets this purpose. Fig. 3 shows the proposed arrangement. For greater clarity, the fence posts are shown positioned at right angles to the direction of beam travel; it might be better to arrange them parallel to that direction.

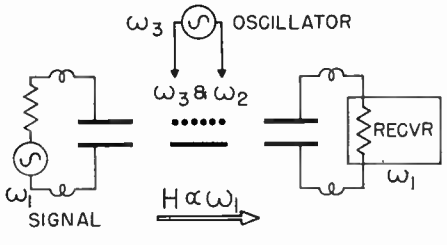


Fig. 3.

As previously stated, the array must be tuned to the beat frequency ω_2, unless the pumping frequency ω_3 is very nearly equal to twice the signal frequency ω_1.

An arrangement of this kind, employing parametric amplification between two fast-wave beam coupling devices, may well result in a very low noise figure.

ROBERT ADLER
Zenith Radio Corp.
Chicago, Ill.

Experimental Characteristics of a Microwave Parametric Amplifier Using a Semiconductor Diode*

The purpose of this letter is to report our observation of the characteristics of a microwave parametric amplifier[1-3] which we have constructed recently. The amplifier employs a back-biased germanium junction diode placed between central posts inside a rectangular cavity as shown in Fig. 1. With proper adjustments of the tuning screws, the cavity with diode in place can be made simultaneously resonant at 3500 mc, 2300 mc, and 1200 mc. We believe these three resonances are the perturbed TE_{103}, TE_{301}, and TE_{101} modes of the empty rectangular box. A loop serves to couple into the 3500-mc resonance and two probes serve to couple in and out of the cavity at either the 1200-mc or the 2300-mc resonance. With proper adjustment they can be made to couple predominantly to the 1200-mc resonance with an insertion loss which depends upon diode bias but is typically 4 to 6 db.

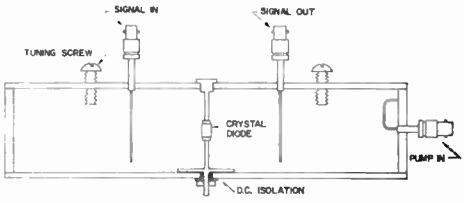


Fig. 1—Cross-section of variable-parameter amplifier.

The diode[4] which was kindly supplied through the Evans Signal Laboratory has a low-frequency zero-bias capacitance of 1 $\mu\mu$f and a spreading resistance of 5 ohms as measured on a standard impedance bridge.

With a pump power at 3500 mc applied to the cavity slightly in excess of 100 mw, the device oscillates at both 1200 mc and 2300 mc either with or without dc diode bias. Below this value of pump power, amplification can be obtained at either of the two frequencies. The majority of our measurements have so far been made on the operation of the device as an amplifier at the lowest of these frequencies using of the order of 5 volts diode bias.

We have been able to obtain net gain up to 40 db, although at such high gains the amplifier is extremely sensitive to load variations since no isolator or circulator is employed. Moreover, there is some noticeable gain variation due to the residual frequency modulation of the 3500-mc pump generator (a Hewlett-Packard signal generator followed by a 1 watt TW tube and filter). The bandwidth at 19 db gain is 1.0 mc and follows the predicted relation-gain times bandwidth squared equals a constant. The saturated power output is approximately 1.5 mw. The dynamic range is over 100 db.

Our preliminary measurement of noise figure at 16 db gain indicates the noise figure is less than 4.8 db.

H. HEFFNER
K. KOTZEBUE
Electronics Labs.
Stanford University
Stanford, Calif.

Noise Figure Measurements on Two Types of Variable Reactance Amplifiers Using Semiconductor Diodes*

The purpose of this letter is to report experimental results on low-noise amplification at uhf and microwave frequencies by *p-n* junction diodes. Some twelve years have passed since amplification by semiconductor diodes was first reported[1] in an experiment involving frequency conversion from 10,000 mc to 30 mc. The noise figures obtained were not encouraging, and the work was discontinued. Recent studies [2-4] show that up-conversion from a low frequency to a high frequency is more likely to produce stable, low-noise amplification.

Low-noise performance also has been predicted for microwave amplification involving a negative resistance type of diode interaction[5] and experimental verification of the gain mechanism of such an amplifier has been achieved.[6]

Measurements which bear out the theoretically predicted low-noise performance of both the negative resistance amplifier and the up-converter are described. The lowest noise figures reported here have been obtained using special diffused silicon *p-n* junction diodes[7] having small capacitance and very low values of series resistance. Further information on the properties and prospec-

* Received by the IRE, March 24, 1958. This work was performed under a supported joint Services research contract N6ONR 225(24) while K. Kotzebue held a Bell Telephone Laboratories Fellowship.

[1] H. Suhl, "Proposal for a ferromagnetic amplifier in the microwave range," *Phys. Rev.*, vol. 106, pp. 384–385; April 15, 1957.

[2] M. E. Hines proposed the use of a semiconductor diode as a variable element and reported some experimental results at the Annual Conference on Electron Tube Research, Berkeley, Calif.; June, 1957.

[3] H. Heffner and G. Wade reported an analysis of the noise figure and gain-bandwidth characteristics of the parametric amplifier at the Annual Conference on Electron Tube Research.

[4] This type of diode was developed by A. Uhlir, Bell Telephone Laboratories, and has been given the number W.E.427A.

* Received by the IRE, April 4, 1958. This work was supported in part by the U. S. Army Signal Corps.

[1] H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 15, ch. 13; 1948.

[2] A. Uhlir, Jr., "Two-terminal *p-n* junction devices for frequency conversion and computation," PROC. IRE, vol. 44, pp. 1183–1191; September, 1956.

[3] "Semiconductor Diodes Yield Converter Gain," Bell Labs. Rec., vol. 35, p. 412; October, 1957.

[4] A. Uhlir, Jr., "Diffused Silicon and Germanium Nonlinear Capacitors," presented at the IRE-AIEE Semiconductor Device Research Conf., Boulder, Colo.; July, 1957.

[5] M. E. Hines, paper submitted for publication.

[6] M. E. Hines, "Amplification in Nonlinear Reactance Modulators," presented at the 15th Annual Conf. on Electron Tube Research, Berkeley, Calif.; June, 1957.

[7] A. E. Bakanowski, N. G. Cranna, and A. Uhlir, Jr., "Diffused Germanium and Silicon Nonlinear Capacitor Diodes," presented at the IRE-AIEE Semiconductor Device Research Conf., Boulder, Colo.

tive uses of these diodes will be found elsewhere.[8]

The fundamental power relations for a nonlinear (lossless) reactance are summed up by a theorem due to Manley and Rowe.[9] When a strong local oscillator or pump of frequency f_p and a signal of frequency f_s are impressed simultaneously on a nonlinear reactance, frequencies equal to mf_p+nf_s are produced. If one considers only the signal and the two lowest sidebands, f_{p-s} and f_{p+s}, and ignores all of the other higher order modulation products, then the power relation due to Manley and Rowe becomes

$$\frac{P_s}{f_s}+\frac{P_{p+s}}{f_{p+s}}-\frac{P_{p-s}}{f_{p-s}}=0 \qquad (1)$$

where P_s, P_{p+s}, and P_{p-s} are the powers at frequencies f_s, f_{p+s}, and f_{p-s}. A positive value of P denotes power generated by the amplifier, and a negative value denotes power absorbed by the amplifier.

In the simplest version of the up-converter, the circuit is tuned to f_s and f_{p+s} only, and (1) becomes

$$P_{p+s}+\frac{f_{p+s}}{f_s}P_s=0. \qquad (2)$$

Under these conditions P_s and P_{p+s} have opposite signs so that power will be emitted at f_{p+s} when signal power is being supplied at f_s; furthermore, the emitted power will exceed the signal power, the power ratio or gain being equal to f_{p+s}/f_s. The relation also shows that gain at the signal frequency is impossible because the difference in signs for P_s and P_{p+s} would require that a positive value of P_s be accompanied by an even larger expenditure of signal power at f_{p+s}.

In the simple negative resistance amplifier, on the other hand, the circuit is tuned only to f_s and its "image" f_{p-s} so that (1) becomes

$$\frac{P_s}{f_s}=\frac{P_{p-s}}{f_{p-s}}. \qquad (3)$$

Here P_s and P_{p-s} have like signs so that amplification at the signal frequency does not require absorption of signal power at another frequency. One therefore can have power gain in the ordinary sense that input power at f_s gives rise to an amplified power output at the same frequency. In this process, power will necessarily also be emitted at f_{p-s}. The amount of gain obtainable is limited only by the properties of the circuit and obeys a constant gain-bandwidth product relation. Gain under these conditions may be thought of as a negative resistance effect.

One also can allow all three signals to be present. In this case, gain can be obtained as a result of a frequency shift as well as from negative resistance effects. The general properties of such an amplifier are analyzed in detail in a forthcoming paper.[5]

[8] A. Uhlir, Jr., "The potential of semiconductor diodes in high-frequency communications," this issue, p. 1099.

[9] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—Part I. General energy relations," PROC. IRE, vol. 44, pp. 904–913; July, 1956.

The experimental results previously reported by Hines[6] were obtained using a point contact germanium diode with pump at 12,350 mc, and amplification was obtained at frequencies near 6175 mc. Under certain conditions of circuit adjustment, oscillations at precisely 6175 mc were also observed. These results are in accordance with the predictions of (3). Experimental results are given on the two types of parametric amplifier discussed above.

Up-Converter

An "up-converter" circuit was assembled in which the signal is introduced at 460 mc and the beating oscillator or "pump" at 8915 mc. The arrangement is shown schematically in Fig. 1. The output at the upper sideband, 9375 mc, was detected by a superheterodyne receiver using balanced point-contact crystals in conventional fashion to give a second-stage noise figure of 7 db. In typical operation, the up-converter gain was 9 db and the over-all noise figure, including noise from the second stage, was 2.5 ± 0.5 db. The first-stage (up-converter) noise figure was therefore 2 ± 0.5 db, which may also be expressed as a "noise temperature" of 170 ± 50°K. These measurements were obtained by standard noise lamp procedures. The over-all noise figure was checked, within the indicated accuracy, by using a refrigerated input load, and also with a signal generator of ±2-db accuracy. About 200 mw of pump power was absorbed, but 2 w of incident power was used to avoid having to provide a good match at the pump frequency. Reasonably broad-band amplification is predicted by the theory, but has not yet been achieved experimentally.

A particular adjustment of the circuit (which has not been reproducible at will) gave 21 db of gain and an over-all noise figure of 1.1 ± 0.5 db (87 ± 40°K). Since the gain in this case exceeded the frequency ratio, 9375/460, this situation must have involved the lower sideband to a considerable extent.

Negative Resistance Amplifier

A negative resistance amplifier was designed to operate near 6 kmc with pump power at 11.7 kmc. The experimental setup is seen in Fig. 2. Amplification is obtained by returning more signal power into the input waveguide of the diode than was incident there, and a circulator is used to provide isolation between ingoing and outcoming waves. The circulator also prevents thermal noise originating in the load from being amplified.

The gain of the amplifier depends on pumping power level as well as on circuit adjustment. The pump requirement is simply one of obtaining a suitable voltage swing across the nonlinear capacitance. Hence a diode with a high value of C_0 (where C_0 is the static capacitance of the diode) will require in general more pump power than one with a low value, and a circuit which is high Q for the pump will be more efficient than a low-Q circuit. The amount of pump power used to obtain the results given below was between 50 and 500 mw with no special attempt made to optimize the pump circuit.

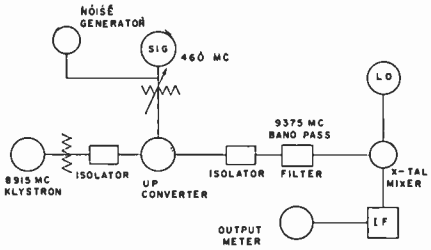


Fig. 1—Up-converter arrangement.

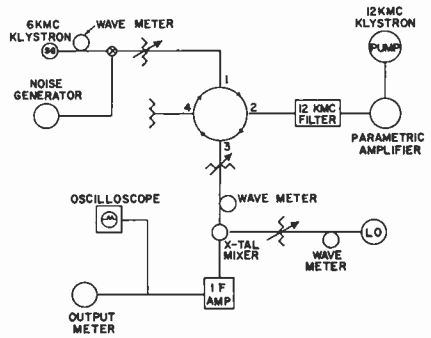


Fig. 2—Negative resistance amplifier arrangement.

Silicon and germanium diffused p-n junction diodes and welded-contact (gold-bonded) germanium diodes were used. Values of C_0 with no dc bias ranged from 0.4 to 2.85 mmf. The silicon junction diodes were usually operated with zero dc bias, but a negative bias voltage of from 1 to 1.5 volts was applied to the germanium diodes to reduce C_0 to a suitable value for optimum amplification. Stable power gains as high as 45 db were measured at $f_s=f_{p/2}$. In general, a constant (bandwidth)·(voltage-gain) product was obtained as expected. With the circuit adjusted to give maximum gain of 18 db at $f_s=(f_{p/2})-10$ mc, a bandwidth of 8 mc was measured to the 3-db down points.

The most important measurement to be reported for the negative resistance amplifier is its noise figure. Because of the nature of the amplification process, noise appears in frequency band f_s at the output due to: 1) input noise at f_s which is amplified, 2) input noise at f_{p-s} which is converted to f_s in the amplification process, 3) noise which originates in the diode at f_s and is amplified, and 4) noise from the diode at f_{p+s} which is converted to f_{p-s} in the amplification process. Because of this generation of "image" frequencies during amplification, the noise performance of this type of amplifier should be better for use with a double-sideband signal centered at $f_{p/2}$ (*i.e.*, with coherent signal frequencies at both f_s and f_{p-s}) than with a signal having the same total power but with no special symmetry about $f_{p/2}$. Amplifier noise figures as low as 3 db for this special kind of double-sideband operation have been obtained, corresponding to a 6-db noise figure for ordinary operation. However, note that (because of noise source 2) above) when the equivalent temperature of the input load is decreased, even with ordinary operation the added noise becomes less than that of an ordinary amplifier having a 6-db noise figure.

The most important source of noise in the diode is that originating as thermal noise in the series resistance, R_s. Analysis shows that the product $q_s = f_s R_s C_0$ should be minimized for best noise performance, and reasonable correlation between amplifier noise figure and q_s was obtained indeed for some ten diodes tested in the negative resistance amplifier. It will be noted that considerably lower noise figures have been obtained in the 400-mc up-modulator than in the 6-kmc amplifiers, so that these data provide further evidence for the importance of reducing q_s. They also show that in the up-modulator case the noise originating at f_{p+s} (where q_{p+s} is much larger than q_s) is not very important in comparison with f_s noise, which is amplified in the up-conversion process.

These preliminary experiments have shown the feasibility and to some extent the practicability of low-noise amplification using variable capacitance diodes. There is little doubt that better circuits will be developed which will make use of improved diodes to produce amplification with noise figures significantly lower than those reported here.

G. F. Herrmann
M. Uenohara
A. Uhlir, Jr.
Bell Telephone Labs.
Murray Hill, N. J.

A Low-Noise Wide-Band Reactance Amplifier*

The object of this letter is to summarize the results of a study of a reactance amplifier operated in the sum-frequency mode. We define a reactance amplifier generically as one in which a periodically varying reactance—obtained here by driving a nonlinear capacitor with a local oscillator—is used to provide amplification of a relatively weak signal. Amplifiers (and also modulators, demodulators, oscillators, harmonic and subharmonic generators, and bistable stages) of this generic type have a long history.[1] Current interest in such amplifiers is based principally on the expectation that low-noise amplification should be obtainable if a nearly ideal reactance is used for the essential element.[2] Our results indicate that very low effective input noise temperatures are indeed realizable.[3]

In the sum-frequency mode of operation, the reactance amplifier constitutes an amplifying modulator in which the output is derived from a circuit resonant to the sum of the local oscillator frequency and the (usually much lower) signal frequency. The amplified output signal is subsequently demodulated in a receiver whose input is tuned to the sum frequency. This form of reactance amplifier also has been described as a noninverting modulator,[4] as an up-converter,[5] and as a noninverting conversion transducer.[6] The maximum available power gain of the amplifier, obtained with a lossless input circuit, is equal to the ratio of sum frequency to signal frequency. Thus appreciable gain is possible; this is important because the minimum achievable over-all effective input noise temperature is set by the post-receiver effective input noise temperature divided by the available gain. The amplifier is inherently stable because the real part of its input and output admittances is positive. Furthermore, its frequency response is formally identical with that of two tuned circuits coupled by a capacitance, so that the overcoupled condition can be utilized if desired. For these reasons, high gain and wide bandwidth are compatible, in contrast with the performance of the negative conductance forms of reactance amplifier designated as parametric amplifiers.[7-9]

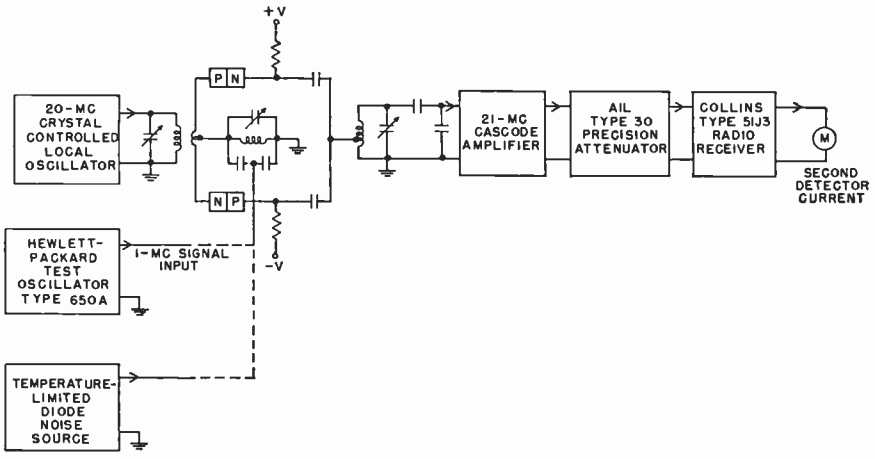


Fig. 1—Setup for measurement of balanced reactance amplifier characteristics.

To ascertain how closely theoretical performance could be achieved, a low-frequency amplifier was set up. The balanced arrangement that was used, shown schematically in Fig. 1, is a low-frequency version of the conventional balanced microwave mixer. It has the desirable features of ease of tuning, cancellation of local oscillator noise, and elimination of local oscillator current through the input and output circuits. The essential nonlinear reactance element is provided by the transition capacitance of a reverse-biased junction diode.[10] Two approximately identical commercially available (Hoffman 1N470) silicon diodes were used in the balanced arrangement. At the low operating frequencies for which the amplifier was designed, these diodes provide a nonlinear capacitance that is nearly ideal with respect to Q and to low reverse current, even at room temperature. The amplifier was supplied by a 600-ohm signal generator and coupled by an output circuit to the essentially open-circuit input terminals of a cascode amplifier. The design was approximately optimized to achieve a minimum over-all effective input noise temperature.

The main performance characteristics are listed in Table I.

TABLE I

Signal frequency	1 mc
Local oscillator frequency	20 mc
Output frequency	21 mc
Bandwidth, referred to signal frequency	10 per cent
Available power gain*	10 db
Amplifier effective input noise temperature*	30°K
Over-all effective input noise temperature	40°K

* These values include the degradation of the output circuit, which was included in the measurement for convenience.

The effective input noise temperatures in Table I were obtained with diodes and circuits operating at room temperature. Supplementary measurements made with the input circuit immersed in liquid nitrogen indicate that thermal noise originating in that circuit is a large part of the residual noise.

Extension of this study to higher frequencies is in progress. It is planned to submit a fuller account of our work at a later date.

We wish to thank A. Barone for helping us with the experimental work.

B. Salzberg
E. W. Sard
Airborne Instruments Lab., Inc.
Mineola, N. Y.

* Received by the IRE, April 4, 1958.

[1] A useful general reference is *Proc. Symp. Nonlinear Circuit Analysis, April 25–27, 1956*, Polytechnic Institute of Brooklyn, Brooklyn, N. Y.; 1957. Many of the papers included in this reference have extensive bibliographies.

[2] A. van der Ziel, "On the mixing properties of non-linear condensers," *J. Appl. Phys.*, vol. 19, pp. 999–1006; November, 1948.

[3] A recently proposed IRE Standard defines effective input noise temperature as $T_e = (F-1)T_0$, where F = noise factor and $T_0 = 290°K$.

[4] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—Part I. General energy relations," Proc. IRE, vol. 44, pp. 904–913; July, 1956.

[5] A. Uhlir, Jr., "Two-terminal *p-n* junction devices for frequency conversion and computation," Proc. IRE, vol. 44, pp. 1183–1191; September, 1956.

[6] C. F. Edwards, "Frequency conversion by means of a nonlinear admittance," *Bell Sys. Tech. J.*, vol. 35, pp. 1403–1416; November, 1956.

[7] H. Heffner and G. Wade, "Noise, Gain and Bandwidth Characteristics of the Variable Parameter Amplifier," presented at the Electron Devices Meeting, Washington, D. C.; November 1, 1957.

[8] S. Bloom and K. K. N. Chang, "Theory of parametric amplification using nonlinear reactances," *RCA Rev.*, vol. 18, pp. 578–593; December, 1957.

[9] T. J. Bridges, "A parametric electron beam amplifier," Proc. IRE, vol. 46, pp. 494–495; February, 1958.

[10] L. J. Giacoletto and J. O'Connell, "A variable-capacitance germanium junction diode for uhf," *RCA Rev.*, vol. 17, pp. 68–85; March, 1956.

Measurement of the Correlation Between Flicker Noise Sources in Transistors*

The flicker noise sources in a transistor can be represented by a noise emf e_s in series with the emitter junction and a noise current generator i in parallel with the collector junction. (See Fig. 1.) According to Fonger's theory[1] of flicker noise in transistors these two quantities should be strongly correlated. It is the aim of this note to present measurements of this correlation.

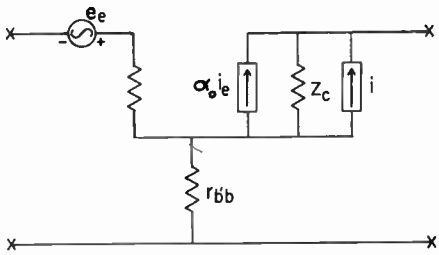


Fig. 1—Flicker noise generators in a junction transistor.

In order to present[2] the measurements in terms of the quantities e_s and i, we put $e_s = e_s' + e_s''$, where e_s' represents the part of e_s that is fully correlated with i and e'' the part of e_s that is uncorrelated with i. If the total flicker noise is now represented by an equivalent emf e_{if} in series with the emitter, we have:

$$e_{if} = e_s'' + \frac{i}{\alpha_o}\left[R_s + R_{eo} + r_{b'b} + \frac{\alpha_o e_s'}{i}\right]. \quad (1)$$

Introducing the flicker noise resistance R_{nf} by equating:

$$\overline{e_{if}^2} = 4kTR_{nf}\Delta f, \quad (1a)$$

one obtains

$$R_{nf} = R_{fl} + g_{fl}(R_s + R_{eo} + r_{b'b} + R_{fc})^2 \quad (2)$$

where R_s is the source impedance, and

$$\overline{e_{fl}''^2} = 4kTR_{fl}\Delta f;\quad \overline{i^2}/\alpha_o^2 = 4kTg_{fl}\Delta f;$$

$$R_{fc} = \frac{\alpha_o e_s'}{i} = \frac{\alpha_o \overline{e_s' i}}{\overline{i^2}}. \quad (2a)$$

Here R_{fl} is the uncorrelated emitter noise resistance, g_{fl} the flicker noise conductance and R_{fc} the correlation resistance. Hence, if $\sqrt{R_{nf} - R_{fl}}$ is plotted as a function of the source impedance R_s, one should obtain a straight line:

$$\sqrt{R_{nf} - R_{fl}} = \sqrt{g_{fl}}(R_s + R_{eo} + r_{b'b} + R_{fc}). \quad (3)$$

The straight line intercepts the zero axis at:

$$R_s = R_{so} = -(R_{eo} + r_{b'b} + R_{fc}). \quad (4)$$

In the case where e_s and i are fully correlated, one has $R_{fl} = 0$ and hence

$$\sqrt{R_{nf}} = \sqrt{g_{fl}}(R_s + R_{eo} + r_{b'b} + R_{fc}). \quad (3a)$$

* Received by the IRE, March 24, 1958. This work was supported by U. S. Signal Corps Contract.

[1] W. H. Fonger, "A determination of $1/f$ noise sources in semiconductor diodes and transistors," in "Transistors I," RCA Labs., Princeton, N. J., pp. 239–297; 1956.

[2] A. van der Ziel, "Noise in junction transistors," this issue, p. 1019.

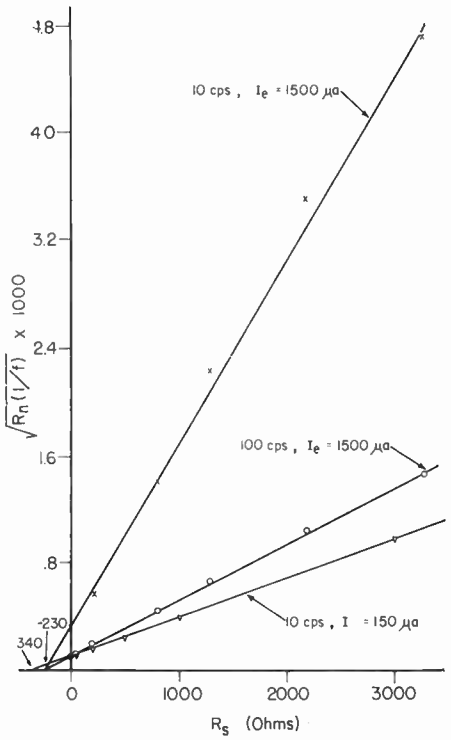


Fig. 2—$\sqrt{R_{nf}}$ vs R_s for a 2N105 transistor.

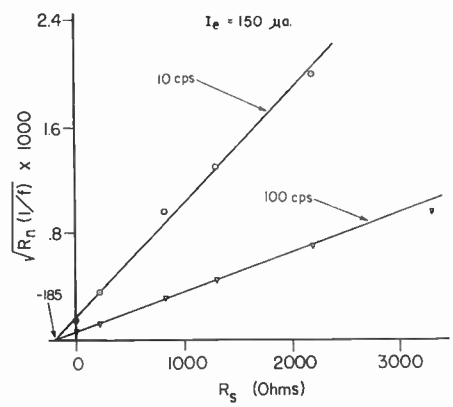


Fig. 3—$\sqrt{R_{nf}}$ vs R_s for a silicon p-n-p transistor.

Measurements of $R_{nf}R_s$ have been made for several transistors. Each experimental value of R_{nf} was obtained by measuring the noise spectrum of the transistor with R_s as a parameter and then subtracting the frequency-independent portion of the noise resistance from the total noise resistance at the frequency of interest.

The results of measurements on a 2N105 p-n-p germanium transistor are shown on Fig. 2; the results of measurements on a silicon p-n-p transistor are shown on Fig. 3. The fact that these curves are linear indicates that R_{fl} is practically zero, so that the noise generators are completely correlated. The slope of the curves increases with decreasing frequency and with increasing current; this is as expected for flicker noise.

Looking at Fig. 2, both the 10 cps, 1500 μa and 100 cps, 1500 μa curves intercept the R_s axis at $R_s = -230$ ohms. When I_e was decreased to 150 μa the intercept moved to −340 ohms. Direct measurement of $r_{b'b}$ for this transistor gave $r_{b'b} = 180$ ohms. The difference between this value and the intercept is attributed to the term $(R_{eo} + R_{fc})$ in (4).

The curves of Fig. 3, measured at $I_e = 150$ μa, show $R_{so} = -185$ ohms. This is also the value obtained for $r_{b'b}$ by direct measurement.

The current dependence of R_{eo} is well known ($R_{eo} \approx 26/I_e$, where I_e is the dc emitter current in ma). The current behavior of R_{fc} is not as well established; the results expected from Fonger's theory are given by van der Ziel.[2]

The measurements indicate that the quantities e_s and i are fully correlated within the limit of accuracy of the measurement.

The author is grateful to Dr. A. van der Ziel for suggesting this experiment and for much helpful advice.

EUGENE R. CHENETTE
Electrical Engineering Dept.
University of Minnesota
Minneapolis, Minn.

On the Effect of Base Resistance and Collector-to-Base Overlap on the Saturation Voltages of Power Transistors*

A recent note[1] presented a method of measuring collector and emitter resistances. This method is an excellent one for transistors having primarily extrinsic base resistance. It is the purpose of this letter to show the differences which are observed on transistors having a relatively large intrinsic base resistance, or having an extended collector to base overlap diode shunting the base resistance to the collector.

The case discussed here is recognized by the fact that the collector voltage, in the saturation region, is clamped to a value just below the base driving voltage in a grounded emitter circuit.[2] Fig. 1 shows the simplified lumped equivalent circuit assumed here. Emitter and base resistances are those normally determined from the geometry of the transistor. We neglect any additional extrinsic base resistance, as its effect is readily added to the base voltage in a more detailed analysis. We assume an ideal transistor with (external) body resistances R_B, R_B, and R_C shunted at the base contact by a base-to-collector diode D with body resistance R_D. The actual case of distributed resistances, particularly in diffused layer transistors or bonded contacts, differs only slightly from the results obtained from the lumped model.

In the practical cases where the base resistance, due to spreading effects or the large

* Received by the IRE, March 4, 1958. This work was supported by the Industrial Mobilization Div., U. S. Signal Corps, Contract No. DA-36-039-SC-72714.

[1] B. Kulke and S. L. Miller, "Accurate measurement of emitter and collector resistances in transistors," PROC. IRE, vol. 45, p. 90, January, 1957.

[2] H. G. Rudenberg and G. Franzen, "An alloy type medium power silicon transistor," 1957 IRE NATIONAL CONVENTION RECORD, pt. 3, pp. 26–31.

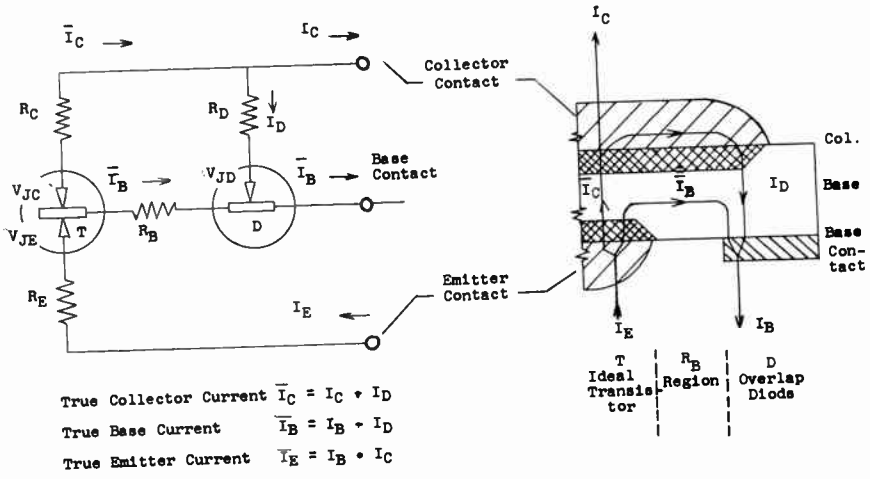


Fig. 1—Equivalent circuit.

contact spacing, is relatively large, the collector voltage can only drop to a value just below the (intrinsic) base voltage referred to the emitter. Then the contact overlap diode D conducts and clamps the collector voltage, so that part of the base input current is directed to the collector contact.

We therefore obtain

$$V_{BE} = R_E I_E + V_{JE} + R_B \bar{I}_B$$
$$= R_E(I_B + I_C) + V_{JE} + R_B(I_B - I_D). \quad (1)$$

Substituting the nominal current gain of the ideal transistor $\beta = \bar{I}_C/\bar{I}_B$ we obtain the relations

$$\bar{I}_B = \frac{I_B + I_C}{\beta + 1}$$
$$\bar{I}_C = \frac{\beta}{\beta + 1}(I_B + I_C)$$
$$I_D = \frac{\beta I_B - I_C}{\beta + 1}, \quad (2)$$

thus leading to

$$V_{BE} = V_{JE} + \left(R_E + \frac{R_B}{\beta + 1}\right)(I_B + I_C), \quad (3)$$

and as V_{JE} is primarily a function of the emitter current $(I_B + I_C)$, we may readily test this result for various combinations of I_B and I_C. Fig. 2 illustrates this result on an alloy transistor having negligible emitter and collector resistances. Eq. (3) contrasts with the case of a transistor without overlap diode. $(R_D = \infty.)$

$$V_{BE} = V_{JE} + R_E(I_B + I_C) + R_B I_B. \quad (4)$$

It is readily realized that both equations agree at full collector current $I_C = \beta I_B$ but differ at $I_C < \beta I_B$ down to $I_C = 0$ where (3) gives a much lower base voltage, which is actually observed.

Now the collector voltage may be derived from either current path

$$V_{CE} = V_{BE} - V_{JD} - R_D I_D$$
$$= V_{BE} - V_{JD} - R_D \frac{\beta I_B - I_C}{\beta + 1}. \quad (5)$$

At the knee of the saturation curve, the current through D is neglible, for $I_C \cong \beta I_B$ and the last term drops out (provided $R_D \cong R_C$). This leads directly to $V_{CE} = V_{BE} - V_{JD}$, and

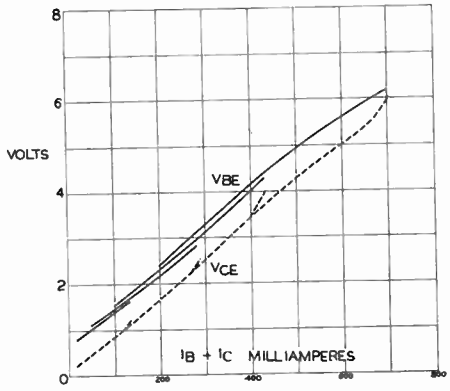


Fig. 2—Saturation voltage.

at low forward currents $(I_D = 0)$ V_{JD} is of the order of 0.2–0.3 volt (germanium) or 0.5–0.7 volt (silicon).

Interpretations of measurements taken at zero collector current differ considerably from the results of Miller. The base resistance and overlap diode lead to current circulating through the transistor even though the external current I_C is zero. Here we have at $I_C = 0$

$$V_{CE} = (V_{JE} - V_{JC}) + R_E I_B + R_C \bar{I}_C$$
$$= (V_{JE} - V_{JC}) + R_E I_B + R_C I_B \frac{\beta}{\beta + 1} \quad (6)$$

so that the resulting resistance is here

$$R_E + \frac{\beta}{\beta + 1} R_C$$

and not just R_E as in Miller's case.

Alternately we may solve from (5) to obtain

$$V_{CE} = (V_{JE} - V_{JD}) + \left(R_E + \frac{R_B}{\beta + 1}\right) I_B$$
$$- R_D I_B \frac{\beta}{\beta + 1}. \quad (7)$$

Here the effective resistance, after subtracting the other junction voltages, is given by

$$R_E + \frac{R_B}{\beta + 1} - \frac{\beta R_D}{\beta + 1}. \quad (8)$$

The choice between (6) and (8) is slight as (5) and (7) are equal and are two equations for solving the network currents.

For cases of negligible R_D there remains an added resistance $R_B/\beta + 1$ obtained from the same experimental conditions which, in the absence of such base resistance, gives only R_E. In the example of Fig. 2 the base resistance can be directly measured between two opposite base tabs, and leads to $R_b = 50$ ohms for both tabs in parallel. It is known that R_E is less than 0.1 ohm for this unit, so that the measured collector saturation resistance of 8.3 ohms is entirely due to the base resistance overlap to the collector.

Similar effects are noted in cases where R_C or R_D are large compared to R_B/β, and these cases can be calculated from analogous derivation. In all the cases where the collector is only driven into apparent saturation by being clamped to the base voltage, the storage time is correspondingly reduced to the pulse recovery time of the base-to-collector diode. The overlap diode here takes the place of an external diode in preventing operation in the (true)[3] saturation region of the transistor.

H. Gunther Rudenberg
Transitron Electronic Corp.
Wakefield, Mass.

[3] J. J. Ebers and J. L. Moll, "Large signal behavior of junction transistors," Proc. IRE, vol. 42, pp. 1761–1772; December, 1956.

Voltage Feedback and Thermal Resistance in Junction Transistors*

It is well known that the static characteristics of a transistor only can be used to give an approximate value of the transistor small signal parameters provided the effects of heating, which invariably occur to some extent, are kept to negligible proportions. However, it is not always appreciated that the voltage-feedback characteristic is dominated by thermal effects at all normal operating currents of a transistor, and only at collector currents of several microamps in low-level transistors or a few milliamps in power transistors can these effects be ignored. It is the purpose of this note to point out that this effect of transistor heating can be used conveniently to measure a transistor's thermal resistance.

For the measurement of thermal resistance it is necessary to consider the transistor in the common-base configuration so that emitter current can be held constant. Two kinds of feedback may be distinguished whose sum represents the total observed feedback. Firstly, there is the usual electrical feedback resulting from collector space-charge layer widening and extrinsic base resistance as described by Early.[1]

* Received by the IRE, December 30, 1957; revised manuscript received, March 27, 1958.

[1] J. M. Early, "Effect of space charge layer widening in junction transistors," Proc. IRE, vol. 40, pp. 1401–1406; November, 1952.

Thus,

$$h_{rb}\text{ (electrical)} = \left.\frac{\partial V_{eb}}{\partial V_{cb}}\right|_{I_e,T_e} \quad (1)$$

where T_e is the emitter junction temperature.

Using Early's notation and results,

$$h_{rb}\text{ (electrical)} = (r_b' + r_b'')g_c' \quad (2)$$

and if emitter efficiency is unity,

$$r_b'' = r_e'/2(1-\alpha'). \quad (3)$$

To a first approximation r_b', the extrinsic base resistance, and α' the common-base current gain are independent of I_c, while the emitter conductance $1/r_e'$, and g_c', the collector conductance, are proportional to it. The variation of h_{rb}(electrical) with I_c is consequently as shown in Fig. 1, curve (a).

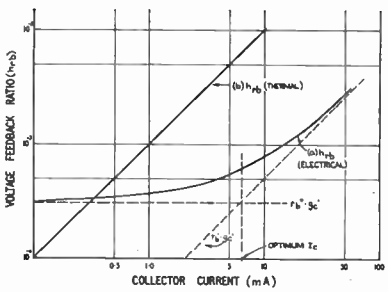


Fig. 1—The dependence of voltage feedback ratio upon emitter current.

It will be explained later that the optimum current for thermal resistance measurement is at the point where h_{rb}(electrical) changes from being constant to being proportional to I_c, that is, where $r_b'=r_b''$.

Secondly there is the thermal feedback. In computing its value the electrical feedback will temporarily be ignored.

Thus,

$$h_{rb}\text{ (thermal)} = \left.\frac{\partial V_{eb}}{\partial V_{cb}}\right|_{I_e} = \left.\frac{\partial V_{eb}}{\partial T_e}\cdot\frac{\partial T_e}{\partial P}\cdot\frac{\partial P}{\partial V_{cb}}\right|_{I_e} \quad (4)$$

where P is the total power dissipated in the transistor. Now, since $I_e \propto I_{eo}\exp(qV_{eb}/kT)$ and I_{eo} increases by about 10 per cent per degree centigrade, it follows that when I_e is constant in germanium transistors

$$\left.\frac{\partial V_{eb}}{\partial T_e}\right|_{I_e} \approx 0.0025 \text{ volts/°C}. \quad (5)$$

If it is assumed that in thermal equilibrium the emitter junction is at the same temperature as the collector junction, then $\partial T_e/\partial P=\theta$, the thermal resistance of the transistor.

Assuming that all the heat is dissipated at the collector junction,

$$\left.\frac{\partial P}{\partial V_{cb}}\right|_{I_e} = I_c + V_{cb}\frac{\partial I_c}{\partial V_{cb}} \approx I_c. \quad (6)$$

(The signs of the right-hand sides of (5) and (6) refer to n-p-n transistors. For p-n-p transistors they are both negative.)

Eq. (4) can now be written

$$h_{rb}\text{ (thermal)} = 0.0025\cdot\theta\cdot I_c. \quad (7)$$

The variation of h_{rb} (thermal) is shown in Fig. 1, curve (b).

Since h_{rb} (total) $=h_{rb}$ (thermal) $+h_{rb}$ (electrical) and since for the measurement of θ only the thermal feedback is required, it is desirable to reduce the electrical feedback as far as possible. Above the optimum current previously referred to the electrical feedback remains a constant and minimum fraction of the total feedback-voltage ratio, so that measurements of h_{rb} should be made within this region. However, since it is the change in V_{eb} that is measured, it is desirable to change V_{cb} (and hence V_{eb}) by as much as possible; I_c therefore should be kept as low as possible so that at the maximum value of V_{cb}, the transistor rated power dissipation, is not exceeded. The value of I_c at which $r_b'=r_b''$ often roughly optimises these conflicting requirements.

The time constants of the two kinds of voltage feedback are very different. The electrical feedback remains constant up to at least a few hundred cycles in most transistors, but the thermal feedback only assumes the value given in (7) after sufficient time has elapsed for thermal equilibrium to be reached, which may be a matter of minutes. It is clear therefore that to determine θ from h_{rb} it is necessary to use a dc method of measurement. (Typically thermal feedback contributes about 10 per cent of the modulus of the total feedback at 10 cycles per second, whilst at dc it may contribute 95 per cent).

The curves given in Fig. 1 are typical of a low-level transistor measured at dc. They refer to a transistor in which $r_b'=100$ ohms, $\alpha'=0.98$, $g_c'=0.5$ μmho when $I_c=1$ma and $\theta=0.4$°C/mw. The relative importance of thermal feedback can readily be seen.

The effect also, of course, appears in the common emitter configuration, but here the thermal feedback ratio is negative although the electrical feedback remains positive. The characteristics shown in Fig. 2, which are

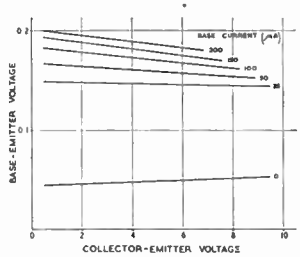


Fig. 2—Common-emitter voltage feedback characteristics.

taken directly from published data, show clearly a negative slope except at the lowest current level. Evidently, therefore, these curves simply record the heat generated by varying the collector voltage and may be misinterpreted if this fact is not appreciated.

J. J. Sparkes
British Telecommun. Res.
Taplow Court,
Taplow, Bucks., England

Microwave Transients from Avalanching Silicon Diodes*

During a study of diffused p-n junction silicon diodes, it was noticed that in a certain range of breakdown current appreciable energy was being radiated in the 9 kmc region. The particular diodes under study were fabricated in a way which results in a maximum impurity gradient in the center of the diode, with the gradient falling off rapidly away from the center. This situation is favorable for breakdown occurring in a single point, with a resulting single microplasm. This was confirmed by observation of a single set of pulses which disappear and are replaced by steady conduction as current is increased.[1]

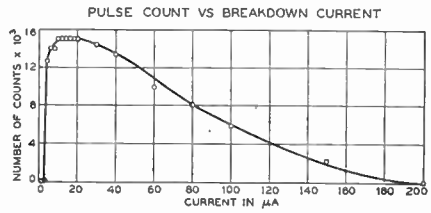


Fig. 1—Pulse count vs average current. The qualitative shape of this curve shows that the breakdown occurs in a single microplasma.

Fig. 1 shows pulse count vs current, with steady conduction setting in at 150 μa.

It was desired to relate in time the occurrence of the pulses and radiation of 9-kmc energy. For this purpose, the circuit of Fig. 2 was arranged and photographs of single simultaneous traces on each oscilloscope were taken. The 545 oscilloscope recorded the occurrence and duration of the current pulses and the 517A recorded the output of the 9-kmc receiver.

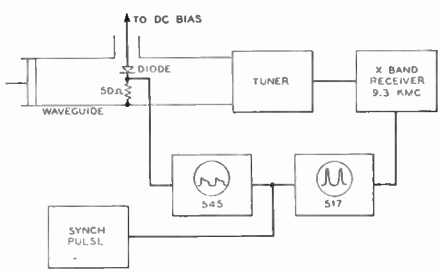


Fig. 2—Block diagram of circuit used for measurements of current and receiver output. The synchronizing pulse was arranged to give single oscilloscope traces which were photographed. The different trigger delay of the two oscilloscopes was accounted for by placing the same signal on the vertical axes and comparing the resulting traces.

Fig. 3 is a composite tracing taken from the oscillograph picture. The traces are rearranged in this way to allow the use of the same time axis in each case. The simultaneity of the turn-on of the current pulse and the radiation of energy is beyond doubt.

The width of the output pulses from the receiver is about ½ μsec in agreement with the receiver bandwidth of about 1 mc. It is interesting to note that the turn-off does not

* Received by the IRE, March 21, 1958.

[1] K. G. McKay, "Avalanche Breakdown in Silicon," *Phys. Rev.*, vol. 94, p. 877; May, 1954.

produce a radiation, indicating that it is much slower than the turn-on of the pulses. This is an expected result since the turn-off time is controlled by passive circuit parameters while the turn-on is controlled by the low dynamic impedance of the microplasma.

Fig. 3—Simultaneous oscillograph tracings of diode current and output of microwave receiver. The dotted curve is the current and shows a slight initial overshoot. The overshoot as well as the observed rise and fall times are characteristic of the measuring circuit. The solid curve is the 9-kmc receiver output. In every case the pulse rise results in a pip from the receiver.

By comparison with an argon lamp noise source, it was determined (for another one of the diodes) that the average power at 9 kmc was 32 db above thermal noise when the pulse rate N was 5×10^6 per second. If the leading edge of the pulse is a step Δi in current, the energy per pulse is

$$\frac{B(\Delta i)^2R}{2\pi^2f^2},$$

where R is the effective impedance shunting the step, and B is bandwidth. The ratio to thermal noise is therefore

$$\frac{(\Delta i)^2RN}{2\pi^2f^2kT}.$$

The pulse height for this diode was 400 μa. For an effective shunting resistance of 10^4 to 10^5 ohms (the estimated value for the tuning that was used), the calculated ratio of average power to thermal power is in the range of 20 to 30 db in good agreement with the measured value.

Thus it is a reasonable conclusion that the radiated energy is the Fourier component in the 9-kmc region. This conclusion requires that the current step occur in the order of 10^{-11} or less seconds, which is in agreement with previous calculations.

The authors wish to acknowledge the help of Dr. N. G. Cranna, who directed the fabrication of the diodes.

J. L. Moll
A. Uhlir, Jr.
B. Senitzky
Bell Telephone Labs.
Murray Hill, N. J.

A Harmonic Generator by Use of the Nonlinear Capacitance of Germanium Diode*

Until recently, silicon diodes have been used for the millimeter wave harmonic generator. The writer used the gold-bonded germanium diode as the harmonic generator of 48 Gc[1] from 24 Gc and fairly good results were obtained.

* Received by the IRE, February 3, 1958.
[1] 1 Gc = 10^9 cycles.

The harmonic generator used in our experiment is of so-called "open guide" type; that is to say, where a germanium crystal and a gallium doped gold whisker are in direct contact in a 48 Gc waveguide. After the whisker is contacted to the crystal, a forming voltage is applied to make a gold-bonded contact. Then a suitable external negative bias voltage must be supplied to the diode in order to obtain the maximum harmonic power. An output of 1.4 mw at 48 Gc is obtained as the maximum output power at the 80-mw input of 24 Gc and the best efficiency obtained is −15.8 db at 10-mw input power. A typical variation of 48 Gc output power and diode current vs supplied negative bias voltage is show in Fig. 1. A remarkable fact is that the maximum output power is obtained in the region where negative diode current flows. Fig. 2 shows the voltage-current characteristic of the diode which is used in this experiment.

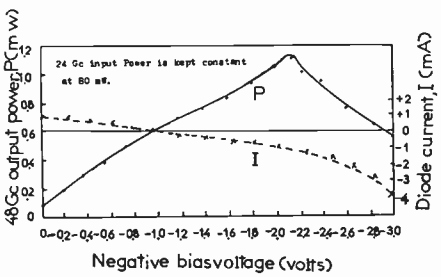


Fig. 1—48 Gc output power and diode current vs supplied negative bias voltage.

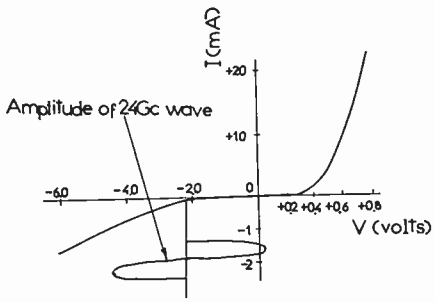


Fig. 2—Voltage-current characteristic of the diode.

This suggests that the most part of the amplitude of 24 Gc wave operates in the reverse region of the diode as shown in Fig. 2. It is believed, therefore, that the harmonic generation is caused by the nonlinearity of diode's barrier capacitance, because the nonlinearity of the reverse resistance is much smaller than that of the forward resistance.

The diode impedance in 24 Gc was measured, and it was found that in the forward bias region of 0 to 0.6 volt only the resistance component varied and in the reverse bias region of 0 to −2.0 volts the resistance component remained constant, while the reactance component varied appreciably. These results might support the above-mentioned reasoning that the harmonic generation is due mainly to the nonlinear barrier capacitance of the germanium diode.

Another merit of this harmonic generator is its large power-handling capacity. This diode could withstand the input power of the order of 300 mw without causing any noticeable change in the diode character. Millimeter wave of several milliwatts in power will be obtained if sufficiently higher input power is available.

I wish to thank Y. Nakamura, Dr. B. Oguchi, and M. Watanabe of the Electrical Communication Laboratory for helpful discussions and suggestions.

Shoichi Kita
Electrical Communication Lab.
Nippon Telegraph and
Telephone Public Corp.
Tokyo, Japan

On Junctions Between Semiconductors Having Different Energy Gaps*

Recently there has been some interest in the properties of junctions between two semiconductors having different energy gaps,[1] such as silicon-germanium alloy of two different compositions. In general, the concentration of holes and electrons will be different on each side of the junction.

Of course, this is also true of a junction between two regions of the same semiconductor, but with different doping densities. In that case, however, there is the condition that the product of hole and electron density be the same on each side of the junction, at equilibrium. This result can be considered as arising from the compensating effects of current due to diffusion and due to the field at the junction; the electron current J_n, for instance, is given by

$$J_n = D_n\left\{\frac{dn}{dx}+\frac{q}{kT}E(x)n\right\} \qquad (1)$$

and a similar equation for J_p. D_n is the diffusion constant, n the electron density, coulombs per unit volume, q the electronic charge, $E(x)$ the field, T the temperature. If $J_n=J_p=0$, by eliminating $E(x)$ from the two resulting equations, one obtains pn =constant. Thus, the junction can be described in terms of the hole and electron densities on each side, and the junction field; and the width of the energy gap, which is taken as fundamental in some treatments, here has disappeared, merely determining, or being a statement of, the electron and hole concentrations on each side of the junction.

When the junction is between two different semiconductors, one can write

$$\left.\begin{aligned} J_n &= D_n\left\{\frac{dn}{dx}+\frac{q}{kT}E_n(x)n\right\}\\ J_p &= D_p\left\{-\frac{dp}{dx}+\frac{q}{kT}E_p(x)p\right\}\end{aligned}\right\}. \qquad (2)$$

E_n and E_p are the fields as "seen by" electrons and holes, respectively, and are not

* Received by the IRE, January 13, 1958.
[1] H. Kroemer, "Quasi-electric and quasi-magnetic fields in nonuniform semiconductors," *RCA Rev.*, vol. 18, pp. 332–342; September, 1957.

necessarily the same. In fact, if $J_n = J_p = 0$, one can obtain

$$\left.\begin{aligned}\int \{E_p(x) - E_n(x)\}\,dx &= C_- + \frac{kT}{q}\ln(pn)\\ \int \{E_p(x) + E_n(x)\}\,dx &= C_+ + \frac{kT}{q}\ln\left(\frac{p}{n}\right)\end{aligned}\right\} \quad (3)$$

C_- and C_+ are constants. It is apparent that if pn is different in different places, E_p and E_n will be different. In fact, the variation in pn implies a variation in the width of the energy gap, and E_p and E_n can be interpreted in terms of the slope of the lower and upper edges of the forbidden band, respectively.

These influences which are different for charges of one sign than for those of the other have been called "quasi-electric" fields.[1] They may appear at first sight to be something of a very strange or "quantum" nature. It may be worthwhile, accordingly, to point out that something like this may be seen to arise from quite simple and elementary considerations.

Suppose that the two kinds of semiconductor have different dielectric constants; in general, this will be true. Then a charge $\pm q$, near the junction, will experience an image force F, given by[2]

$$F = \frac{1}{4\pi\epsilon_1}\left(\frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2}\right)\frac{q^2}{4a^2} \quad (4)$$

ϵ_1 and ϵ_2 being the permittivities of the two semiconductors (the charge is in semiconductor 1) and a the distance of the charge from the junction. The force is such as to drive charges of either sign into the semiconductor whose permittivity is greater. Thus, this image force could be regarded as a "quasi-electric" field, E_p and E_n being equal in magnitude and opposite in sign. Thus, effects of this sort can arise on very simple classical grounds. It is interesting to notice, in this connection, that there is often a correlation between dielectric constant and energy gap in semiconductors. Fig. 1 shows this relation for the series carbon (diamond), germanium, silicon, and grey tin. The values used are from information believed to be the most recent. It is apparent that the energy gap decreases with increasing dielectric constant, more or less as the square. This correlation is in the right direction, for one would expect carriers to concentrate in the material of higher dielectric constant, just as one expects their density to be greater in the material of smaller energy gap.

Analogous "quasi-magnetic" fields could arise similarly.[1] Consider a charge moving parallel to a junction between two semiconductors of different magnetic permeabilities. Again there would be an image force, in a direction perpendicular to the plane of the junction. The force would be in the same sense for charge of both signs, just as happens with an ordinary magnetic field; but, since the square of the velocity of the charges would enter into the relations, the "quasi-magnetic" field "seen" by various carriers would be proportional to their respective mobilities.

[2] G. P. Harnwell, "Principles of Electricity and Electromagnetism," 2nd ed., McGraw-Hill Book Co., Inc., New York, N. Y., pp. 72–74; 1949.

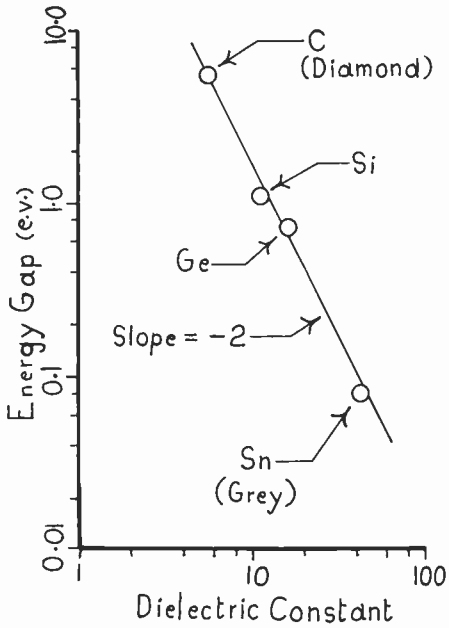


Fig. 1—Energy gap E_g vs dielectric constant for several semiconductors of similar crystal structure. There is approximately an inverse square relationship.

In conclusion, there is a point in connection with all this which it may be interesting to consider. Suppose that there is a junction between two different semiconductors. The one will have a fairly wide energy gap, and be (say) n type, for definiteness. The other semiconductor will have a quite narrow energy gap. Then, as can be seen in Fig. 2, the form of the energy zones will be about the same whether the material with the narrow energy gap is p type, n type, or intrinsic. Accordingly, it might as well be intrinsic, and junctions of this sort might have the generic name e-i_+ (extrinsic to super intrinsic). This would add another generic type of junction to the p-n junction and the L-H (low to high doping, such as n to super-n) junction.[3]

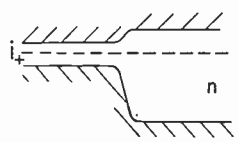


Fig. 2—Energy band scheme of a particular case of an e-i_+ (here an n-i_+) junction. The Fermi level is shown dashed, and the diagram is plotted in terms of electron energy, as is usual.

An n-i_+ junction, as shown in Fig. 2, might be expected to have some interesting properties. For instance, the hole density on the i_+ side would be greater than that on the n side. Accordingly, if the i_+ side is driven positive, it will inject holes into the n side. This should give a result similar to the forward characteristic of a p-n junction. If, on the other hand, the i_+ region is driven negative, since the i_+ region has a greater density of electrons than has the n region, the characteristic would be similar to that of an n-n_+ junction.[3]

It is interesting to think that a contact between a metal and a semiconductor might be considered as a limiting case of an e-i_+ junction, the metal being the i_+ part, in which the energy gap in the i_+ material had disappeared entirely. Ordinarily, one does not think of a metal as an intrinsic material. However, for every conduction electron above the Fermi level in the metal, there should be a vacancy below. While it is not suggested that these vacancies take part in metallic conduction to any appreciable extent, they might be able to act like holes to the extent of being injected into a contacting semiconductor. Certainly, many metal point contacts on semiconductors act in the way suggested above for e-i_+ junctions, and it may be interesting to think of the point contacts as a special case of these junctions.

Of course, there may be surface or interface effects, such as barriers, double layers of charge, etc., at any junction between two different materials. The effects considered here would be in addition to any interface effects.

Some of the work on which the discussion is based was done when the writer was at the National Research Council, Ottawa, Canada.

H. L. Armstrong
Pacific Semiconductors, Inc.
Culver City, Calif.

[3] J. B. Arthur, A. F. Gibson, and J. B. Gunn, "Carrier accumulation in germanium," *Proc. Phys. Soc. B*, vol. 69, pp. 697–704; July, 1956.
——, ——, ——, "Current gain at L-H junctions in germanium," *Proc. Phys. Soc. B*, vol. 69, pp. 705–711; July, 1956.

Arc Prevention Using *P-N* Junction Reverse Transient*

In the November issue I found the above article[1] by Miller.

The survey of literature indicates that the method of using semiconducting devices for arc suppression has been developed by earlier investigators and actual devices have been built using this principle. In particular I would like to quote the paper by Parrish[2] of the International Rectifier Corporation, El Segundo, Calif. This paper discusses in great detail the design and theory of arc quenching devices by means of germanium or selenium diodes. It also contains references to previous work in the same field. Since Miller does not mention the above referenced work, I assume that his duplication of effort stems from the unawareness of similar earlier work.

Berthold Zarwyn
Chief Physicist
Link Aviation, Inc.
Binghamton, N. Y.

* Received by the IRE, January 29, 1958.
[1] W. Miller, Proc. IRE, vol. 46, pp. 1546–1547; November, 1957.
[2] F. W. Parrish, "Arc suppression with semiconductor devices," *Elec. Mfg.*, vol. 57, pp. 127–131, 344–346; June, 1956.

Improved Keep-Alive Design for TR Tubes*

I was very interested to read the above paper by Dr. Gould.[1] At first glance, his explanation of intermittent crystal failure behind a TR switch appears basically opposed to that previously advanced;[3] his mechanism is an anode effect whereas the glow-arc transition is essentially a cathode effect. I hope, however, to show that his results are not inconsistent with our work on this subject. First it is relevant to summarize the evidence in favor of the "cathode effect" hypothesis.

1) If the keep-alive voltage is observed on a free running oscilloscope, occasional instabilities are observed which correspond to a drop to a few tens of volts. This is fairly well established[2,3] and has been observed in many laboratories. This instability is consistent with the discharge behavior in a glow-arc transition.

2) Initially, with a clean, well-baked assembly, these instabilities do not happen. Their occurrence seems to be bound up with the formation of an oxide film on the keep-alive, which typically occurs in the first 100 hours of life.[4]

3) With rhodium keep-alives, no oxide films and no instabilities of this type are observed.[2]

4) In argon alone[4] or argon plus hydrogen[5] these instabilities do not occur.

All these results are mainly applicable to the "abnormal glow" condition of keep-alive discharge. Gould's keep-alives were in the "normal glow" condition, where transitions from glow to arc are very unlikely to occur.[4] Our experiments with "normal glow" keep-alives were discontinued because of the practical success of using two keep-alives in the abnormal glow condition. Dr. Gould's ingenious solution of the main problem of the normal glow keep-alive discharge, namely localizing the discharge in the required position, would be expected to give prolonged crystal protection. There is thus no scientific disagreement between us.

I should like to ask three questions on points that were not specifically mentioned in the article. First, is the low power performance of the TR cell as a function of life adversely affected by this arrangement? Sputtered (charged) particles follow lines of force and thus should be deposited on the outside of the cone where their effect on matching and insertion loss could be great. Second, does the interaction loss change very greatly as the water vapor is cleaned up? Third, are relaxation oscillations, which usually occur at some time with normal glow discharges, present or troublesome?

D. Walsh
Engineering Lab.
Oxford University
Oxford, Eng.

* Received by the IRE, May 7, 1957.

[1] L. Gould, Proc. IRE, vol. 45, pp. 530–533; April, 1957.

[2] T. J. Bridges, P. O. Hawkins, and D. Walsh, "Keep-alive instabilities in a TR switch," Proc. IRE, vol. 44, pp. 535–538; April, 1956.

[3] J. C. French, "Electrode deterioration in transmit-receive tubes," *J. Res. Nat. Bur. Stand.*, vol. 45, pp. 310–315; October, 1950.

[4] D. Walsh, A. W. Bright, and T. J. Bridges, "Electrode deterioration in keep-alive discharges in transmit-receive switches," *Brit. J. Appl. Phys.*, vol. 7, pp. 31–35; January, 1956.

[5] D. Walsh, "The electron affinity of hydrogen in a microwave gas discharge," *J. Electronics*, vol. 1, pp. 444–448; January, 1956.

WWV Standard Frequency Transmissions*

Since October 9, 1957, the NBS radio stations WWV and WWVH have been maintained as constant as possible with respect to atomic frequency standards maintained and operated by the Boulder Laboratories, NBS. On October 9, 1957, the U.S.A. Frequency Standard was 1.4 parts in 10^9 high with respect to the frequency derived from the UT 2 second (provisional value) as determined by the U. S. Naval Observatory. The atomic frequency standards remain constant and are known to be constant to 1 part in 10^9 or better. The broadcast frequency can be further corrected with respect to the U.S.A. Frequency Standard as indicated in the table below. This correction is *not* with respect to the current value of frequency based on UT 2. A minus sign indicates that the broadcast frequency was low.

The WWV and WWVH time signals are synchronized; however, they may gradually depart from UT 2 (mean solar time corrected for polar variation and annual fluctuation in the rotation of the earth). Corrections are determined and published by the U. S. Naval Observatory.

WWV and WWVH time signals are maintained in close agreement with UT 2 by making step adjustments in time of precisely plus or minus twenty milliseconds on Wednesdays at 1900 UT when necessary; no step adjustment was made at WWV and WWVH.

WWV Frequency†

1958 March 1500 UT	Parts in 10^9
1	−2.6
2	−2.6
3	−2.7
4	−2.8
5	−2.9‡
6	−2.9
7	−2.9
8	−3.0
9	−3.0
10	−3.0
11	−3.0
12	−2.9
13	−2.9
14	−2.8
15	−2.8
16	−2.8
17	−2.9
18	−3.0
19	−3.0
20	−3.1
21	−3.1
22	−3.1
23	−3.1
24	−3.1
25	−3.1
26	−3.1
27	−2.9
28	−2.8
29	−2.7
30	−2.7
31	−2.7

† WWVH frequency is synchronized with that of WWV.

‡ Decrease in frequency of 0.5×10^{-9} at 1900 UT at WWV.

W. D. George
Radio Standards Lab.
Natl. Bur. of Standards
Boulder, Colo.

* Received by the IRE, April 16, 1958.

Contributors

R. W. Aldrich was born in Melrose, Mass. on December 9, 1920. He received the B.S. degree in electrical engineering from Columbia University in 1942 and the Ph.D. degree in applied physics from Cornell University in 1951. He was a wartime staff member of the Radiation Laboratory at Massachusetts Institute of Technology. From 1951 to 1956 he was a member of the General Electric Electronics Laboratory staff and subsequently has been a member of the General Electric Advanced Semiconductor Laboratory, Syracuse, N. Y.

R. W. ALDRICH

Dr. Aldrich is a member of the American Physical Society and Sigma Xi.

❖

James B. Angell (S'45–A'47–M'55–SM'57) was born in Staten Island, N. Y., on December 25, 1924. He received the S.B. and S.M. degrees of the cooperative course in electrical engineering in 1946 and the Sc.D. degree in electrical engineering in 1952 at the Massachusetts Institute of Technology, Cambridge, Mass.

J. B. ANGELL

From 1946 to 1951, he worked as a research assistant in the Research Laboratory of Electronics at M.I.T., studying noise in tracking radars. Since 1951, he has been with the Research Division of Philco Corporation, where he has worked on microwave communication systems, various military projects, and transistor circuit applications. At present, he is a research section manager leading a group studying the basic circuit application and evaluation of transistors and other solid-state devices.

He serves on the IRE and AIEE Committees on Solid-State Devices and the Joint Subcommittee on Semiconductor Devices.

Dr. Angell is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the AIEE, and the Franklin Institute.

❖

Stephen J. Angello (SM'52) was born in Haddonfield, N. J., on March 2, 1918. He received three degrees at the University of Pennsylvania, Philadelphia: the B.S.E.E. in 1939, the M.S.E.E. in 1940, and the Ph.D. in physics in 1942, with thesis work in the field of semiconductors.

Dr. Angello has been with the Westinghouse Electric Corporation in various capacities since 1942. First, he served as a research engineer in the Headquarters Laboratories, then as the manager of a development group in the Materials Engineering Department, then as Engineering Manager of the Youngwood, Pennsylvania Semiconductors Department, and recently as consulting engineer reporting to the Director of Research.

S. J. ANGELLO

He was responsible for the silicon rectifier development work which resulted in the formation of the Semiconductor Department. Currently, he is directing thermoelectric device development work at the Research Laboratories in Pittsburgh, Pa.

Dr. Angello is a member of the American Physical Society and the American Institute of Electrical Engineers.

❖

John Bardeen was born in Madison, Wis., on May 23, 1908. He received the B.S. and M.S. degrees in electrical engineering from the University of Wisconsin and the Ph.D. degree in mathematical physics from Princeton University. He was a post-doctoral Fellow at Harvard University and from 1938 to 1941 taught physics at the University of Minnesota. During the war years, he was at the Naval Ordnance Laboratory in Washington, D. C. From 1945 to 1951, he was a research physicist at the Bell Telephone Laboratories. Since then, he has been Professor of physics and of electrical engineering at the University of Illinois. Most of his research has been on the theory of the solid state.

Dr. Bardeen is co-winner of the 1956 Nobel Prize in Physics.

For a photograph of Dr. Bardeen, see page 952 of this issue.

❖

George Bemski was born in Warsaw, Poland, on May 20, 1923. He received the B.A. degree in 1949 and the Ph.D. degree in physics in 1953, both from the University of California in Berkeley. In 1953, he became a member of the technical staff at Bell Telephone Laboratories, Murray Hill, N. J. Since then he has been investigating the electrical properties of silicon crystals.

G. BEMSKI

Dr. Bemski is a member of the American Physical Society and Sigma Xi.

Walter H. Brattain was born in Amoy, China, on February 10, 1902. He is co-winner of the 1956 Nobel Prize in Physics with Dr. John Bardeen and Dr. William Shockley. The prize was awarded for "investigations in semiconductors and the discovery of the transistor effect."

He received the B.S. degree from Whitman College in 1924, the M.A. degree from the University of Oregon in 1926, and the Ph.D. degree from the University of Minnesota in 1929. Since 1929, Dr. Brattain has been a member of the Bell Laboratories technical staff, where he is a research physicist. His chief field of research has been the surface properties of solids. His early work was concerned with thermionic emission and adsorbed layers on tungsten. He continued on into the field of rectification and photoeffects at semiconductor surfaces, beginning with a study of rectification at the surface of cuprous oxide and followed by similar studies of silicon. Since World War II, he has continued in the same line of research with both silicon and germanium.

His chief contributions to solid-state physics have been the discovery of the photoeffect at the free surface of a semiconductor, the invention of the point-contact transistor jointly with Dr. Bardeen, and work leading to a better understanding of the surface properties of germanium, undertaken first with Dr. Bardeen and later with Dr. C. G. B. Garrett.

He has also carried on research on electron collisions in mercury vapor, piezoelectric frequency standards, magnetometers, and infrared detectors.

Dr. Brattain received the honorary Doctor of Science degree from Portland University in 1952, from Whitman College and Union College in 1955, and from the University of Minnesota in 1957. In 1952, he was awarded the Stuart Ballantine Medal of the Franklin Institute, and in 1955 the John Scott Medal. The degree at Union College and the two medals were received jointly with Dr. John Bardeen, in recognition of their work on the transistor.

Dr. Brattain is a Fellow of the American Physical Society, the American Academy of Arts and Sciences, and the American Association for the Advancement of Science, and a member of the Franklin Institute, Phi Beta Kappa, and Sigma Xi.

For a photograph of Dr. Brattain, see page 953 of this issue.

❖

Mason A. Clark (SM'56) was born in Ladysmith, Wis., on August 13, 1921.

Mr. Clark received the B.S. degree in electrical engineering in 1947 from Northwestern University, Evanston, Ill., and the M.S. degree in physics in 1949 from the same institution.

Following a period of work on gas discharges, spectroscopy, and infrared detectors

at Northwestern University, he was employed as a member of the technical staff of Bell Telephone Laboratories, Murray Hill, N. J., from 1952 to 1956. During this time at Bell Laboratories he was engaged primarily in the development of low- and high-frequency power transistors.

M. A. CLARK

He is now head of Development at Pacific Semiconductors, Inc., Culver City, Calif.

❖

Esther M. Conwell was born in New York, N. Y. She received the B.A. degree, magna cum laude, from Brooklyn College in 1942, the M.A. degree from the University of Rochester in 1945, and the Ph.D. degree from the University of Chicago in 1948. She has taught physics at the Universities of Rochester and Chicago and at Brooklyn College. During the war, she worked at Western Electric Company and on the Manhattan District Project. After a year at Bell Laboratories, Dr. Conwell joined the Research Laboratories of Sylvania Electric Products, Inc., in 1952. During this period she has been working on the theory of semiconductors.

E. M. CONWELL

❖

Reimer Emeis was born on August 15, 1920, in Rendsburg, Germany. He was conferred the degree of physics by the Darmstadt Institute of Technology, Darmstadt, Germany, in 1952.

R. EMEIS

Since that time, Mr. Emeis has been affiliated with the Siemens-Schuckertwerke Aktiengesellschaft Laboratory in Pretzfeld, West Germany. He did some of the original work in the floating zone technique for silicon. Presently, he is concerned with silicon devices.

❖

Lawrence J. Giacoletto (S'37–A'42–M'44–SM'48–F'58) received the B.S. degree in electrical engineering from Rose Polytechnic Institute, Terre Haute, Ind., in 1938 and the M.S. degree in physics from the State University of Iowa, in 1939, where he held an appointment as research assistant. From 1939 to 1941, he was a postgraduate student and teaching fellow in the Department of Electrical Engineering at the University of Michigan, where he received the Ph.D. degree in 1952.

Dr. Giacoletto was associated with the Collins Radio Company during the summers of 1937 and 1938 and with the Bell Telephone Laboratories in 1940. From 1941 to 1945, he was on military duty with the Signal Corps Engineering Laboratory concerned with development activities in the fields of radio, navigational, and meteorological direction-finding equipments. He returned to inactive status as a Major in the Signal Corps Reserve, in May, 1946. From June, 1946 to June, 1956, he worked on electronic and semiconductor devices as a research engineer with RCA Laboratories, Princeton, N. J. Since then, he has been with the Scientific Laboratory, Ford Motor Company, Dearborn, Mich., as manager of the Electrical Department.

L. J. GIACOLETTO

Dr. Giacoletto is a member of the American Association for the Advancement of Science, American Physical Society, Gamma Alpha, Iota Alpha, Phi Kappa Phi, Tau Beta Pi, and Sigma Xi.

❖

Rolf Gremmelmaier was born in Berghausen (Baden), Germany on November 10, 1924. He received the Dr.rer.nat. degree from the Technische Hochschule Karlsruhe, Germany in 1952.

R. GREMMELMAIER

Since 1952 Dr. Gremmelmaier has been employed at the Solid State Department of the Research Laboratory of the Siemens-Schuckertwerke, Erlangen, Germany.

❖

John Halpern (M'56) was born in New York, N. Y. on July 27, 1931. He received the B.S. degree in 1951 from the City College of New York, New York, N. Y. and the M.A. degree in physics in 1955 from Columbia University, New York, N. Y.

J. HALPERN

He spent 1953 to 1955 as a research assistant with the IBM Watson Scientific Laboratory, where he was engaged in the study of ferroelectric crystals. In 1954–1955 he was an instructor in the Physics Department in the evening school of the City College of New York. Since 1955 he has been a member of the device section of the solid state physics group at the Lincoln Laboratory, M.I.T., Lexington, Mass., and has been working on high frequency devices.

Mr. Halpern is a member of the American Physical Society and Phi Beta Kappa.

❖

Herbert W. Henkels (M'53–SM'57) was born on August 15, 1922 in Philadelphia, Pa. He received the Bachelor's degree in chemical engineering in 1943. After naval electronics training and service, he received the Ph.D. degree in physics from the University of Pennsylvania. He worked as a research associate in the Moore School of Electrical Engineering on thermodynamics of resistors, computing machine reading and writing problems, selenium semiconductors, rectifiers, and photocells.

H. W. HENKELS

He has been with Westinghouse Electric Corporation since 1953. In 1955 he was manager of the Semiconductor Engineering Section; in 1956, manager of development in the Semiconductor Department; and from 1956 to the present, was manager of advanced development in the same department. He is currently working on silicon and germanium rectifiers and transistors, switches, III–V compounds and devices, and long-range development liaison.

Dr. Henkels is a member of AIEE, Sigma Tau, Pi Mu Epsilon, and Sigma Xi.

❖

Robert A. Henle (S'49–A'51–M'57) was born in Virginia, Minn., in 1924. He received the B.S. degree in electrical engineering from the University of Minnesota in 1949 and the M.S.E.E. degree from the same university in 1951. He joined the International Business Machines Corporation in, 1951, as a technical engineer in a research group concerned with the possible application of transistors to computers. Continuing to work with investigations of transistor circuits. he was placed in charge of a transistor servoamplifier project in 1955.

R. A. HENLE

He is now manager of transistor circuit development in the Product Development Laboratory, Poughkeepsie, N. Y.

❖

Adolf Herlet was born on August 30, 1921, in Mettlach, Germany. He received the degree of Dipl. phys. from the

University of Göttingen in 1948, and the degree of Dr. rer. nat. from the University of Bonn in 1951.

A. HERLET

After he was awarded the doctoral degree, he became associated with the Siemens-Schuckertwerke Aktiengesellschaft Laboratory in Pretzfeld, West Germany.

He has been engaged in the theory and development of semiconductor devices.

❖

David D. Holmes (A'48–M'55) was born in Portland, Me., on August 12, 1926. He received the B.S.E.E. degree from the University of Maine in 1946, and from 1946 to 1947, was with Hazeltine Electronics Corporation, where he worked on television receiver circuits. From 1947 to 1949, he was an instructor in electrical engineering at the University of Nebraska. He received the S.M.E.E. degree from Massachusetts Institute of Technology in 1950, following which he joined the technical staff of RCA Laboratories, Princeton, N. J. From 1950 to 1957, Mr. Holmes worked in the fields of color television and transistor applications. In 1957 he assumed his present position of manager, Radio Research Laboratory, RCA Laboratories, with activities at Princeton, N. J., and at Rocky Point and Riverhead, N. Y.

D. D. HOLMES

Mr. Holmes is a member of Sigma Xi.

❖

Nick Holonyak, Jr. (S'51–A'55) was born November 3, 1928 in Zeigler, Ill. He received the B.S. in E.E. degree in 1950, the M.S. degree in 1951, and the Ph.D. degree in 1954 from the University of Illinois. While a graduate student at the University of Illinois, he was a teaching assistant, a research assistant in microwave tubes and in semiconductors and transistors, and held the Texas Instruments Fellowship in transistor physics. He joined the transistor development department of Bell Telephone Laboratories in 1954 and worked on diffused-impurity silicon devices. He was inducted into the U. S. Army in 1955, and served with the Signal Corps at Fort Monmouth, N. J., and in Japan. Following his discharge from the Army, he joined the Advanced Semiconductor Laboratory of the General Electric Company and has been involved in studies of power and signal silicon *p-n-p-n* structures.

N. HOLONYAK, JR.

Dr. Holonyak is a member of the American Physical Society, Sigma Xi, Eta Kappa Nu, Tau Beta Pi, Pi Mu Epsilon, and Phi Kappa Phi.

❖

John E. Iwersen (M'58) was born in New York, N. Y., on April 2, 1928. He received the B.S. degree in chemistry from Wagner College, Staten Island, N. Y., in 1949. He then attended The Johns Hopkins University receiving the M.A. and Ph.D. degrees in chemistry in 1951 and 1955, respectively.

J. E. IWERSEN

Since 1955, Dr. Iwersen has been a member of the technical staff at Bell Telephone Laboratories, Murray Hill, N. J., working on silicon high-frequency power transistors.

He is a member of the American Chemical Society.

❖

Dietrich A. Jenny, for a photograph and biography, please see page 790 of the April, 1958 issue of PROCEEDINGS.

❖

Pieter J. W. Jochems was born on February 15, 1924 in Nijmegen, The Netherlands. During 1942 he was engaged in chemical work at the "Nyma" factory, and he subsequently studied radio and television engineering.

P. J. W. JOCHEMS

Mr. Jochems joined the Philips Research Laboratories at Eindhoven in 1945, and since 1948, he has been working in the transistor predevelopment group.

❖

Frank Keywell was born in Detroit, Mich., on March 16, 1923. He received the B.A. degree in physics-meteorology from the University of California at Los Angeles in 1944. During World War II, he served with the United States Air Force in the fields of weather forecasting and aircraft operations. He received the M.S. degree in 1951 and the Ph.D. degree in 1954 in physics from the University of Southern California. His thesis work, done under an ONR contract, was related to high-vacuum sputtering of metals due to ion bombardment and a statement of the phenomenon in terms of radiation damage in metals. After graduation, he joined the Bell Telephone Laboratories, Murray Hill, N. J., where he has done work in the field of silicon junction transistor development by means of gaseous diffusion techniques.

F. KEYWELL

Dr. Keywell is a member of the American Physical Society and Sigma Xi.

❖

John G. Linvill was born in Missouri in 1919. He received the A.B. degree in mathematics from William Jewell College in 1941, and the S.B., S.M., and Sc.D. degrees in electrical engineering from M.I.T. in 1943, 1945, and 1949, respectively.

J. G. LINVILL

From 1949 until 1951, Dr. Linvill was assistant professor of electrical engineering at M.I.T. From 1951 until 1955, he was a member of the technical staff at the Bell Telephone Laboratories, where he did research in transistor circuits. Since 1955, Dr. Linvill has been at Stanford University, where he is Professor of electrical engineering and head of the solid-state electronics program.

Dr. Linvill is a member of Sigma Xi and Eta Kappa Nu.

❖

Ian M. Mackintosh (M'57) was born on April 20, 1927, in Nottingham, England. He received the B.Sc. and Ph.D. degrees from the University of Nottingham in 1953 and 1956, respectively. His doctoral work was on the physical properties of intermetallic semiconductors.

I. M. MACKINTOSH

He spent several years with the Engineering Department of the British Post Office, engaged in communication engineering. In 1956, he joined the technical staff of the Bell Telephone Laboratories in Murray Hill, N. J., where he has been concerned primarily with the development of silicon switching devices.

Dr. Mackintosh is the author of several publications in the field of intermetallic semiconductors.

He is a member of the British Institution of Electrical Engineers.

Oscar W. Memelink was born in Blinju, Indonesia, on April 11, 1927. He received a degree in physical engineering in 1952 from the Technische Hogeschool of Delft.

O. W. MEMELINK

From 1952 to 1955, he was engaged in work on photoconducting compounds for the Institute of Applied Scientific Research. Since 1955, he has been with Philips Research Laboratories at Eindhoven, The Netherlands, where he is concerned with transistor physics.

❖

George C. Messenger (A'53) was born in Brattleboro, Vt. on July 20, 1930. He received the B.S. degree in physics from Worcester Polytechnic Institute, Worcester, Mass., in 1951 and the M.S. degree in electrical engineering from the University of Pennsylvania, Philadelphia, Pa., in 1957.

G. C. MESSENGER

Mr. Messenger has been a member of the Philco Research Division since 1951. He has worked primarily on the design of semiconductor devices, on microwave mixer diodes, and on high-frequency transistors. He is currently research section manager in charge of several areas including transistor development and microwave diodes. He has published several papers related to these studies.

Mr. Messenger is a member of the American Physical Society.

❖

John L. Moll (A'51–M'57) was born on December 21, 1921, in Wauseon, Ohio. He received the B.S. degree in engineering physics in 1943 and the Ph.D. degree in electrical engineering in 1952 from Ohio State University, Columbus, Ohio.

J. L. MOLL

He was associated with the mathematics department and the Research Foundation of Ohio State University before joining the technical staff of Bell Telephone Laboratories, Murray Hill, N. J., in 1952. There he has been concerned with the development and analysis of silicon transistors and silicon *p-n-p-n* crosspoints. He presently is in charge of a group working on solid-state electronics research.

He has accepted an appointment as associate professor of electrical engineering at Stanford University, Stanford, Calif., beginning with the Fall, 1958, term.

Dr. Moll is a member of Sigma Xi.

❖

Jack A. Morton (A'36–SM'53–F'53) was born in St. Louis, Mo., on September 4, 1913. He received the B.S.E.E. degree from Wayne University, Detroit, Mich. in 1935 and the M.S. degree from the University of Michigan, Ann Arbor, Mich., in 1936. He joined the Bell Telephone Laboratories, Murray Hill, N. J. in 1936 and continued with part-time postgraduate studies at Columbia University, New York, N. Y.

J. A. MORTON

His first work at the Bell Telephone Laboratories was on coaxial cable and microwave-amplifier circuit research. During the first part of World War II, he was a member of a group engaged in improving the signal-to-noise performance of radar receivers. Mr. Morton later entered vacuum-tube development and designed a microwave tube now being used in the New York-San Francisco microwave system for telephone and television transmission. Shortly after the invention of the transistor, he was placed in charge of all developmental work on this and related semiconductor devices. In 1952, he was appointed assistant director of electronic apparatus development, including transistor and related developments, and in October, 1953, director of transistor development in charge of all development on semiconductor devices. In June, 1955, he was named director of device development with the over-all responsibility of a large general department engaged in fundamental development and development for manufacture of electron tubes, solid-state devices, and electromechanical and passive devices.

In the several fields in which he has been active, Mr. Morton has been noted for producing engineering ideas, practical achievements, and deeper physical understanding. For instance, in his early work, which was of electrical circuit engineering nature, he pioneered in several directions. He conceived and realized high-frequency transmission measuring methods which exceeded the ranges of previously existing means by no less than a factor of ten and which were widely used for measuring broad-band feedback amplifiers for ten years. Later he aided the development of the grid-return amplifier at microwave frequencies, an achievement which was not only interesting scientifically but also, by extending the range of early radars, had an important effect on the course of World War II in the Pacific.

In 1948 he received an Eta Kappa Nu honorable mention award as a nationally outstanding young engineer, and in 1951 a University Alumni Award from Wayne University. He also received a Distinguished Alumnus Citation in October, 1953 from the University of Michigan. He was awarded the degree of Doctor of Science by the Ohio State University in 1954. in 1956, he also received an honorary degree of Doctor of Science from Wayne University. He has given some forty-five papers and lectures and fifteen patents have been issued in his name.

Dr. Morton is a member of Eta Kappa Nu, Alpha Delta Psi, MacKenzie Honor Society, Phi Kappa Phi, and Sigma Xi.

❖

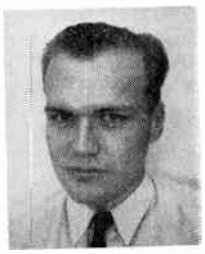
Herbert Nelson received the B.S. degree from Hamlin University, St. Paul, Minn., in 1927, and the M.S. degree in physics from the University of Minnesota in 1929. From 1929 to 1930, he was an engineer with the Westinghouse Lamp Company in Bloomfield, N. J. He transferred to RCA in 1930 and was engaged in electronic research and development at the Tube Plant in Harrison, N. J., until 1953, when he joined the semiconductor activity at the RCA Laboratories in Princeton, N. J.

H. NELSON

Mr. Nelson is a member of Sigma Xi and of the American Physical Society.

❖

James T. Nelson was born in Portland, Ore., on March 28, 1924. He received the B.A. degree in physics from Reed College in 1950. In 1952 he received the M.A. degree and in 1955 the Ph.D. degree from the University of Oregon, where he also worked as a research fellow on the development of a scintillation counter for the detection of soft X rays and an investigation of the electrical and optical properties of intermetallic compounds.



J. T. NELSON

Dr. Nelson became a member of Bell Telephone Laboratories, Murray Hill, N. J., in 1955 and has been working on the development of a silicon power transistor and diffused base germanium transistors.

He is a member of Sigma Xi.

❖

Donald F. Page (A'54) was born in Hamilton, Ont., Canada, on July 2, 1929. He received the degree of B.Sc. in electrical engineering from Queen's University, Kingston, Ontario, in 1951.

He then was employed by the Canadian Defence Research Board in the Electronics

Laboratories, Defence Research Telecommunications Establishment, Ottawa. In October, 1954, he registered in the electrical engineering department of the Imperial College of Science and Technology, University of London, London, England, where he received the College Diploma (DIC) in 1955 for work concerning the use of transistors in superregenerative circuits.

D. F. PAGE

He is now a research assistant at Imperial College, on leave from the Canadian Defence Research Board.

❖

William J. Pietenpol (M'52) was born on July 15, 1921 in Colorado. He received the B.S.E.E. degree in 1943 from the University of Colorado, Boulder, Colo. In 1946 he continued his study at Ohio University, Columbus, Ohio, where he held three fellowships: Eastman Kodak, 1946; Westinghouse, 1947; and Research Foundation, 1948. He received the Ph. D. degree in December, 1949. From 1943–1946, he was employed by RCA, Lancaster, Pa., to work on the development of photoelectric tubes. Since January, 1950, he has been with a transistor design group at Bell Telephone Laboratories, Murray Hill, N. J. Dr. Pietenpol has been primarily concerned with the development of the *p-n* junction diodes and grown *n-p-n* junction transistors. In October, 1953, he was appointed transistor development engineer and transferred to Allentown, Pa., where he was responsible for the development of transistors for manufacture. He assumed his present post, director of development-semiconductor devices, on June 1, 1955.

W. J. PIETENPOL

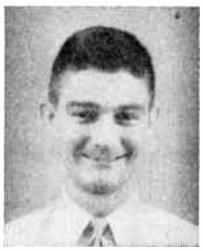
A frequent contributor to transistor literature, he has appeared before numerous technical groups to present papers relating to his work.

He is a member of Sigma Xi, Tau Beta Pi, and Phi Beta Kappa.

❖

Robert L. Pritchard (S'45–A'51–SM'55) was born in Irvington, N. J., on September 8, 1924. He graduated from Brown University, Providence, R. I., in 1946 with the B.S. Degree in engineering, and received the Ph.D. Degree in acoustics at Harvard University, Cambridge, Mass., in 1950.

From 1950–1957, he was with the General Electric Research Laboratory, Schenectady, N. Y., where he was engaged in research on the transistor from an electric-circuit point of view. In 1957, he joined Texas Instruments Inc., Dallas, Texas, where he now is a member of the Development Department engaged in studying the electric-circuit properties of semiconductor devices.



R. L. PRITCHARD

Dr. Pritchard is a member of the Acoustical Society of America, Sigma Xi, and Tau Beta Pi.

❖

Robert H. Rediker (A'53) was born in Brooklyn, N. Y. on June 7, 1924. He received the B.S. degree in electrical engineering in 1947 and the Ph.D. degree in physics in 1950 from the Massachusetts Institute of Technology, Cambridge, Mass.

R. H. REDIKER

During 1950–1951, Dr. Rediker was a research associate in cosmic rays in the Physics Department of M.I.T. In April, 1951 he became a staff member of the Lincoln Laboratory, M.I.T., Lexington, Mass., where he worked on transistorized computer circuits.

During the academic year 1952–1953, he was a research associate at the Physics Department of Indiana University, Bloomington, Ind. Since June, 1953, he has been a staff member at Lincoln Laboratory, where he now heads the semiconductor device section of the solid state physics group.

Dr. Rediker is a member of the APS and Sigma Xi.

❖

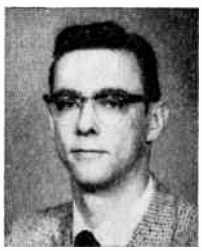
Bernard Reich (M'57) was born in New York, N. Y. on January 7, 1926. He received the B.S. degree in physics from the College of the City of New York in 1948. From 1948 to 1954, he did graduate work in electrical engineering and mathematics at Rutgers University, attending on a part-time basis.

B. REICH

Mr. Reich has been employed at the U. S. Army Signal Engineering Laboratories since 1948. From 1948 to 1954, he served as radiation physicist, and was Radiological Safety Officer during 1951–1953. At present, he is working as chief of the Device Engineering Section, Solid-State Devices Branch, Evans Signal Laboratory, Belmar, N. J.

❖

David E. Sawyer (M'55) was born in Boston, Mass., on February 6, 1927. He served with the U. S. Navy from 1945 to 1946 and later from 1950 to 1952.



D. E. SAWYER

He received the A.B. degree from Clark University, Worcester, Mass., in 1953 and the M.S. degree in physics from the University of Illinois, Urbana, Ill., in 1955.

Since June, 1955, he has been a member of the device section of the solid-state group of Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass.

❖

William Shockley (SM'51–F'55), one of three recipients of the 1956 Nobel Prize in Physics for his work on semiconductors and discovery of the transistor effect, was born in London, England, on February 13, 1910. He received the B.S. degree in 1932 from the California Institute of Technology. He continued his studies at Massachusetts Institute of Technology on a teaching fellowship, where he received the Ph.D. degree in physics in 1936.

In September of that year he joined Bell Telephone Laboratories, where his work included vacuum-tube and electron-multiplier design, radar development, solid-state physics, magnetism, and semiconductors.

In 1942, Dr. Shockley was assigned to Columbia University Division of War Research as Director of Research for the Anti-Submarine Warfare Operations Research Group. In 1944, he became a consultant to the Office of the Secretary of War.

Returning to Bell Telephone Laboratories in 1945, he became director of transistor physics, and in collaboration with W. H. Brattain and John Bardeen, worked on the development of the transistor.

At present, Dr. Shockley is director of the Shockley Semiconductor Laboratory of Beckman Instruments, Inc., Mountain View, Calif.

He was awarded the Presidential Medal for Merit in 1946 and the Morris Liebmann Award in 1952, and in 1951 was elected to the National Academy of Sciences. He received the honorary degree of Doctor of Science from the University of Pennsylvania in 1955 and from Rutgers University in 1956. He is a member of the American Physical Society, American Academy of Arts and Sciences, Tau Beta Pi, and Sigma Xi.

For a photograph of Dr. Shockley, see page 954 of this issue.

Friedolf M. Smits was born in Stuttgart, Germany, on November 10, 1924. He received the Dr. rer. nat. from the University of Freiburg, Germany, in 1950. From 1950 to 1954, he was a research associate at the University of Freiburg. In 1954, he joined the technical staff of Bell Telephone Laboratories, Murray Hill, N. J. He is now concerned with material, process, and exploratory device studies.

F. M. SMITS

Mr. Smits is a member of the American Physical Society.

❖

Eberhard Spenke was born in Bautzen, Germany, on December 5, 1905. He studied physics and mathematics at Bonn, Göttingen, and Königsberg, and was graduated as a doctor of philosophy in 1928 from the University of Königsberg. Since 1929, he has been with Siemens-Halske and Siemens-Schuckertwerke Aktiengesellschaft and, since 1946, he has been at the Pretzfeld, West Germany, Laboratory.

E. SPENKE

He has worked on acoustics, thermistors, gas discharges, the perturbation theory of betatron, and crystal rectifiers.

❖

James P. Spratt was born in Philadelphia, Pa., on September 3, 1934. He received the B.S. degree in electronic physics from St. Joseph's College, Philadelphia, in 1956.

J. P. SPRATT

Mr. Spratt joined the Philco Research Division as a junior engineer immediately after graduation. He spent one year as a member of the Physics of Metals Group, investigating such bulk semiconductor parameters as resistivity and lifetime. For the past year he has been a staff physicist, concerned primarily with the study of the effects of radiation on transistors.

❖

Jerome J. Suran (A'52–SM'55) was born in New York, N. Y., on January 11, 1926. After having served for three years with the U. S. Army during World War II, he received the B.S.E.E. degree from Columbia University, New York, N. Y., in 1949, and continued graduate studies there and at the Illinois Institute of Technology, Chicago, Ill.

J. J. SURAN

From 1949 to 1952, he was employed in the field of control systems design and development by J. W. Meaker and Co., New York, N. Y., and in the field of FM communication research and development by Motorola, Inc., Chicago, Ill. Since 1952, he has been active in the area of solid-state circuits, and is currently Manager of the Advanced Circuits Component of the Electronics Laboratory, General Electric Company, Syracuse, N. Y.

Mr. Suran is co-author of "Principles of Transistor Circuits" and "Transistor Circuit Engineering." He is a member of the AIEE and the Research Society of America.

❖

Donald E. Thomas (A'47–SM'47) was born in Hanover Township, Pa., on May 12, 1907. He was awarded the B.S. degree in electrical engineering in 1929 by Pennsylvania State University, University Park, Pa., and the M.A. degree in 1932 by Columbia University, New York, N. Y.

D. E. THOMAS

He joined Bell Telephone Laboratories in 1929, and was assigned to submarine telephone cable systems development. Just before World War II, he became engaged in the development of sea and airborne radar and continued in this field until he left for military duty in 1942. Upon rejoining Bell Laboratories in 1946, he was active in the development and installation of the first deep sea repeatered submarine telephone cable system. From 1950 to 1954, he worked on the development of transistor devices and circuits for special applications. Currently, he is making evaluation and feasibility studies of new types of semiconductor devices at Bell Laboratories in Murray Hill, N. J.

During World War II, he was a member of the Joint and Combined Chiefs of Staff Committees on Radio Countermeasures. From 1948 to 1953, he was a civilian member of the Department of Defense's Research and Development Board Panel on Electronic Countermeasures. He is a member of Tau Beta Pi and Phi Kappa Phi.

❖

Clarence G. Thornton (SM'54) was born in Detroit, Mich., on August 3, 1925. He obtained the B.S., M.S., and Ph.D. degrees in physical chemistry at the University of Michigan, Ann Arbor, Mich., in 1944, 1949, and 1952. He taught at the University of Michigan from 1949 to 1952 and was associated with a classified research program conducted by the Engineering Research Institute for the U. S. Army. In addition, he spent a year setting up high-speed computing methods for the Fourier synthesis and harmonic analysis of sinusoidal functions. From 1951 to 1954 he worked at Sylvania Electric Company as a section head in the Semiconductor Engineering Laboratory. He joined the engineering staff of Philco Corporation-Lansdale Tube Company in 1955. He is now manager of the Semiconductor Product Development Laboratories in the Lansdale Tube Company Engineering Laboratories.

C. G. THORNTON

Dr. Thornton is a member of the American Chemical Society, the American Crystallographic Association, the AAAS, Alpha Chi Sigma, Sigma Xi, Phi Lambda Upsilon, and Phi Kappa Phi.

❖

Leo J. Tummers was born in Leiden, The Netherlands, on August 22, 1922. In 1950 he received a degree in physical engineering from the Technische Hogeschool of Delft. He joined the Philips Research Laboratories in Eindhoven in 1949. Following his graduation he served two years in the Royal Dutch Navy, and then returned to Philips Laboratories, where he is presently engaged in research on semiconductor devices.

L. J. TUMMERS

Mr. Tummers is a member of the Dutch Physical Society and the Royal Institute of Engineers.

❖

Arthur Uhlir, Jr. (A'57) was born in Chicago, Ill. on February 2, 1926. He received the B.S. and M.S. degrees in chemical engineering from Illinois Institute of Technology in 1945 and 1948, respectively, and the S.M. and Ph.D. degrees in physics in 1950 and 1952, respectively, from the University of Chicago, where he was an AEC predoctoral fellow from 1949 to 1951.

A. UHLIR, JR.

He was a process analyst at Douglas Aircraft, Chicago, Ill., during 1945. From 1945 to 1948, he was engaged in fluid mechanics research at Armour Research Foundation. Since 1951, he has been in the Transistor Development Department of Bell Telephone Laboratories. There he has worked on point-contact transistor theory, semiconductor surface protection, and electrochemical properties of semiconductors, and has developed an electrolytic micromachining technique for metals and semiconductors. He is now working on microwave semiconductor devices.

Dr. Uhlir is a member of the American Physical Society, Sigma Xi, Phi Lambda Upsilon, and the American Association for the Advancement of Science.

❖

A. van der Ziel, for a photograph and biography please see page 621 of the March, 1958 issue of PROCEEDINGS.

Karel M. van Vliet was born in Dordrecht, The Netherlands, on December 27, 1929. From 1946–1956, he attended the Free University, Amsterdam, where he received the Ph.D. degree in physics in 1956. He was a research fellow at the Free University from 1953–1956, engaged in studies of electrical fluctuations and in photoconductivity. He held a post-doctoral fellowship at the University of Minnesota from 1956–1957; since 1957 he has been an assistant professor of electrical engineering.

K. M. VAN VLIET

Dr. van Vliet is a member of the American Physical Society, the Dutch Physical Society, and Gamma Alpha.

James L. Walsh (A'52) was born in Providence, R. I. in 1925. He received the B. S. degree in electrical engineering from Rhode Island State College in 1949, and the M.A. degree in physics from Hofstra College in 1952. The same year he joined the International Business Machines Corporation as a design engineer working on data processing machines. He has since had experience in the design and application of transistor circuiting for computers.

J. L. WALSH

He is now a development engineer in the IBM Product Development Laboratory, Poughkeepsie, N. Y., in charge of a transistor circuit design group.

Scanning the Transactions

A 111-year-old paper is providing circuit theorists with important new insights into the analysis of network problems. We have reference to an epoch-marking but strangely neglected paper by Kirchhoff, published in 1847, which laid down the basis of electrical circuit theory. Kirchhoff, renowned for his voltage and current laws, enunciated in this celebrated paper what is known as Kirchhoff's Rule, which concerns itself with what we know now as network topology and linear graphs. This rule, although well known to circuit theory specialists, has so many implications and uses in the analysis and synthesis of RLC networks that they have yet to be fully explored. Indeed, most of the attempts to evaluate Kirchhoff's Rule in the light of modern network theory have occurred just in the last decade. Strangely enough this famous rule has appeared in English only once, in Maxwell's 1892 edition of *Electricity and Magnetism*. Therefore, the full paper has now been translated and published, together with two exceedingly valuable interpretive discussions of it, in a current issue of TRANSACTIONS. This must be the oldest and certainly one of the most historically important papers to appear in an IRE publication.

It is interesting to note in passing that, as great as Kirchhoff's contributions were in the field of electrical circuits, he is best known to the world at large for his work in physics. His crowning achievement was the invention, with Bunsen, of the spectroscope, and with it the discovery that chemical elements could be identified by the spectral characteristics of the light they emitted during combustion. Among other important results, this discovery revolutionized astronomy and laid the foundation of the modern field of astrophysics by making it possible to analyze the composition of stars and, from Doppler shifts of spectral lines, to calculate their velocities. (G. Kirchhoff, "On the solution of the equations obtained from the investigation of the linear distribution of galvanic currents," (translated by J. B. O'Toole), IRE TRANS. ON CIRCUIT THEORY, March, 1958. L. Weinberg, "Kirchhoff's third law," *ibid.* F. Reza, "Some topological considerations in network theory," *ibid.*)

Engineering writing and speech has received increasing attention in recent years. A great many engineering companies have come to be technical publishers of no small magnitude. We know of at least one engineering organization whose publication output during World War II exceeded that of any commercial publishing house in the United States. Technical societies, too, have become publishers of the first order. IRE publications, for example, totaled more pages last year than did *Life* magazine. This upsurge in technical literature has resulted in the formation of several societies through which professional writers and editors in technical fields may exchange ideas and advance themselves and their arts.

Last year another organization was formed in this field, the IRE Professional Group on Engineering Writing and Speech. The PGEWS has now issued its first TRANSACTIONS containing the papers presented at its first national meeting last Fall. It is significant that 70 per cent of those attending the meeting were *not* professional writers or editors, but just plain ordinary engineers. Therein lies the big difference between PGEWS and the other societies—it exists primarily to serve the nonprofessionals.

It has been stated that the English language contains 600,000 words, compared to only 250,000 for German and 170,000 for Spanish. Whether this makes "good English" easier or harder to come by than other languages is a moot point. But there can be no denying that better English is needed. This need was well expressed in the keynote address at the PGEWS meeting: "Few things are more important to the engineer than the ability to write and speak clearly. It is largely through what a man says or through what he puts down on paper that we can judge his thoughts and his contributions. Sometimes things which are really good and important can be obscured through being badly expressed. It is very hard to tell sloppy thinking from sloppy writing, and at times the two may be almost the same thing. . . . Certainly, the Professional Group on Engineering Writing and Speech is representative of as important a field as there is in modern engineering and, indeed, in modern life." (J. R. Pierce, "The challenging field of engineering writing and speech," IRE TRANS. ON ENGINEERING WRITING AND SPEECH, March, 1958.)

Internationality was never more in evidence than in the current issue of PGMTT TRANSACTIONS. Some sort of record for an IRE publication was undoubtedly set when better than one-third of the 17 papers in the issue proved to be from abroad. Contributions came from Japan, England, France, India, and last but not least, a lady engineer in Poland. (IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, April, 1958.)

Home music rooms have become extremely popular in recent years among hi-fi enthusiasts. The big push in stereophonic disk recordings, expected this year, will no doubt be an added stimulus to those seeking the ultimate in music listening. It has reached a point where many music devotees have become self-styled acoustical experts, designing their music rooms with an eagle eye (or should we say "ear") for such factors as reverberation, room resonances, sound diffusion and noise. Devoted attention is given to the location and dimensions of the room, size of the windows, wall materials and room furnishings. Slanted ceilings with exposed beams and walls broken up by bay windows, alcoves, etc., are resorted to in order to avoid parallel surfaces within the room. Those who like to worry about such things as sound leaking from the room can even buy special sound attenuating doors. And the real purists go so far as to worry about noise leaks through electrical outlet boxes, heating pipes and nearby plumbing fixtures. Fortunately for the rest of us, though, good music can still be thoroughly enjoyed under the very primitive conditions of our low-fi living rooms. (W. B. Snow, "Application of acoustical engineering principles to home music rooms," IRE TRANS. ON AUDIO, November–December, 1957.)

Difficulties in implementing automation was the subject of some plain talk at last year's First National Symposium on Production Techniques. Most of the talk centered on the component parts industry and lack of standardization. A representative of a prominent components manufacturer reported that more than 75% of their orders are for less than 100 units. To make matters worse, between 25 and 40 thousand different capacitors are manufactured each month and 80 to 100 new designs created each week by this one firm. This tremendous multiplicity of types and ratings makes mechanization of the components industry almost economically impossible.

The wide variety of components in use also has a deleterious effect on the application of automatic assembling techniques. Although some progress has been made in designing components specifically for automatic insertion machinery, there remains an urgent need for greater standardization, not just of parts, but of printed wiring panels.

The above presupposes that automation is wanted and needed. The supposition is supported by the estimate that the electronics industry will be asked to produce 67% more gear in 1960 than it did in 1953, with a labor force that will be only 10% larger. However, at least one symposium speaker warned

that automation seems to be more a matter of fashion than common sense and that some of the present applications are unjustified economically. Perhaps we are spending too much effort automating the trivial and the inexpensive. (IRE TRANS. ON PRODUCTION TECHNIQUES, April, 1958.)

A new kind of dielectric modulator has been suggested by a theoretical study of electromagnetic wave propagation through a dielectric slab. If the dielectric takes the form of a ferroelectric or ferrimagnetic material, a varying external electric or magnetic field would cause a variation in the propagation velocity characteristics of the dielectric, producing a velocity modulated electromagnetic wave. It appears from the study that a novel modulator of this sort would be feasible at microwave frequencies, at least for low-power levels. (F. R. Morgenthaler, "Velocity modulation of electromagnetic waves," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, April, 1958.)

Cooling of airborne electronic equipment is one of the more serious problems facing designers today. It also offers a striking example of the growing importance of the mechanical engineer to his colleagues in the electronics field. The relatively slow and uncrowded aircraft of World War II did not pose much of an equipment cooling problem. However, as more and more electronics has been added in aircraft—much of it miniaturized, thus increasing the number of watts per cubic inch—and as modern aircraft operated at higher speeds and altitudes, the cooling of airborne electronic equipment has become an increasingly difficult problem. The answers to this problem are now coming in large measure from mechanical engineers with an extensive knowledge of heat transfer techniques. Thus conduction, radiation, natural convection, forced convection and evaporative cooling methods have become an intimate part of electronic equipment design. (Cooling Airborne Electronics Issue, IRE TRANS. ON AERONAUTICAL AND NAVIGATIONAL ELECTRONICS, March, 1958).

The family of pulse-type modulation systems now has a new member. The family is already quite large, including among its members such old-timers as PAM (amplitude), PCM (code), PCM (count), PDM (duration), PFM (frequency), PPM (position), PTM (time) and PWM (width). The new arrival is WPCM. It stands for weighted pulse-code modulation, and represents an interesting modification of its nineteen-year-old brother, pulse-code modulation. In PCM, a widely used technique in binary data transmission systems, the message signal is sampled periodically. The amplitude of each sample is then represented by a binary number which is transmitted in the form of a group of positive and negative pulses, representing binary numbers 0 and 1. In conventional PCM systems the binary pulses all have the same amplitude. Thus in receiving them over a noisy channel, the probability of error is the same for all pulses. However, the pulse representing one digit of a binary number has a different importance in reconstructing the signal than the pulse representing the next digit. In weighted PCM, therefore, the amplitudes of the pulses are adjusted, or "weighted," so that the probability of error is in keeping with the value of each pulse. While equipment complexity may dim somewhat the practicality of the system, it will nevertheless be of substantial theoretical interest as a means of improving communication system performance. (Edward Bedrosian, "Weighted PCM," IRE TRANS. ON INFORMATION THEORY, March, 1958.)

IRE's role in solving receiver radiation problems stands as an outstanding tribute to the valuable, though usually unglamorous, contributions which the IRE technical committees are rendering to the profession and industry. The spurious radiation generated by the local oscillators of FM and TV sets has long presented a worrisome problem to the industry. After some six years of investigation the FCC, in December, 1955, established limits on the maximum allowable radiation of all receivers in the 30 to 800 mc range, specifying that radiation measurements were to be carried out as prescribed by IRE standards on the subject. Thus two IRE standards are now a part of the law of the land. The FCC action has since stimulated a great deal of additional work on the part of industry on methods of reducing spurious radiation and improving the accuracy of radiation measurements. The IRE, through its Committee on Radio Frequency Interference, is continuing to play a very important role by compiling and evaluating radiation measurement data for the entire industry. It is expected that the committee's evaluations, when completed, will lead to further important improvements of the measurement art. (W. G. Peterson, "Local oscillator radiation from TV and FM sets," IRE TRANS. ON BROADCAST AND TELEVISION RECEIVERS, March, 1958.)

Report of the Secretary—1957

TO THE BOARD OF DIRECTORS
THE INSTITUTE OF RADIO ENGINEERS, INC.

Gentlemen:

Submitted herewith is the Secretary's Report for the year 1957. The statistics in it show the increasingly ramified activities of your IRE and the continued growth in all departments. Membership (see Fig. 1 and Table I) went up some 17%. Of interest was the 22% increase of those residing in foreign countries compared with 16.5% of those residing within the United States. Significantly, the number of members holding professional grades now exceeds those having non-professional status. Also, members of Professional Groups now exceed in numbers those that have membership in the IRE, because many members belong to more than one Group.

Publication volume has moved up some 20%. The production of the STUDENT QUARTERLY has been strengthened by the addition of an assistant to the headquarters editorial staff.

Technical activities continue unabated. The number of Section meetings went up almost 8%, the number of Professional Group meetings 4.6% and the number of Student Branches 12%.

Respectfully submitted,

Haraden Pratt

Haraden Pratt
Secretary

February 26, 1958

It is with deep regret that this office records the death of the following members of the IRE during the year 1957.

Fellows

Cowan, Frank A. (M'30, SM'43, F'55)
Englund, Carl Robert (A'17, F'28, L'52)
Farrington, John F. (A'19, M'29, F'31)
Finch, James Leslie (A'19, M'28, SM'43, F'54)
Graf, Alois W. (A'26, M'44, SM'45, F'55)
Jensen, John C. (M'19, SM'43, F'56, L'54)
Pierce, George Washington (M'13, F'15, L'48)
Reisz, Eugen (F'54)
Tolson, William A. (M 28, SM 43, F 56)
Webb, Wilbur L. (A'35, SM'44, F'54)

Senior Members

Arnold, Prescott N. (A'29, SM'46)
Campbell, Robert L. (M'45, SM'47)
Emery, Ernest John (M'29, SM'43)
Franks, Christopher J. (M'38, SM'43)
Freeman, Edward E. (M'40, SM'43)
Gilbert, John J. (A'16, SM'50, L'52)
Gillson, Morley H. (SM'46)
Howard, George E. (SM'47)
Junken, Lawrence H. (A'25, SM'55)
King, Marion E. (SM'52)
LeLoup, Theodore E. (SM'56)
McKee, Douglas Allen (A'45, SM'57)
Metzger, Frederick W. (SM'57)
Mumtaz Ud Din, Sheikh (A'48, SM'52)
Oliver, George Edward (M'50, SM'51)
Overhold, Ralph Jr. (A'29, SM'51)
Ready, William A. (A'25, SM'45)
Siegmund, Humphreys O. (SM'53)
Southwell, John D. (A'47, M'48, SM'53)
Stark, Rawson E. (A'35, SM'55)
Stroebel, John C., Jr. (A'14, M'26, SM'43)
Tierney, Walter Louis (M'26, SM'43)
White, William (M'50, SM'55)
Wilcox, James Franklin (A'36, SM'47)
Zarky, Bert (M'45, SM'54)

Members

Balmer, David W. (A'46, M'52)
Bell, Richard A. (A'46, M'46)
Brennan, Daniel J. (S'54, M'56)
Bruce, Kenneth C. (A'55, M'56)
Bueche, Harry S. (A'46, M'55)
Butler, Eugene R. (S'49, A'49, M'55)
Charlebois, Joseph C. (A'48, M'55)
Cotton, Thomas D. (M'54)
Dahse, Clem Alfred (S'48, A'49, M'54)
Dakin, Oron C. (A'45, M'49)
DeLong, Oscar A., Jr. (A'20, M'55)
Elliott, Clifford I. (A'45, M'55)
Fredine, Earle L. (A'41, M'45)
Frevert, Andrew W. (M'47)
Haas, Max L. (A'50, M'56)
Hedgecock, Edward B. (M'54)
Hiltner, Edward B. (S'42, A'45, M'55)
Hood, Chauncey R. (A'47, M'55)
Kelsey, Philip C. (M'53)
Korby, Irving (S'48, A'50, M'55)
Land, Barney E. (M'57)
Martin, Frank R. (M'56)
Nakamoto, Tatsuo (M'55)
Nordyke, Cutlar J. (S'54, M'56)
Nowick, Chester A. (M'53)
Owen, E. Robert (A'45, M'55)
Plusc, Igor (M'45)
Proud, Eric (M'49)
Ragni, Victor F. (A'49, M'55)
Read, James H. (S'54, M'56)
Reichert, Homer A. (A'47, M'55)
Reisman, Emil (A'46, M'55)
Richardson, Arthur W. (A'50, M'56)
Siegel, David T. (M'45)
Smith, Don C. (S'50, A'51, M'56)
Tarkoff, Philip B. (A'45, M'46)
Turnbull, John Mason, Jr. (M'56)
Whilldin, Frank W. (M'54)
Williams, John C. (A'51, M'56)
Winn, William F. (S'50, A'52, M'56)
Wood, Perry E. (A'51, M'56)
Wright, James W. (A'23, M'55)

Voting Associates

Curtis, Richard C. (A'30, VA'39)
Wallace, Gordon S. (A'17, VA'39, L'52)

Associates

Afferton, Percy C. (A'55)
Browning, William I., Jr. (A'55)
Chiappinelli, Bruno A. (A'53)
Dexter, Lyman Henry (S'57, A'57)
Douglass, William A. (A'54)
Dunn, James F. (S'52, A'53)
Erickson, Charles G. (S'52, A'54)
Finke, Walter W. (A'52)
Giolitto, Amedeo P. (A'46)
Herder, Martin R. (A'55)
Holt, Todd N. (A'53)
Hudson, Ivan (A'53)
Ito, Yoji (A'54)
Kamys, Joseph E. (A'54)
Kimons, Ellis (A'51)
Kruse, James E. (A'53)
Liewald, Vernon R. (A'53)
Masuch, Robert A. (S'50, A'51)
Ollick, Edward (A'53)
Orbell, Reginald J. (A'35)
Patrick, Anderson T. (A'54)
Perna, Frank J. (A'51)
Pesca, Frank J. (A'53)
Scheuer, Rudolph D. (A'53)
Schiesel, Ervin E. (A'56)
Snyder, Reed E. (A'41)
Spearman, John C. (A'56)
Wackid, Charles B. (A'53)

Students

Banks, Richard Norman (S'56)
Freshwater, Philip L. (S'56)
Sommer, Roger M. (S'56)
Young, James A. (S'56)

Fiscal

A condensed summary of income and expenses for 1956 is shown in Table II, and a balance sheet in Table III.

Editorial Department

PROCEEDINGS OF THE IRE

Several steps were taken to increase the value of the PROCEEDINGS to the general membership during 1957. The Editorial Board initiated a grass-roots survey of PRO-

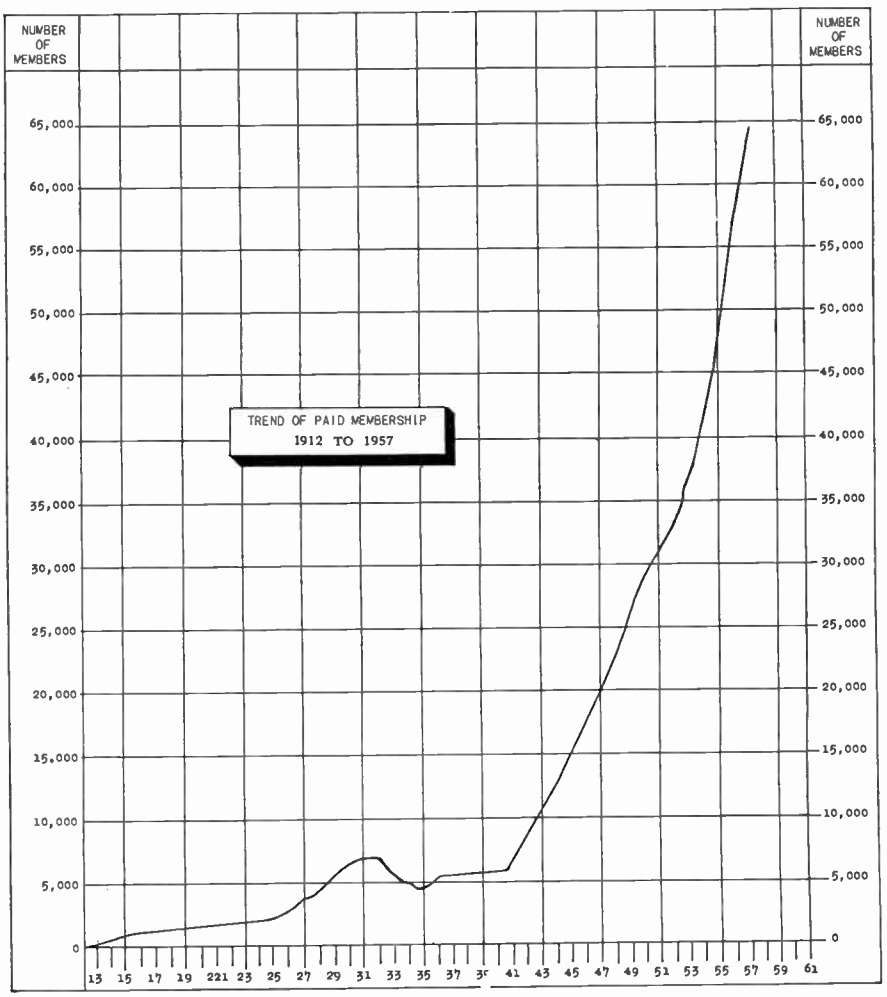


Fig. 1

TABLE I

Grade	As of Dec. 31, 1957		As of Dec. 31, 1956		As of Dec. 31, 1955	
	Number	% of Total	Number	% of Total	Number	% of Total
Fellow	700	1.1	635	1.1	565	1.2
Senior Member	7,685	11.9	6,486	11.6	5.643	11.9
Member	26,115	40.3	19,110	34.6	13,360	28.2
Associate	16,827‡	25.9	18,879†	34.0	20,492*	43.2
Student	13,446	20.8	10,384	18.7	7,328	15.5
TOTALS	64,773		55,494		47,388	

* Includes 396 Voting Associates.
† Includes 388 Voting Associates.
‡ Includes 352 Voting Associates.

TABLE II

SUMMARY OF INCOME AND EXPENSE, 1957

Income		
Advertising	$1,280,580	
Member Dues and Convention	1,306,940	
Subscriptions	154,165	
Sales Items: Binders, Emblems, etc	127,909	
Investment Income	34,707	
Miscellaneous Income	1,469	
TOTAL INCOME		$2,905,770
Expense		
Proceedings Editorial Pages	$ 401,001	
Advertising Pages	668,543	
Directory	232,702	
Section Rebates	65,757	
Student Program	107,321	
Professional Group Expense	170,647	
Sales Items	119,340	
General Operations	498,929	
Convention Cost	401 609	
TOTAL EXPENSE		$2,665,849
Reserve for Future Operations—Gross		$ 239,921
Depreciation		20,033
Reserve for Future Operations—Net		$ 219,888

TABLE III

BALANCE SHEET—DECEMBER 31, 1957

Assets		
Cash and Accounts Receivable	$ 448,883	
Inventory	40,463	
TOTAL CURRENT ASSETS		$ 489,346
Investments at Cost	1,405,446	
Building and Land at Cost	912,317	
Furniture and Fixtures at Cost	215,277	
Other Assets	62,440	
TOTAL		2,595,480
TOTAL ASSETS		$3,084,826
Liabilities and Surplus		
Accounts Payable	$ 83,343	
TOTAL CURRENT LIABILITIES		$ 83,343
Deferred Income	745,409	
Professional Group Funds on Deposit	171,436	916,845
TOTAL LIABILITIES		1,000,188
Reserve for Depreciation		64,757
Surplus Donated	595,286	
Surplus	1,424,595	
TOTAL SURPLUS		2,019,881
TOTAL LIABILITIES AND SURPLUS		$3,084,826

CEEDINGS and other publication policies, conducted by each IRE Section among its own members, to obtain the guidance of the membership with respect to present and future operations. In addition, the *Correspondence* section was expanded, with the result that 146 letters to the Editor were published in 1957 as compared with 83 the previous year. The tempo of publishing special review and historical papers also increased, with ten such papers appearing during the year.

The total PROCEEDINGS output, shown in Table IV, was lower than the previous year, due to the fact that none of the several special issues under preparation fell within the calendar year. Despite the lower PROCEEDINGS total, the over-all total for all IRE publications soared to a new high of 115 issues totaling 15,556 pages, a 20% increase over the previous year, due principally to the inauguration of a new publication, the IRE WESCON CONVENTION RECORD. It is interesting to note that over the past five years, as IRE membership doubled, publication volume has nearly tripled.

IRE CONVENTION RECORDS

The practice of publishing a CONVENTION RECORD containing papers presented at the IRE National Convention, begun in 1953, was continued. The 1957 IRE NATIONAL CONVENTION RECORD, containing 254 papers and 19 abstracts totaling 1,924 pages, was issued in ten parts. Approximately 53,000 paid members of Professional Groups received free of charge a copy of that Part pertaining to the field of interest of their Group.

A new annual publication, covering the papers presented at the Western Electronic Show and Convention (WESCON), and modelled after the IRE NATIONAL CONVENTION RECORD, was inaugurated in 1957. The new publication, called the IRE WESCON CONVENTION RECORD, contained 209 papers and 24 abstracts totaling 1,796 pages, and was issued in ten Parts. As with the NATIONAL CONVENTION RECORD, paid members of Professional Groups received a free copy of the Part pertaining to the field of their Group.

TABLE IV

VOLUME OF PROCEEDINGS PAGES

	1957	1956	1955	1954
Editorial	1,868	1,996	2,060	1,884
Advertising	2,700	2,800	2,372	2,072
TOTAL	4,568	4,796	4,432	3,956

VOLUME OF TRANSACTIONS PAGES

	1957	1956	1955	1954
Groups Publishing	24	23	21	18
No. of Issues	75	69	56	51
No. of Pages	5372	5044	3504	3714

IRE STUDENT QUARTERLY

To insure continuity of operation and maintenance of quality of the STUDENT QUARTERLY in the future, Assistant Editor Paul Lucey was added to the Editorial Department staff to take over from the Editor most of the publication burdens. Four issues, totaling 208 pages, were sent free to all IRE Student members during the year.

IRE DIRECTORY

The 1958 IRE DIRECTORY was published in October, containing 1,112 pages including covers, of which 465 were membership listings and information, and 647 were advertisements and listings of manufacturers and products.

CONFERENCE PUBLICATIONS

The *Proceedings of the 1957 Western Joint Computer Conference*, sponsored jointly by the IRE, AIEE and Association for Computing Machinery, was published by the IRE Editorial Department in June, 1957. The issue contained 244 pages. The IRE also published the 292-page *Proceedings of the 1957 Electronic Components Symposium*, sponsored by IRE, AIEE, Electronic Industries Association, and West Coast Electronic Manufacturers Association.

Technical Activities

Technical Committees

During 1957 the 26 Technical Committees and their 103 subcommittees held 274

meetings, of which 246 were held at IRE Headquarters and 28 throughout the nation.

The six Standards listed herewith, having been approved by the Standards Committee and the IRE Board of Directors, were published in the PROCEEDINGS in 1957, and reprints are now available to the public:

57 IRE 7.S1 January—
Standards on Electron Tubes: Physical Electronics Definitions, 1957.
57 IRE 14.S1 March—
Standards on Piezoelectric Crystals—The Piezoelectric Vibrator: Definitions and Methods of Measurement, 1957.
57 IRE 7.S2 July—
Standards on Electron Tubes: Definitions of Terms, 1957.
57 IRE 21.S1 August—
Standards on Letter Symbols and Mathematical Signs, 1948 (Reprinted 1957)
57 IRE 21.S2 November—
Standards on Reference Designations for Electrical and Electronic Equipment, 1957.
57 IRE 21.S3 December—
Standards on Graphical Symbols for Semiconductor Devices, 1957.

IRE is directly represented on 33 Committees of the American Standards Association and sponsors three; the ASA Sectional Committee on Radio and Electronic Equipment, C16; the ASA Sectional Committee on Sound Recording, Z57; and the ASA Sectional Committee on Nuclear Instruments, N3. Three IRE Standards received approval of the American Standards Association as American Standards in 1957, and are now available overseas through the International Standards Organization.

IRE Technical Committees actively participated in international standardization in 1957 by reviewing and preparing comments on documents for the United States National Committee of the International Electrotechnical Commission.

Appointed IRE Delegates on Other Bodies

The IRE appointed delegates to a number of other bodies for the one-year period—May 1, 1957 to April 30, 1958 (as listed on page 1440 of the October, 1957 issue of the PROCEEDINGS.

The Annual Spring Meeting of the International Scientific Radio Union (URSI) was held on May 22, 23, 24, and 25, 1957 at the Willard Hotel at the National Bureau of Standards in Washington, D. C. It was cosponsored by the following IRE Professional Groups: Antennas and Propagation, Circuit Theory, Information Theory and Microwave Theory and Techniques. The XIIth General Assembly of URSI was held on August 22 through September 5, 1957 in Boulder, Colorado. Special meetings were held on the status of the International Geophysical Year.

Numerous responses to the questions under study by the various CCIR Study Groups have been received in IRE during 1957. Lists of all material received from these organizations were distributed quarterly to the chairmen of all Technical Committees, Professional Groups and Definitions and Measurement Subcommittees. During 1957 the Executive Committee of the U. S. Preparatory Committee of CCIR held 23 meetings and the 14 Study Groups held approximately 8 meetings.

The Joint Technical Advisory Committee

The Joint Technical Advisory Committee and its Subcommittees held a total of 10 meetings in addition to the annual dinner.

Volume XIV, the cumulative Annual Report of the *JTAC Proceedings* was published in 1957. This includes in Section I—official correspondence between the Federal Communications Commission and the Joint Technical Advisory Committee (IRE-EIA). Also included were other items of correspondence pertinent to the activities of the JTAC. Section II of the report contained approved Minutes of Meetings of the Joint Technical Advisory Committee for the period July 1, 1956 to June 30, 1957.

A new Subcommittee for cooperation with the Cooperative Interference Committees (57.1) was formed on January 24, 1957 to enlist the cooperation of the IRE Sections and Professional Groups in the publicizing of the "Cooperative Interference Committees" project of the JTAC.

The JTAC Subcommittee on Study of Forward Scatter Propagation (55.1) is still working on its final report.

The JTAC Subcommittee on Study of Interference From Arc Welders (54.2) disbanded, its major objectives having been achieved. The FCC complimented JTAC for their cooperation and assistance in endeavoring to solve the reduced cycle methods of operation and for their constructive services in investigating the feasibility of reducing interference from radio frequency stabilized arc welders.

The JTAC Subcommittee on Study of Single Sideband Transmission (56.1) reported that thirty papers had been submitted for publication in the December, 1956 special issue of the IRE PROCEEDINGS. A similar study is being conducted by CCIR Study Group XIII on possible use of single sideband equipment in Aeronautical and Maritime Mobile Services which was adopted at the VIIIth Plenary Assembly in Warsaw. The Subcommittee is awaiting receipt of proposed rules and standards now being formulated by the FCC in connection with Further Proposed Rule-Making concerning single sideband utilization.

The JTAC Subcommittee on Study of Spurious Radio Emissions (52.2) presented a report on "The Principles Affecting the Probability of Serious Interference" which will be included in the next Annual Volume of the *Proceedings of JTAC*. This study by the Subcommittee has been completed and the committee disbanded.

The International Electrotechnical Commission (IEC)

The International Electrotechnical Commission held its annual meeting in Moscow July 2—July 12, 1957. IRE did not actively participate in this meeting, since there were no meetings of IEC Technical Committee 12 on Radio Communication, or Technical Subcommittee 12-1 on Measurements scheduled at the Moscow meeting.

A list of all documents and material received in the Office of the IRE Technical Secretary from the IEC was distributed to the chairmen of all Professional Groups, Technical Committees and Subcommittees.

Professional Group System

General. There are currently 28 Professional Groups operating actively within the IRE. Groups 25, 26, 27 and 28 were organized during this year and cover the specialized fields of Education, Engineering Writing and Speech, Radio Frequency Interference and Human Factors in Electronics.

Approximately 55% of all IRE members have taken advantage of the Professional Group System which now has a total membership of approximately 67,000. Included are 4,440 Student members of the IRE who have joined the Groups at the special Student member rate of $1.00 annually. Under the newly instituted Affiliate Plan, 111 scientists and medical doctors, whose major interests lie in fields other than electronics, have affiliated with several of the Professional Groups.

All of the Groups except G-28 have levied publications fees and their members are receiving the pertinent Group TRANSACTIONS regularly. In addition, a large number of company, university and public libraries have subscribed to the TRANSACTIONS of all the Groups. There is also a demand for individual Group subscriptions and individual copies of the TRANSACTIONS from outside sources.

Financial and editorial assistance were among the many services rendered by Headquarters to the Groups during 1957. The Office of the Technical Secretary provided administrative services for Group operations, the planning of meetings, advance publicity and the recording and mailing for all activities, including 840 mailings to Group members.

Symposia. The procurement of papers and actual management of national symposia are entirely in the hands of the Professional Groups. Each of the Groups with the exception of the newly organized ones sponsored one or more technical meetings during this year, in addition to Technical Sessions at the IRE National Convention, the WESCON, the National Electronics Conference and other joint meetings, for a total of 43 meetings of national import in 1957.

Publications. During the year, 24 Groups published 75 TRANSACTIONS containing 5,372 pages. Since publication began in 1951, 311 issues (20,704 pages) have appeared.

Professional Group Chapters. 212 Professional Group Chapters have been organized by Group members in 48 IRE Sections. Chapter growth is continuing at a healthy rate. The Chapters are meeting regularly and sponsoring meetings in the fields of interest of their associated Groups in the various Sections.

Section Activities

We were glad to welcome five new Sections into the IRE during the past year. They are as follows: Central Pennsylvania (formerly Centre County Subsection), Shreveport (former Subsection), North

Carolina (formerly Piedmont Subsection), Virginia, and Western Massachusetts (formerly Berkshire County Subsection).

The total number of Sections is now 94.

The Berkshire County, Centre County, Piedmont and Shreveport Subsections became full Sections in the year 1957. The North Carolina-Virginia Section and the Hampton Roads Subsection were dissolved, following the formation of the new North Carolina and Virginia Sections.

The Subsections of Sections now total 28, the following being formed in 1957: Lehigh Valley (Philadelphia), Kitchener-Waterloo (Hamilton), Nashville (Huntsville), Northern Vermont (Boston), Santa Barbara (Los Angeles), Eastern North Carolina (North Carolina), and Burlington (Cedar Rapids).

A growing major activity of many Sections and the larger Subsections in recent years is the publication of a local monthly Bulletin to fulfill the need for announcing to the Section member the increasing activities of the Section, including (1) Section meetings (2) Professional Group Chapter meetings and (3) information on the local and national level of interest to the Section member.

Forty-five of the Sections and Subsections now issue monthly publications.

Student Branches

The number of Student Branches formed during 1957 was 16. The total number of Student Branches is now 151, 113 of which operate as joint IRE-AIEE Branches, and 3 as Student Associate Branches.

Following is a list of the Student Branches formed during the year: Arizona State College, Bradley University, Central Technical Institute, Fairleigh Dickinson University, Hofstra College, Howard University, Lamar State College, Louisiana Polytechnic Institute, Ohio Northern University, Ryerson Institute of Technology, University of Saskatchewan, Southwestern Louisiana Institute, Union College (Schenectady), Valparaiso University, Wentworth Institute, and University of Wichita.

Books

Engineering College Research Review—1957, ed. by Renato Contini

Published (1957) by Engineering College Research Council, New York University, University Heights, New York 53, N. Y. 355 pages+52 index pages+xxx pages. 9×6. $2.00.

This publication appears biennially as a review of the research, sponsored and otherwise, being carried on by the engineering colleges in this country. According to the introduction, it "provides a unique source of information to all those who are in engineering college research so that they can know of the research conducted elsewhere which is similar or related to their own interests. It provides a guide to sponsors of research, whether they be government agencies, industrial organizations, foundations or individuals, by identifying the institutions which are conducting research in areas of interest to the sponsor. Students can be assisted in their choice of institution and even in their choice of field of specialization and thesis topics by a study of these publications. The engineering colleges themselves can benefit from the knowledge here displayed of the policies applicable to research and of the research programs at other institutions."

In order to carry out these worthy purposes, the *Review* provides the following information about 141 administrative units associated with 112 different educational institutions: the names of the research officers, a short statement of the research policies governing the actions of the institutions, information concerning the number of research personnel engaged on either a part-time or full-time basis, a breakdown of the sources of the institutions' research income according to whether it comes from government-military, government-nonmilitary, industry, or other sources, a listing of the active research projects which in this 1957 *Research Review* includes projects for the period of the fiscal year 1955–1956, an estimate of the number of equivalent full-time persons employed on the projects, and a listing of the short courses and conferences held at the respective institutions during the 1955–56–57 period.

As is always the case in such compilation, there is a certain lack of uniformity of reporting and some of the material differs in meaning from institution to institution. There is variation in coverage of disciplinary fields, but electrical engineering and electronics projects are always reported. Not all the institutions report completely on all the information requested. However, such a volume cannot help but be useful in connection with the purposes set forth in the quotation at the beginning of this review. It is recommended for perusal by all those thus concerned.

H. H. Goode
Univ. of Michigan
Ann Arbor, Mich.

Transistor Electronics by David DeWitt and A. L. Rossoff

Published (1957) by McGraw-Hill Book Co., 330 W. 42 St., N. Y. 36, N. Y. 352 pages+17 pages of appendix+4 pages of bibliography+6 index pages +xii pages. 188 figures. 6×9. $8.00.

The authors state in the preface that their "principal aim is to prepare the engineer to use the transistor with a confidence based on a quantitative understanding of its operating mechanisms" while maintaining a level of treatment "appropriate to electrical engineering senior or to the graduate engineer." Accordingly, the book covers the transistor field, including the quantitative treatment of all the important physical processes governing the behavior of the transistor, as well as the main considerations and design principles related to its application in circuits, without assuming that the reader is equipped with specialized knowledge in semiconductor physics.

Chapter 1 is a succinct discussion of fundamental processes in semiconductors. Starting with the Bohr atom, the reader's attention is guided to the wave equation, energy-band theory of crystals, statistics of semiconductors, diffusion, drift, Boltzmann and Einstein relationships, and Hall-effect. In deriving important quantitative relationships pertaining to such complex subjects, the authors make excellent use of intuitive and descriptive approaches and analogies.

Chapter 2 deals with the *pn* junction. Expressions are derived for the saturation current, barrier width and capacitance for both step- and graded-junctions, junction breakdown and temperature dependence of diode characteristics.

Chapters 3 and 4 deal with the junction-transistor. Following a qualitative description of the principal device mechanisms, a low-frequency small-signal analysis based on the diffusion equation is given. Transport factor, emitter and collector efficiencies, high-level injection effects, emitter and collector resistances, extrinsic base resistance, high frequency limitations are discussed.

Chapters 5 to 8 are devoted to low-frequency properties and applications. The circuit properties of the three transistor configurations, the nature, measurement and use of matrix parameters are discussed. Operating point stabilization and thermal behavior are treated in fair detail. A section deals with operation in the collector multiplication region. Various classes of audio power amplification, multistage RC- and transformer-coupled amplifiers and feedback amplifiers are treated in more or less detail. An extremely short discussion of feedback oscillators is also given.

Chapters 9 and 10 deal with small-signal transients and high-frequency amplification. Based on the diffusion equation, transistor driving-point and transfer parameters are derived and various equivalent circuits, using transmission lines and also simplified types with lumped elements are given. Transistor operating and transient properties are discussed and, very briefly, broad- and narrow-band amplifiers.

Chapters 11 to 13 deal with large-signal transient, high-frequency and switching properties and applications, starting with transistor mechanisms (such as minority carrier distributions for various states, high-

level injection, etc.). Mixer and demodulator properties are analyzed. The treatment of transistor switching circuits (bi-stable, monostable and astable circuits, logic circuits, power converters) is qualitative and very short.

Chapters 14–16 deal with radio receivers, special transistor types (including a quantitative analysis of the drift transistor) and with transistor noise respectively.

In a field as new and active as that of transistor electronics, a reviewer can always find fault with one or the other portion of a new book. This is particularly easy by emphasizing minor errors, by selecting a controversial area, by attacking the method of presentation rather than its content, or by declaring parts of the book obsolete due to recent developments. This book, like any other, could be subjected to biting criticism but this would be an unfruitful task.

The principal merit of the book and what distinguishes it from others dealing with transistor electronics is that it discusses device behavior and circuit performance consistently in terms of physical mechanisms and properties. In order to achieve this the authors had to give an unusually complete discussion and characterization of the pertinent physical processes and, due to the complexity of the field, had to use simplification and approximation abundantly. The authors deserve great credit for doing this without becoming scientifically crude or degrading the subject and by maintaining a constant, high scholarly level of treatment. The book reflects a thorough understanding of transistor behavior and is written in a mature and cultured manner reminding one of the best, today almost extinct, tradition of classical technical writing.

The device-oriented portions of the book are considerably superior in quality and completeness to the circuit-oriented portions. (The discussion of oscillators, high-frequency amplifiers, switching circuits, etc., is too brief to serve as a basis of circuit design.) Nevertheless, the book will be appreciated by the circuit designer who truly desires to understand what he is doing and why. The engineer engaged in device design and evaluation will welcome this volume with enthusiasm.

A. P. Stern
General Electric Co.
Syracuse, N. Y.

Radio Aids to Air Navigation by J. H. H. Grover

Published (1957) by Philosophical Library, Inc., 15 E. 40 St., N. Y. 16, N. Y. 131 pages+3 appendix pages+2 index pages+x pages. 63 figures. 8¾×5¾. $6.00.

This book is a sort of encyclopedia of systems of radio navigation presently available or soon to be made available, presented from the user's standpoint in an elementary manner as far as technical matters are concerned. An user having a general knowledge of navigation and radio will understand clearly what results can be obgained in the use of these systems, and will be aware of their qualities and limitations.

As none of the early aids developed have been retired completely from service, the book covers practically every radio aid which has been in existence. It gives more treatment to those which found their origin in Great Britain, probably because of a better access to and availability of the material presented. The U. S. developed systems, particularly those developed in the last five to ten years, are only briefly covered. The author, however, tries to be impartial in the treatment of this subject and reports on the availability as well as on the performance of the various systems without getting into controversial comparisons.

After a brief introduction covering general principles, we find the medium-frequency systems, the DF loop, the automatic DF, then the medium frequency radio ranges, including Consol. In very-high-frequency systems, it covers direction finders, VOR and marker beacons, hyperbolic aids, Loran, Decca, Gee. Under pulse systems, the author groups the various radars for airborne use, covers the distance measuring equipment, and in Chapter 8 goes on to aids to air traffic control with the use of ground based radars, secondary radars, storm warning radar, airfield surface movement indicators. The aids to approach and landing are treated separately and include the ILS (but omitting the latest version of it involving many improvements and designated as the MRN-7 & 8 of the U. S. Air Force), the ground control approach radar system. Chapter 10 mentions as trends in future aids, several systems including TACAN. TACAN is given a short treatment, not extensive enough in our opinion in view of its wide use by the U. S. Military Services and NATO and of the fact that VORTAC, which is TACAN plus VOR, became the standard en route navigation system of the U. S. for civil aviation.

The book finishes with some information on charts, documents and regulations. It contains 135 pages of useful information to the radio engineer not specializing in radio navigation and to the users for whom it had been written.

Henri Busignies
Federal Telecommunication Labs.
Nutley, N. J.

Elements of Magnetic Tape Recording by N. M. Haynes

Published (1957) by Prentice-Hall, Inc., 70 Fifth Ave., N. Y. 11, N. Y. 380 pages+9 index pages+xii pages. Illus. 9¼×6¼. $7.95

This book provides a general non-mathematical treatment of all phases of magnetic recording, starting with a brief introduction to the theory of magnetism and ending with a summary of the fundamental electronic circuits used in magnetic recording equipment

The book is divided into four parts. Part I, the introduction, gives a very brief but enlightening summary of the theory of magnetism introducing the domain theory with special emphasis on the crystal lattice structure and orientation of the particles comprising a magnetic recording tape. From there, magnetic circuit fundamentals are presented and major and minor hyteresis loops are described with emphasis placed on biasing and demagnetization principles. Part I also includes a very brief and elementary introduction to electroacoustic fundamentals, and concludes with a description of the physical constituents, properties and manufacture of recording tape.

Part II provides the principles of magnetic recording, giving a brief description of the recording and playback processes. It includes descriptions of current magnetic recording heads, multiple track recording and tape editing techniques.

In Part III the author presents the basic elements of a tape driving mechanism pointing out the advantages of various basic designs and including descriptions of several manufactured units including special purpose equipment. The author concludes this part with a chapter describing procedures for basic maintenance and repairs.

Part IV contains a resume of typical circuits used in magnetic recording equipment.

The book appears to be directed principally toward the requirements of the recording technician or home recording enthusiast. However, since the book contains a vast amount of information which appears nowhere else under one cover, the book is recommended for reading to anyone associated with magnetic recording. The author's style provides easy reading and maintains the reader's interest. Unfortunately, the author has oversimplified much of the material and in places the simplified presentation of the theory tends to cause the reader to make erroneous assumptions.

F. A. Comerci
New York Naval Shipyard Material Lab.
Brooklyn, N. Y

Logical Design of Digital Computers by Montgomery Phister, Jr.

Published (1958) by John Wiley & Sons, Inc., 440 Fourth Ave., N. Y. 16, N. Y. 392 pages+8 appendix pages+6 index pages+xvi pages. Illus. 9½×6½. $10.50

The stated intent of the author is to provide a basic text for a course in the technique of digital computer logical design. He considers his audience to be uninitiated (to digital computers), but interested engineers, mathematicians, and physicists.

The book begins with a brief introduction to digital computers and then establishes realistic guides which define the scope of the logical design problem. In introducing a subject it is usually difficult to decide how much knowledge to assume of one's intended readers. About the only question that could logically be brought up in this case is whether a course in logic design should be a first brush with digital computers. The brevity and clarity, however, will appear refreshing even to those long associated with the subject.

Boolean algebra, and the manipulation of boolean functions is next considered. The formal introduction of boolean algebra with digressions into short-cut methods was appreciated by the reviewer and will probably satisfy most students' requirements. This knowledge is then applied to flip-flops which introduces the time variable and application equations. Asynchronous memory units are not considered nor are dynamic pulse memories (contrasted to d.c. level memories typified by flip-flops). The compounding of units is then developed and the synthesis problem tackled, the Huffman, Mealy, Moore techniques being used throughout. Memory and input-output equipment are next discussed. This will most likely provide a welcome respite for those readers who find boolean expressions growing tedious. This is only

temporary, however, for the arithmetic unit is next up for consideration. The four common arithmetic operations are covered as well as the comparison operation. A short chapter on the error problem follows. Finally, the overall logical design problem is tackled wherein a general purpose and a special purpose computer are designed. The last chapter provides a three-page description of the logical designer's responsibility after the design is complete. Appendices on the solution of simultaneous boolean equations and the relationship between binary and reflected codes are included. At the end of most of the chapters suitable problems are provided and a well selected bibliography is given.

The author has definitely realized a text that should enhance the learning of techniques which will enable the reader to handle logical design problems of a limited scope. He has mentioned other aspects of logical design such as asynchronous systems, but has not provided sufficient directions to allow the student to easily locate this knowledge.

The book does fill a very definite need in the computer-education field and should be considered by instructors for inclusion in digital computer course work. Due to the many detailed examples in the text and the extreme clarity of presentation, apt students will be able to progress satisfactorily without an instructor. Thus, this book will be of significance to most persons working in the digital computer field who have never gone through a complete or formal digital computer logical design problem.

R. A. TRACY
Burroughs Corp. Res. Center
Paoli, Pa.

L'Automatique des Informations by F. H. Raymond

Published (1957) by Masson et Cie, 120 Blvd. Ste-Germain, Paris 6, France. 185 pages. 51 figures. 9×5½.

The title and subtitle of this 188-page book can be literally translated: *The Automation of Information. Principles underlying the machines operating on information (calculating machines in particular).*

In the reviewer's opinion, a more adequate title would be: *Exercises in the Mathematical Expression of Some Well-Known Conditions Encountered in the Study of Data Processing Devices.*

The author, F. H. Raymond, is introduced by M. Chalvet, Chief Engineer of the French Naval Artillery. Mr. Chalvet, probably a good mathematician and cannoneer himself, candidly admits that he is somewhat flabbergasted at the mathematical fireworks Mr. Raymond deems necessary to shoot in the air in order to explain what a computer really is.

From the introduction we learn that F. H. Raymond is one of the foremost electronicians in France. He is professor at the Conservatoire des Arts et Metiers and at the Faculte des Sciences de Toulouse, former president of the Association des Ingenieurs Electronciens, and he has been very active in promoting and organizing international conventions such as: "Journees Internationales de Calcul Analogique" at Brussels, Belgium, 1955; and "Congres International de L'Automatique" at Paris, France, 1956. An undefatigable writer, Mr. Raymond is also manager of a corporation manufacturing electronic devices, especially for the military.

Early in his book, Mr. Raymond states that, in accordance with R. P. Russo, he believes that "the present evolution of data processing machines proceeds more from a deeper understanding of ideas already utilized (but not expressed with luminous clarity, this implying the use of some abstract language), than from invention."

The above remark and its context makes it clear that the book is written for the pure theoreticians attracted by the opportunity for more mathematical developments offered by the computer problem rather than for the engineers who wish primarily to develop computers as such.

The book is divided into ten parts: Introduction, 3 pages; A Panoramic Look at Automation, 28 pages; Numerical Representations of Information and Coding, 25 pages; Fine Analysis of a Digital Machine Structure, 25 pages; What Is an Automatic Calculator, 16 pages; Basic Digital Structure. From the Concrete to the Abstract, 37 pages; Programming, 17 pages; Conclusion, 3 pages.

The impression given by this table of contents is probably not altogether wrong. The book seems to have been assembled hastily from accumulated notes, of which many are indeed well worked out, but the typographical errors and discrepancies between figures and text are too many for pleasurable reading.

Between pages 70 and 85 is a quite comprehensive discussion of the various error detecting and error correcting codes including the representation of an n-digit word in an n-dimension space. References are made to the works of M. Gloess, H. Stibitz, Sainte-Lague, R. W. Hamming, R. N. Gilbert, W. H. Klautz, L. Brillouin, etc.

On page 111, following considerations on the nature of Boolean operations, is an interesting remark: "All operations that can be performed by a digital machine are equivalent to operations performed in a groupoid, on natural numbers." Unfortunately, the author informs us that he has not the time, at present, to develop the useful consequences that one would expect from his statement.

There are many other catching things to be found in Mr. Raymond's book. Although its sparks do not seem to shed much light on the computer problem, it might be helpful reading to create the required mood at such times when the ominous deadline compels the reluctant mind to abandon the clarity of the general concept and to break down into a maze of precise but not very lucid intricacy.

J. P. JEANNENEY
IBM Corp.
Poughkeepsie, N. Y.

The Management Approach to Electronic Digital Computers by J. S. Smith

Published (1957) by Essential Books Inc., 16-00 Pollitt Dr., Fairlawn, N. J. 205 pages+16 appendix pages+4 index pages+2 bibliography pages+xi pages. 31 figures. 8½×5½. $6.30.

Automation and Management by J. R. Bright

Published (1958) by Harvard Business School, Div. of Research, Boston, Mass. 280 pp.+13 appendices. 16 figures. $10.00.

These two books, the one having its source in England and the other in America, complement each other in this reviewer s estimation to an extent that amply warrants their being teamed. Mr. Smith, chartered accountant and management consultant, Founder Member and member of Council and the British Institute of Management, approaches the application of electronic data processing systems from a background based on long tradition in the field of business management that takes electronic data processing in stride with a refreshing and illuminating view of essential aspects uncluttered by entrancement with the technology or the business management shop-talk too frequently encountered in our current literature on data processing and automation.

Mr. Bright, assistant professor of business administration at the Harvard Business School and well known as an editor, writer and consultant in the American industrial field of automation of manufacturing, treats his subject with equal facility, candor and competence based on many years' experience as an industrial engineer as well as from more than three years of recent research under the auspices of the Harvard Business School, aimed discerningly at analysis of the management implications of recent technological developments in both the data processing and the production areas.

The first ten chapters of Mr. Smith's book constitute an excellent introduction for the non-technical as well as the technical reader to the field of electronic computers. The following chapters, eleven through seventeen, then give a better appreciation of the actual computing equipment in business organizations of various sizes and kinds, albeit from the British point of view. It is regrettable that Chapter Fourteen on available business computers fails to include several prominent American types, although the author indicates knowledge of such Americana by mention elsewhere in his book. A Puckish attitude regarding a few more or less sacred cows of technology is welcome. His comment that, "The argument for not using magnetic tape for inputting original data is far from negligible," is definitely arguable but constructively provocative; and his reference to "the new miracle factor HP $2x$" ("Hocus Pocus Twice Multiplied") of Mr. James ("Scotty") Reston (*not* Weston) of the *New York Times* in the introduction to Chapter Fourteen regarding commercial computer specifications, suggests no malice intended by the above-noted omission.

In his chapters on the impact of integrated data processing on organizational structures and on the size of business which can use a computer, Mr. Smith makes a strong case for the system concept and for adequate methods and operations study; a significant quotation on the latter: "In my experience any business which has not got an efficient organisation and methods department constantly reviewing its work can at any stage make a review of its organisation and reduce its clerical staff by at least ten per cent." The final chapter, "The Social Significance of the Computer," gives food

for thought in economics as well as it indicates sound social principles.

In *Automation and Management* Mr. Bright similarly gives force to the system concept and graphically illustrates the importance of thorough analysis and advance study in planning for automation, illuminating the principles by case study data on thirteen automation programs in such diverse manufacturing activities as automobile manufacturing plants, a feed mill, a fertilizer plant, a coal mine, and in electrical and electronic equipment production. These case studies are treated under three headings, "The Nature of Automatic Manufacturing," "Experiences With Automation" and "Critical Areas of Automation," the last being especially important with regard to the management, personnel and sales aspects—with implications that are in both the economic and social areas. While Mr. Bright modestly disclaims to provide an instruction manual on automation to suit any and all cases, he definitely provides a pattern and sound way of thinking that should serve as a management guide toward the proper and profitable objective.

J. J. LAMB
Consulting Engineer
South Norwalk, Conn.

The Measurement of Colour, 2nd ed. by W. D. Wright

Published (1958) by The Macmillan Co., 60 Fifth Ave., N. Y. 11, N. Y. 242 pages+12 appendix pages +6 index pages+ix pages. 83 figures. 8¼×5¼. $10.75.

Dr. W. D. Wright of the Technical College of Science and Technology is an authority in the field of color (or colour as he would have it). Merely to list his publications would require more space than is taken by this review. An important one of these publications was his *The Measurement of Colour* which appeared in 1944, and which was added immediately to the libraries of practicing colorimetrists.

The book reviewed here is a second edition of the 1944 work. Except for the title, the author, and the evident authority of the text, there is little in common between the two editions. As a result of some rearrangement of the order in which the material is presented and of complete rewriting of about two-thirds of the text, Dr. Wright has produced an easily readable and up-to-date treatise on color measurement.

In his first two chapters, Dr. Wright discusses the physical problems of the production, reflection, and absorption of radiant energy in the visible spectrum, and the properties of the eye in so far as they relate to the perception of color. With this background established, he treats the principles of photometry and of colorimetry as an introduction to the 1931 CIE system of colorimetry.

The remainder of Dr. Wright's text is concerned with the practical problems of color measurement and specification. Following a description of spectrophotometers and colorimeters, he discusses physical color standards and their orderly arrangement in color atlases.

The final chapters of *The Measurement of Colour* are devoted to the colorimetry of color reproduction, and to the applications of colorimetry in sundry scientific and commercial problems. The television engineer will find the discussion of color television and the material on the assessment of color pictures of particular interest. Dr. Wright has succeeded in condensing his treatment of these two problems into a few pages of highly enlightening text.

The usual tables for colorimetric calculations are included in the four appendices to the book.

Dr. Wright has done a real service to the colorimetrist by revising and bringing up to date this book. Quite apparently, he has read most of the post-war work in this field and distilled the important elements from numerous publications. In consequence, the student who wishes to know the current thinking in the colorimetric field can do no better than to turn to this book.

This book shows every evidence of careful proofreading. One unfortunate lapse is to be found in Plate 4, where it is obvious that the color patches have been transposed so that the most highly saturated color is associated with the spectral reflectance characteristic corresponding to the least saturated color.

W. T. WINTRINGHAM
Bell Telephone Labs.
Murray Hill, N. J

Abstracts of IRE Transactions

Aeronautical & Navigational Electronics

VOL. ANE-5, NO. 1, MARCH, 1958

General Aspects of Cooling Airborne Electronic Equipment—J. Kaye and H. Y. Choi (p. 4)

A brief introduction and review is presented of the problems of protecting and cooling airborne electronic equipment in present and future high-speed devices. The thermal characteristics associated with high-speed flight of manned and unmanned vehicles are related to the thermal problems of irreversible electrical and magnetic components which possess upper bounds of temperature for reliable operation. Various types of cooling devices, fluid flow processes, and techniques are presented and discussed, and the trends of future development are reviewed briefly.

Problems Arising in High-Speed Aircraft due to Cooling Requirements of Electronic Equipment—N. A. Carhart (p. 10)

High-speed aircraft require the successful operation of large quantities of electronic equipment. The problem of providing a suitable environment for such equipment is of interest to both airplane and electronic designers. A proposal is offered by which cooling provisions may be standardized for airplanes of widely varying performance. Basic design parameters for the development of such a system are outlined, as are the economic factors involved.

Temperature Limits, Ratings, and Natural Cooling Procedures for Avionic Equipment and Parts—J. P. Welsh (p. 15)

The results of adequate cooling of electronic parts are gains in part life and reliability. An engineering compromise between ideal electronic part temperature and the thermal point of diminishing return must be evaluated not only with respect to desired life, but also in terms of the electronic circuit and cooling efficiencies. This paper outlines the flow of heat within, through, and from heat producing electronic parts in terms of internal thermal limitations, part surface and environmental ratings, and cooling indices. Natural heat flow design data pertinent to conduction cooling of heat sources, tube shields, the placement and mounting of parts, and "sink connectors" are presented.

Forced-Air Direct-Contact Cooling of Airborne Electronic Equipment—T. Jordan (p. 25)

A brief explanation of the prlnciples of forced-air direct-contact cooling is given, and it is shown that the heat density of the electronic part being cooled determines essentially the range of applicability of this method. The discussion then progresses to cooling problems arising when this concept is applied to airborne electronic equipment. Several techniques are presented for obtaining high heat transfer coefficients and good air distribution with small pressure drops. The use of laminar flow in all of these techniques is shown to be effective, and examples are given of practical applications, such as the use of tube shields and special solenoid designs. Applications of thermal insulation to forced-air direct-contact cooling in aircraft are also treated, and the shortcomings, as well as the advantages of forced-air direct-contact cooling, are pointed out.

Cold Plate Design for Airborne Electronic Equipment—M. Mark (p. 30)

A solution to the problem of cooling certain types of high heat-dissipating airborne electronic equipment is the modification of the equipment chassis to incorporate a simple plate-fin heat exchanger, sometimes called a "cold plate." In this paper equations for cold plates are developed and the effect of variations in fin configuration on performance are discussed. Theoretical and experimental results based on tubes mounted on a cold plate are found to compare very favorably. The application of the cold-plate technique is illustrated for power transistors. This method of cooling is shown to be useful and efficient.

Why and How Should High-Speed Aircraft Electronics Be Liquid Cooled?—Walter Robinson (p. 36)

Based on weight penalty comparisons of ultimate heat sinks for electronic equipment cooling, the use of expendable evaporants and fuel is indicated for high supersonic flight. Centralized ram air should be the alternate coolant during subsonic flight. System integration is best accomplished with a recirculating liquid transfer system, which is relatively easy to control and which is characterized by small pumping power, line size, and heat gain from high temperature environments. Because of these features close temperature control of dispersed components and cooling of remote high power units is best achieved by liquid coupling, regardless of the type of ultimate heat sink.

Part temperature rises in high voltage equipment can be minimized by use of dielectric liquids. Although this permits some reduction in ultimate coolant weight penalty, the reduction is usually not great enough to offset the equipment weight increases that are due to liquid filling. Vapor-filled or air-filled units with minimal liquid contents and liquid transport to part surfaces by capillary action or mechanical means are superior.

Electronic assemblies that are to be series cooled in sealed liquid transfer systems should be designed for conduction, forced air convection, or radiation heat transfer from the parts, and high power units should have integral liquid cooled heat exchangers, placed in separate transfer system branches. Internal heat transfer in such units may be attained by conduction through flexible metal or rubber jackets and electrical insulators, by air convection in standard modules, and by liquid film cooling. In this paper, design and performance details of liquid coupling and dielectric liquid cooling are discussed and are illustrated with examples of actual electronic packages and cooling system components.

An Evaporative-Gravity Technique for Airborne Equipment Cooling—M. Mark, M. Stephenson, and C. Goltsos (p. 47)

In airborne electronic packages, for either thermal or electrical reasons cooling air often is not ducted directly over the components but is passed through a heat exchanger. Consequently, the thermal path between the heat exchanger and the components must be of low impedance to result in efficient heat transfer. The high heat transfer coefficients obtainable as a liquid boils and condenses permit an effective reduction of the temperature drop between the electronic components and the heat exchanger. In this paper the development and design of an evaporative system utilizing gravity return flow is discussed, and the test results of such a system are compared with those obtained utilizing a conventional metallic conductive paths technique. Where heat dissipation or cooling air inlet temperature is high, the evaporative-gravity (ev-grav) system is shown to be the most effective.

Temperature Measurement—W. M. Rohsenow and R. J. Nickerson (p. 52)

The quality of heat transfer data depends, to a large extent, on the accuracy of temperature measurements. Errors in temperature measurement may be due to calibration errors and/or application errors. This paper describes some common methods of temperature measurement and estimates the precision commonly attainable with these devices. The calibration of thermometers is also discussed. The problems of surface temperature measurement, fluid temperature measurement, and bulk temperature measurement are discussed in some detail and some specific examples are included.

Radio Rack Cooling in Present Commercial Aircraft—T. A. Ellison (p. 58)

Limited heat dissipation from the necessarily compact electronic equipment installations of airline aircraft has been a problem to commercial air carriers for many years. To deal with this problem, United Air Lines, in August, 1955, began equipping fleet aircraft with a forced air electronic equipment cooling system, with the following objectives: 1) improvement of electronic equipment reliability, 2) reduction of crew discomfort caused by electronic-equipment generated heat in the flight deck area, and 3) evacuation of any smoke from the flight deck area which might arise from electronic equipment failure.

Observation of electronic equipment maintenance records during the subsequent transition period from convection to forced air equipment cooling shows characteristic failure rate trends which indicate the dependence of equipment reliability on operating temperature. These patterns occur independently in fleet records for the Convair 340, and DC-6 and DC-6B fleets, in time phase with the progress of conversion to forced air cooling in the fleet.

These records, in conjunction with other comparative testing of forced cooled and convective cooled equipment installations indicate that forced air cooling provides an effective answer to airline equipment heat problems.

Audio

Vol. AU-5, No. 6, November–December, 1957

A Magnetic Recording Pickup Head with Crossed Cores for Hum Balance—Marvin Camras (p. 153)

This paper describes a simple means of hum balance for a magnetic recording pickup head. Hum level improvements of approximately 20 db are obtained in both single and double coil pickup heads by the method described in this paper.

Application of Acoustical Engineering Principles to Home Music Rooms—W. B. Snow (p. 154)

This paper describes the important factors to be considered in the construction of a *home music room*. The effects of reverberation, room resonances, sound diffusion, and noise are considered. Special consideration is given to the control of acoustical conditions, and a room in which many of these factors have been taken into account, is described.

On Making Accurate Measurements with a Harmonic Distortion Meter—J. R. MacDonald (p. 161)

It is pointed out that an avoidable error occurs in making total harmonic distortion measurements by the usually recommended method with an instrument incorporating an average-responding output meter. The error arises from the fact that the average value of a full-wave rectified distortion signal can be reduced by adding a small phase-shifted fundamental component to the distortion. The error is always in the direction to make equipment seem better than it actually is and can amount to ten per cent or more. It can be readily avoided by changing the instrument operating procedure.

Circuit Theory

Vol. CT-5, No. 1, March, 1958

Our Spotlight Falls on Network Topology—W. R. Bennett (p. 3)

On the Solution of the Equations Obtained from the Investigation of the Linear Distribution of Galvanic Currents (Translated by J. B. O'Toole)—G. Kirchhoff (p. 4)

An English translation is given of Kirchhoff's classic paper: "Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung Galvanischer Ströme geführt wird," *Annalen der Physik und Chemie*, vol. 72, pp. 497–508, 1847.

Kirchhoff's "Third and Fourth Laws"—Louis Weinberg (p. 8)

The epoch-marking publication of Kirchhoff in 1847 is considered in this paper. The rule

given by Kirchhoff deserves to be understood by almost all engineers, for it shows how to determine in a simple and straightforward way the system functions of any passive network that contains no transformers. Preliminary to the consideration of the rule and its dual, the basic concepts of combinatorial topology are summarized in the present paper. The rules are then stated in the detail necessary to show that their application can be carried out in a routine and mechanical manner. Application of the rules to problems in analysis and synthesis is illustrated in a number of examples. Included among these examples are the derivation of the necessary and sufficient conditions on RLC networks that have been previously given by Fialkow and Gerst, the analysis of an infinite graph, and the determination of the elements of the chain matrix (or the open-circuit impedance or short-circuit admittance matrix) by essentially a single calculation. Finally, the Feussner method for simplifying the use of the rules is discussed.

Some Topological Considerations in Network Theory—F. Reza (p. 30)

Elements of the theory of linear graphs which are readily applicable to communication theories are reviewed. These concepts are used to examine the geometrical significance of Kirchhoff's laws. Topological rules are discussed and illustrated which aid not only in formulating solutions for the resistive networks considered by Kirchhoff but also apply to RLC networks, switching networks, and probabilistically defined networks. References for further study are given.

General Topological Formulas for Linear Network Functions—C. L. Coates (p. 42)

This paper presents a development of topological formulas for the vertex admittance functions of a network which includes components that are mutually coupled. The results are proper generalizations of those which have been presented previously for networks without mutually coupled components.

With each network, which excludes generators but which includes mutually coupled coils, linear vacuum tubes, and transistors, is associated a linear graph G. Each edge (element) of G is either a single edge or a pair edge which belongs to an edge pair of G. For each graph G there is a companion graph $\underline{G}$ such that G and $\underline{G}$ constitute a graph pair.

The vertex driving point and transfer admittance functions of G are topologically related to G by complete tree sets and complete two-tree sets. A complete tree (two-tree) set is a set of edges of G the corresponding subgraphs of which are trees (two-trees) of both G and $\underline{G}$. Corresponding to each such set is a weight and an admittance. The admittance is the product of the admittance weights of the edges which belong to the set. The weight determines the sign of the admittance and depends upon the pair edge members of the set, their orientations and their topological arrangement in the corresponding subgraphs of G and $\underline{G}$.

The vertex driving point admittance function associated with the vertex pair p_i, p_σ of G is

$$y_{i,\sigma} = \frac{V(Y)}{W_{1,V-1,V}}.$$

$V(Y)$ denotes the sum, over all possible complete tree sets of G, of the product of the complete tree admittance and the associated weight. $W_{1,V-1,V}$ denotes a corresponding sum of products for the complete two-tree admittance and associated weight. Similar expressions are given for the vertex transfer admittance functions of G.

Synthesis of RC Grounded Two-Ports—E. S. Kuh (p. 55)

This paper is concerned with the problem of synthesizing a minimum phase RC transfer function into a grounded 2-port. The network is to be inserted between two prescribed resistive terminations and the gain obtained is maximized. The method is simpler and more practical than the previous available techniques for synthesizing an unterminated, or terminated at one end, RC network.

On the Compactness of an RC Quadripole—Paul Slepian (p. 61)

Consider an RC quadripole with open-circuit impedances $z_{11}(s)$, $z_{12}(s)$, and $z_{22}(s)$, and let s_v be a pole of any of these functions. If $k_{11}^{(v)}$, $k_{12}^{(v)}$, and $k_{22}^{(v)}$ are the respective residues of these three functions at s_v, then it is well-known that $k_{11}^{(v)}k_{22}^{(v)} - k_{12}^{(v)} \geq = 0$. If the inequality is an equality, then s_v is called a compact pole; if every such pole is compact, the network is called compact. In this paper two new properties of compactness are exhibited and discussed.

It is shown that if the network is grounded, noncompactness implies certain degeneracies in the determinant of the nodal admittance matrix and its cofactors.

If the network is terminated with a resistance R, it is shown that for all but a finite number of values of R, the overall terminated network is compact. Thus, a noncompact resistance terminated RC quadripole can be approximated to any degree of accuracy by a compact network of this type, which implies that noncompactness is not detectable by terminal measurements of the open-circuit impedances.

Some Relations Between Frequency and Time-Domain Errors in Network Synthesis Problems—Igor Gumowski (p. 66)

Bounds are derived for frequency and time-domain errors of a transient synthesis problem. The maximum error produced by a general input is related to the error of the impulse input. By considering distributions, a meaning is proposed for the time-domain error in the case when the error-transfer function is not Fourier-transformable in the ordinary sense. Rectangles and bell-shaped curves are used in approximating the error-transfer function.

A Transistor Univibrator with Stabilized Pulse Duration—D. J. Hamilton (p. 69)

The use of an elementary characterization of the transistor leads to a very simple model of the basic emitter-coupled univibrator, or monostable multivibrator. Based on this model calculations of waveforms exhibit a reasonable degree of accuracy. Furthermore, the simplicity of the model is conducive to a facile understanding of the operation of the circuit and its disadvantages. A method for circumventing the adverse effects of temperature upon the quasi-stable duration is readily deduced, and may be extended to permit an increase in quasi-stable duration without an increase in jitter.

Experimental and calculated waveforms are given and experimental curves are given which indicate the effectiveness of the temperature stabilization.

Synthesis of Sampled-Signal Networks—P. M. Lewis II (p. 74)

The problem of designing sampled-signal networks with a finite number of passive, but not necessarily lumped, circuit elements is discussed, together with realizability conditions and a general synthesis procedure. The resulting networks consist of resistors and open- and short-circuited transmission lines, all of which have the same delay. A method for finding a realizable networks function that approximates any arbitrarily given magnitude characteristic is presented, and an approximate low-pass to sampled-data-network transformation is included.

Reviews of Current Literature—On a New Type of Adjustable Equalizer—J. Oswald (in French). Reviewed by G. L. Ebbe (p. 78)

Theorie des Circuits de Telecommunication—V. Belevitch (in French). Reviewed by H. J. Carlin (p. 80)

Abstracts of Articles on Circuit Theory—On the Direction and Curvature of the Impedance Loci of Real and Idealized Two Poles—H. Wolter (in German) (p. 81)

On the Effect of the Coefficients of the Transfer Function of a Transmission System on the Behavior in the Time Domain—R. Hofmann and W. Walcher (in German) (p. 81)

Tabulation of Linear Circuits—W. Klein (in German) (p. 82)

Approximations of Logarithms by a Power Law; Application to Electrodynamic Problems—W. Rehwald and O. Zinke (in German) (p. 82)

Errors in Measurement of Pulse Frequency Spectra by Means of Stagger Tuned Resonant Circuits—(p. 82)

Theoretical and Experimental Investigation of Distortion in Low Frequency Junction Transistor Four Poles—G. Spescha and M. J. O. Strutt (in German) (p. 82)

A Network Theorem—E. Green (p. 82)

Filter Technique Development in France During the Last Ten Years—J. Colin (in French) (p. 82)

Synthesis of Voltage Transfer Functions—P. M. Lewis II (p. 82)

Transient Correction by Means of All Pass Networks—J. Pinson (p. 83)

On the Use of Delay Lines as Network Elements—L. E. Franks (p. 83)

Correspondence (p. 83)

PGCT News (p. 86)

Engineering Writing and Speech

Vol. EWS-1, No. 1, March, 1958

Information Theory

Vol. IT-4, No. 1, March, 1958

The method of Part I is applied to the problem of finding the characteristic function for the probability distribution of $\int_0^t \Sigma_{jk} \chi_j(\tau) K_{jl}(\tau) \chi_l(\tau)\, d\tau$, where $\chi_j(\tau)$ denotes the jth com-

ponent of a stationary n-dimensional Markoffian Gaussian process. The problem is reduced to the problem of solving $2n$ first-order linear differential equations with initial conditions only. For the case of constant K, the explicit solution is given in terms of the eigenvalues and the first $2n-1$ powers of a constant $2n \times 2n$ matrix. For the case of a symmetric correlation matrix which commutes with K, the problem is reduced to the one-dimensional case treated in Part II. For the case $K_{ij}(t) = \delta_{il}\delta_{jl}e^{-t}$, where the functional represents the output of a receiver consisting of a lumped circuit amplifier, a quadratic detector, and a single-stage amplifier, the solution has been obtained in a form which is more explicit than that provided by the earlier methods.

The Axis-Crossing Intervals of Random Functions—II—J. A. McFadden (p. 14)

This paper considers the intervals between axis crossings of a random function $\xi(t)$. Following a previous paper, continued use is made of the statistical properties of the function $x(t)$ and the output after $\xi(t)$ is infinitely clipped. Under the assumption that a given axis-crossing interval is independent of the sum of the previous $(2m+2)$ intervals, where m takes on *all* values, $m = 0, 1, 2, \cdots$, an integral equation is derived for the probability density $P_0(\tau)$ of axis-crossing intervals. This equation is solved numerically for several examples of Gaussian noise. The results of this calculation compare favorably with experiment when the high-frequency cutoff is not extremely sharp. Under the assumption that the successive axis-crossing intervals form a Markoff chain in the wide sense, infinite integrals are found which yield the variance $\sigma^2(\tau)$ and the correlation coefficient κ between the lengths of two successive axis-crossing intervals. These parameters are obtained numerically for several examples of Gaussian noise. For bandwidths at least as small as the mean frequency, κ is large. For low-pass spectra, κ is small, yet the statistical dependence between successive intervals may be strong even when the correlation κ is nearly zero.

Recursion Formulas for Growing Memory Digital Filters—Marvin Blum (p. 24)

A growing memory digital filter is defined by considering the input (y_{v-u})–output (Z_m) relationship in the form $Z_m = \sum_{u=0}^{m} W_{um} y_{m-u}$, $m = 0, 1, 2, \cdots$ where W_{um} is the weighting sequence of a linear time varying digital filter. Contained herein are a derivation of an optimum growing memory smoothing and prediction filter in the least squares sense for polynomial input functions (of degree $= K$) and a theorem on the class of time invariant sequence W_u, which are solutions of a difference equation of finite order, and an application of the theorem to the synthesis of samples correlated noise by digital processes, using recursion formulas. The recursion formulation represents a practical solution to the generation of a correlated noise sequence on line during simulation studies on digital computers.

The Fluctuation Rate of the Chi Process—R. A. Silverman (p. 30)

The chi process is defined as a natural generalization of the chi distribution of statistical theory. A formula is derived for the expected number of level crossings per second of the chi process. The formula contains as a special case the familiar expression for the fluctuation rate of the envelope of Gaussian noise.

Loss of Signal Detectability in Band-Pass Limiters—R. Manasse, R. Price, and R. M. Lerner (p. 34)

A loss of signal detectability results from ideal symmetric limiting of a very narrow-band signal in narrow-band Gaussian noise. At small input snr this loss can be expressed in terms of the degradation of the effective signal energy-to-noise power per cycle ratio. Proceeding from results derived previously by Davenport and Price, an expression is derived for this degradation in terms of the autocorrelation function of the input noise. The degradation is evaluated for three typical input noise autocorrelation functions and is found to be quite small in all these cases. It is seen that the degradation can be made to vanish by appropriately shaping the spectrum of the input noise. Regulation of the input noise spectrum to obtain conditions of limiter operation assumed in this paper may often prove to be a convenient method for reducing loss of signal detectability in band-pass limiters.

A Table of Bias Levels Useful in Radar Detection Problems—James Pachares (p. 38)

A table of bias levels $\lambda = \lambda(m, p)$ corresponding to specified probabilities of false alarm, 10^{-p}, when integrating over m independent noise pulses using a radar system with a square-law detector receiver, is given for $p = 1(1)12$ and $m = 1(1)150$. The notation $p = 1(1)12$ means that p varies from 1 to 12 in steps of unity.

Weighted PCM—Edward Bedrosian (p. 45)

A modified form of pulse-code modulation, called weighted pcm, is described in which the relative amplitudes of the pulses within the pulse-code groups are adjusted so as to minimize the noise power in the reconstructed signal due to errors in transmission. A performance analysis shows the knee of the output signal-to-noise ratio curve to be moved 1.4 db to the left for a weighted seven-digit pcm system. An information rate study reveals that the maximum improvement which can ever be achieved by any encoding process over a conventional seven-digit pcm system is only 8 db. The importance of selecting a suitable system worth criterion is emphasized by showing that weighting increases the information rate relative to an rms fidelity criterion but decreases it on a pure equivocation basis.

Radar Detection Probability with Logarithmic Detectors—B. A. Green, Jr. (p. 50)

It is shown that use of a logarithmic (rather than square-law) detector in a search radar system inflicts a loss of sensitivity equivalent to a power loss of the order of one db under typical conditions. Curves of probability of detection vs relative range are given for various false alarm probabilities and various numbers of pulses integrated. The power loss (in db) is roughly proportional to the logarithm of the number of pulses integrated.

Correction (p. 53)

Envelopes and Pre-Envelopes of Real Waveforms—J. Dugundji (p. 53)

Rice's formula for the "envelope" of a given signal is very cumbersome; in any case where the signal is not a single sine wave, the analytical use and explicit calculation of the envelope is practically prohibitive. A different formula for the envelope is given herein which is much simpler and easier to handle analytically. We show precisely that if $\hat{u}(t)$ is the Hilbert transform of $u(t)$, then Rice's envelope of $u(t)$ is the absolute value of the complex-valued function $u(t) + i\hat{u}(t)$. The function $u + i\hat{u}$ is called the pre-envelope of u and is shown to be involved implicitly in some other usual engineering practices.

The Hilbert transform $\hat{u}$ is then studied; it is shown that $\hat{u}$ has the same power spectrum as u and is uncorrelated with u at the same time instant. Further, the autocorrelation of the pre-envelope of u is twice the pre-envelope of the autocorrelation of u.

By using the pre-envelope, the envelope of the output of a linear filter is easily calculated, and this is used to compute the first probability density for the envelope of the output of an arbitrary linear filter when the input is an arbitrary signal plus Gaussian noise. An application of pre-envelopes to the frequency modulation of an arbitrary waveform by another arbitrary waveform is also given.

Microwave Theory & Techniques

VOL. MTT-6, NO. 2, APRIL, 1958

A Wide-Band Double-Vane Torque-Operated Wattmeter for 3-CM Microwaves—A. L. Cullen, B. Rogal, and S. Okamura (p. 133)

This paper describes a torque-operated wattmeter for waveguide, capable of measuring power in the range of 10 to 200 watts in the wavelength range 3.05–3.45 cm, with an accuracy of about 2 per cent over most of the wave band.

The instrument is an absolute standard since its calibration depends only on measurements of mass, length, and time. Negligible power is absorbed, and the instrument is insensitive to mismatch.

Traveling-Wave Resonators—L. J. Milosevic and R. Vautey (p. 136)

In the first part of the paper, the principles are given which have led to the conception of the traveling-wave resonator, and the calculations enabling its operation to be understood are presented.

The second part describes the apparatus in detail and examines it, bearing in mind its use as high-power testing equipment.

Reflection Coefficient of *E*-Plane Tapered Waveguides—Katsu Matsumaru (p. 143)

This paper treats the reflection of linearly and sinusoidally tapered waveguides. In the first part, reflection coefficients of linearly tapered waveguides for dominant modes are calculated. Graphs of the vswr of tapers for different impedance ratios are plotted showing that the vswr does not go to unity at multiples of a half wavelength. In the second part, reflection coefficients of sinusoidally tapered waveguides are calculated. Experimental data verify the theory for both kinds of tapers of various lengths at 4 kmc band.

Linear tapers perform almost as well as exponential tapers, and better than shorter hyperbolic tapers. The reflection coefficients of sinusoidal tapers can be about half as small as that of the linear tapers, and these tapers compare favorably with the Dolph-Tchebycheff and the Willis taper of improved design.

Propagation in Ferrite-Filled Microstrip—M. E. Brodwin (p. 150)

The propagation constant of a ferrite-filled microstrip is measured as a function of the longitudinal static magnetic field. The results agree with the analysis by Van Trier of the infinite parallel plane waveguide filled with gyromagnetic media. The analysis is extended to anisotropies greater than 0.5. A simple relationship between propagation constant and anisotropy for the quasi-TEM mode with small spacing ($X \ll \lambda_g$) is noted. Cutoff spacings for higher modes are calculated. An apparatus for the measurement of propagation constant independently of interface reflections is described.

An Improved Method for the Determination of *Q* of Cavity Resonators—Amarjit Singh (p. 155)

The various Q factors and circuit efficiency of a cavity resonator can be evaluated from standing-wave measurements on a transmission line or waveguide coupled to the resonator. In the usual method, measurement errors near the

half-power points have an unduly large influence on the result. This paper describes a method in which this type of error is avoided.

In the new method, vswr and position of minimum at various frequencies are plotted on a Smith chart and a circle is drawn through the points. This circle is suitably rotated around the center of the chart and a value of equivalent susceptance is read off for each frequency. The graph of susceptance vs frequency is a straight line, from whose slope the Q factors are evaluated.

The underlying theory of the above method is discussed and typical experimental results are presented. Charts of parameters required in the calculations are given.

Characteristic Impedances of the Slotted Coaxial Line—Jadwiga Smolarska (p. 161)

The characteristic impedance for the two possible TEM modes is calculated for a slotted coaxial line whose outer walls have a zero thickness. Conformal mapping is used in the calculations. The characteristic impedance for a slotted coaxial line is calculated in an approximate way for outer wall thickness different from zero.

Velocity Modulation of Electromagnetic Waves—F. R. Morgenthaler (p. 167)

This paper deals with electromagnetic wave propagation through dielectric media whose propagation constants vary as a function of time.

If the parameters of the medium cannot respond to changes in the electric and magnetic fields of the propagating wave, the fields within such media will be linear. Maxwell's equations are solved for cases in which the scalar permittivity and permeability vary independently with time. When the impedance is constant, an exact solution is obtained. When the impedance varies, a closed form approximation is found since an exact solution is not always possible. The field energy and electromagnetic momentum are derived for a velocity transient and it is seen that, in general, the energy changes and the momentum remains constant.

The frequency deviation that results when a monochromatic wave is passed through a section of dielectric with nonconstant velocity of propagation is discussed in detail. An approximate solution is obtained for the case in which the electrical length of such a section is small; it is found that essentially linear phase modulation occurs. The general solution is found for the case in which the electrical length of section is long and the permittivity of the medium sinusoidally modulated. The optimum length found to give the greatest frequency deviation is shown to be generally impracticable.

It appears that ferroelectric or ferrimagnetic velocity-modulated dielectrics are feasible, at least for low-power modulators.

Heat Loss of Circular Electric Waves in Helix Waveguides—J. A. Morrison (p. 173)

This paper presents a theoretical calculation of the eddy current losses of circular electric waves in a closely-wound helix waveguide. The wire diameter is assumed large compared to the skin depth, but small compared to the guide diameter and the operating wavelength, so that the fields near the wire are quasi-static and may be determined by conformal mapping.

When the wires are in contact, the waveguide wall is effectively a metal surface with grooves of semicircular cross section, the current flow being parallel to the direction of the grooves. The power loss for this case is computed to be about 8.5 per cent higher than in a waveguide with smooth metal walls. When the wires are not in contact, the wall is treated as a grating of parallel, round wires. The increase in power loss over a smooth surface is approximately 22.5 per cent when the wires are separated by a distance equal to their diameter.

The Expansions of Electromagnetic Fields in Cavities—Kaneyuki Kurokawa (p. 178)

In the theory of cavity resonators, the assumptions are frequently made that every irrotational function can be represented as the gradient of a scalar and that every divergenceless function can be represented as the rotation of a vector. These are, however, not necessarily correct. This paper corrects these misleading assumptions and describes "the theory of cavity resonators" which supplement the classical theory of Slater.

Broad-Band Calorimeters for the Measurement of Low and Medium Level Microwave Power—I. Analysis and Design—M. Sucher and H. J. Carlin (p. 188)

Design considerations for a group of broad-band calorimetric power meters, capable of accurately measuring low (zero to one milliwatt) and medium (zero to 100 milliwatts) power levels over a frequency range from zero to 75,000 mc, are presented. The power meters are of the nonadiabatic, twin, dry-load type and utilize the substitution of dc power. The conflicting requirements imposed upon the design by the need to realize broad-band performance, adequate sensitivity, reasonably short response time, negligibly small rf-dc equivalence error, freedom from "zero" drift and from other types of error are discussed. An analysis is given of the known sources of error which enables the accuracy of the individual instruments to be reliably estimated.

Broad-Band Calorimeters for the Measurement of Low and Medium Level Microwave Power—II. Construction and Performance—A. V. James and L. O. Sweet (p. 195)

The construction and performance of a series of rugged, broad-band twin-Joule calorimeters, using dry loads, are described. These calorimeters operate over the frequency range of 0 to 75,000 mc. The over-all measurement error, computed as the rms value of the maximum individual errors from known independent sources, is shown to lie between 1 and $2\frac{1}{2}$ per cent for power levels between 1 and 100 mw. Power measuring techniques are discussed and a method using the heating and cooling cycle of the calorimeter is described in detail. Power comparison measurements between the calorimeters and several bolometer mounts illustrate the increasing inefficiency of bolometer mounts with increasing frequency.

Amplitude Stabilization of a Microwave Signal Source—G. F. Engen (p. 202)

Recent developments in the microwave field have provided new tools for use in regulating the output amplitude of a microwave signal source. An amplitude or power stabilizer has been constructed at the National Bureau of Standards Boulder Laboratories, using the recently developed self-balancing dc bolometer bridge and a commercially available, electrically controlled, ferrite attenuator which achieves power stabilities of a few parts in 10^4 per hour.

Use of a high directivity directional coupler permits stabilization of the forward traveling component of the signal, thus providing the equivalent of a *matched*, stable generator. In practice, a broad-band source match of vswr less than 1.05 is achieved, and this figure may be further improved, at a given frequency, by suitable tuning. In addition, the device has applications as a precision broad-band attenuator, since known changes in power level may be achieved by switching certain of the associated dc components.

A Simple Artificial Anisotropic Dielectric Medium—R. E. Collin (p. 206)

The anisotropic properties of an infinite stack of thin dielectric sheets separated by another set of thin sheets with a different dielectric constant is investigated. It is shown that the anisotropic properties are brought about because of the two distinct modes of propagation which can exist in such a stacked array of sheets. The limiting forms of the wave solutions and second-order results for the equivalent dielectric constants are given.

Calculation and Measurement of the Noise Figure of a Maser Amplifier—J. C. Helmer and M. W. Muller (p. 210)

The noise performance of regenerative amplifiers is reviewed and equations are obtained which serve to interpret a measurement of noise from an ammonia molecular beam maser amplifier.

The measurement is accomplished by means of a double heterodyne system in which a detuned maser oscillator serves as second local oscillator. The measured noise figure is 3.5 ± 0.5 db, as predicted by theory for the slightly undercoupled circuit used. Although no beam noise is observed, the experimental uncertainty places an upper limit of 40°K on the spontaneous emission noise temperature of the ammonia beam.

Propagation in Dielectric Slab Loaded Rectangular Waveguide—P. H. Vartanian, W. P. Ayres, and A. L. Helgesson (p. 215)

Propagation in dielectric loaded rectangular waveguide is investigated theoretically for varying slab thickness and dielectric constant. The slabs are placed across the center of the waveguide in the E plane. This geometry is found to offer bandwidths in excess of double that of rectangular waveguide for dielectrics having dielectric constants of approximately 18. Power handling capacities which are double or triple that of standard waveguide are achievable using the dielectric loaded waveguide. In addition to the theory, design curves of bandwidth, guide wavelength, cutoff wavelength, impedance, power handling capacity, wall losses, and dielectric losses are presented and compared to experiment were possible.

Parallel-Coupled Transmission-Line-Resonator Filters—S. B. Cohn (p. 223)

This paper describes the synthesis of bandpass transmission-line filters consisting of series of half-wavelength resonant conductors such as strips. The design differs from the usual end-coupled strip configuration in that successive strips are parallel coupled along a distance of a quarter-wavelength. The resulting coupling between resonators is partly electric and partly magnetic. Several important advantages are gained by this arrangement: 1) the length of the filter is approximately half that of the end-coupled type; 2) the gaps are larger and therefore less critical; and 3) the insertion-loss curve is symmetrical on a frequency scale with the first spurious response occurring at three times the center frequency of the pass band.

Formulas are derived for the parallel-coupled-resonator transmission-line filter that permit accurate design for Tchebycheff, maximally flat, or any other physically realizable response. The formulas are theoretically exact in the limit of zero bandwidth, but frequency-response calculations show them to give good results for bandwidths up to about 30 per cent. An experimental strip-line filter of this type has been constructed, and the data given in this paper show that excellent performance has been obtained.

A New Class of Broad-Band Microwave 90-Degree Phase Shifters—B. M. Schiffman (p. 232)

In the type of circuits considered here, the input power is divided equally between two channels whose outputs are caused to have a very nearly 90° phase difference over a broad frequency range. Networks suitable for application at low frequencies which perform the above function have been widely investigated. This report describes a new type of 90° differential phase shifter which has a constant resistance input, and which is useful over band-

widths as large as 5:1 in the microwave region.

Production Techniques

PGPT-3, April, 1958

(*Papers Presented at the First National Symposium on Production Techniques, June 6–7, 1957, Washington, D. C.*)

The Use of Commercially Available Automation Equipment—D. D. Israel (p. 1)

As chairman of the first session: "Management, How to Prepare for and Implement Automation," the author points out that the session not only focusses upon a subject that all recognize as of primary interest to many facets of the electronics business; but that it also draws a precision bead on the strengths and weaknesses of our progress to date, and upon the problems that confront us in the further development of our automation programs.

The author outlines Emerson's operating policy to buy, not build, automation equipment. The program of expediting and "applying needles" is described. Six important advantages of buying equipment are discussed at some length: (1) recourse of exchange, return, or correction, (2) benefit of professional consulting services, (3) access to professional specialists of competitors, (4) art learned faster because competition puts its best foot forward to sell, (5) objectivity avoids too much mechanization, and (6) best design may be chosen, section by section.

The Use of Company-Developed Automation Machinery—George Harrigan (p. 3)

Steps in the implementation of automation at Admiral are outlined. The selection, development and testing of new component-part forms are described at some length. Standardization and design details are covered. Stressed is uniformity in product and procedure and recent mechanized equipment improvements. In conclusion, the author prognosticates upon the future gains from automation development.

A Comptroller's View of the Problems Encountered in Adapting to Automation—J. A. Cadwallader (p. 5)

The automation experiences of GE's Light Military Electronic Equipment Department are outlined. Military job-shop assembly quantities are established at 200 units. In balancing direct-labor savings against machine costs, it is emphasized that indirect and intangible savings should be properly evaluated. Examples are given. The five basic problem areas of obsolescence, depreciation, factory overhead ratios, cost ratios, and capital investment are discussed. The automation of clerical areas is described and answers are given to some of the associated perplexing questions. These questions are outlined as having to do with people, lack of flexibility, overuse of data processing information, and conversion. Encouragement for the future is voiced.

Problems in Manufacturing Component Parts for Automation—A. H. Postle (p. 9)

It is pointed out that by 1960 both printed wiring and automation will be badly needed by the dynamic electronics industry. Outlined is the recent history of the swing to automation by the component parts manufacturer. Those attempts that have yielded less than uniform products, are laid at the door of lack-of-standardization. Some of the present standardization needs are enumerated. Stressed is the thesis that tighter levels of incoming quality are required for automation. Emphasis is placed upon the composite assembly approach using the ceramic base plate. The author concludes with the hope that in the near future some universally accepted system will evolve, so that a component-parts mechanization-increase can effect a total industry saving.

Some Personnel Problems of Automation—Edgar Weinberg (p. 11)

The author presents two basic principles of strategy for orderly transition to automation. Also given are eight generalized solutions to the attendant personnel problems—solutions that were suggested by research studies being conducted by the Bureau of Labor Statistics of the U. S. Department of Labor. Specific examples are given from the case studies of: a TV manufacturer issuing printed wiring and automatic assembly equipment, an insurance company and an airline both using electronic computing systems, and a bakery and an oil refinery representing other industries that are more experienced in automation. The author indicates that, as difficult as the problems of extensive technological change may be, they can still be handled in a constructive and orderly fashion. Answers are suggested to such questions as: how are workers informed about changes; what is the extent of displacement; how were workers reassigned; what was the change in total employment; do new jobs require greater skill; has there been any upgrading or downgrading; and what is the attitude of workers toward the changes? He concludes with an 1835 quotation, "That machines do not, even at their first introduction, invariably throw human labor out of employment must be admitted; and . . . that they never produced that effect . . . ."

A Prediction of the Future Machinery of Automation—R. W. Daniels (p. 14)

Based on the past experience of a company engaged in this field, their own results and the results of others are assessed. Machines, past and present, are discussed covering the last three years. It is concluded that there are none in actual production that can be truly called "automation systems," but that some have reached a significant level of the assembly process. Emphasized is the importance of component-part and packaging standardization, particularly in the direction of narrowing the range of physical sizes and configurations, and in the evolving of simpler physical forms. It is pointed out that effective packaging of parts can eliminate a great deal of complications at the assembly machine.

Three methods of automatic check of machine performance are discussed from the economics point of view. Automatic set-up, approaching the true automation concept, is predicted—again, within economic restrictions. Automatic soldering and automatic testing are treated at some length. The economic considerations of a relatively limited future market are outlined. The rate at which future refinements will be integrated is pictured as dependent primarily upon the economic justification for their development. The author concludes with the prediction that the immediate future will not see a radical change; but that the production segment of the electronics industry must keep pace, as we stand on the threshold of a second industrial revolution.

Reasons for Avoiding an Automation Program at This Time—Jacob Rabinow (p. 17)

The speaker makes the statement that much of the present work in automation seems to be a matter of fashion rather than of hard, common sense thinking. It is charged that projects like the NBS "Tinkertoy," and some of the applications of printed circuits are unjustified economically, or from the point of view of the needs of national defense. It appears to the speaker that a great deal of effort is spent in automating the trivial and the inexpensive.

He emphasizes that there are great dangers in designing equipment to use the modular approach, where the intention is to service the equipment by replacing complete sub-assemblies rather than the individual component parts. The thesis is presented that electronics is too young an industry to be put in a straight jacket of standardized assemblies at this date. The admonition is made that automation is not new, and that as always it should be applied only where very careful cost studies indicate its desirability.

Introductory Remarks—E. R. Gamson (p. 18)

A Production Design Concept for Electro-Mechanical Components—G. D. Gruenwald and William Attride (p. 19)

A unique application of worm gearing to electro-mechanical shaft assembly servos is shown, including a survey of basic requirements, and a comparison of alternate configurations. Possible application of mass production techniques to small lot quantities through design standardization is described. The savings in engineering effort by the selection of standard components having predictable characteristics such as frequency response, backlash, life, etc., are outlined; and additional savings are described, brought about by the use of preprinted vellums.

A Small Three-Dimensional Printed-Wiring Module—A. C. Ansley (p. 26)

The advantages of individual-circuit modular construction are reviewed. Armed Services opinion is quoted, indicating that more modular construction is the most important present trend. A new modular approach is introduced, and several shape-variations which were investigated, are shown. The finally-adopted "PLUS Module" standard is described as consisting of a square socket-mounted top and another square plug-mounted base, tied together by two printed-wiring boards that are slotted at their center and crossed (to form a third-dimensional PLUS symbol). The advantages of using standard printed-wiring techniques and standard component parts, are stressed. Details are discussed, including: a design for either dip or hand-soldering methods, a moisture trap solution for a 32-pin connector, heat-dissipation provisions, and epoxy-resin sealing to cope with extreme conditions of vibration and humidity. Possible constructional variations are mentioned.

Commercial and economic considerations are outlined, and a break-even point is set at 20 units, as compared with hand-wired modules. The author rates the space-factor as comparable with flat plug-in printed-wiring boards that use the minimum practical spacing; and gives the standard dimensions (in a shielding can) as $1\frac{3}{8}$ inches square and $2\frac{3}{4}$ inches high. In conclusion, he expresses the hope that the new "PLUS Module" will help to solve some of the problems of adapting electronic equipment to modular construction.

"Fountain" Soldering—Its Background and Development—Thomas Stearns (p. 31)

Old, tried, and true principles are described as they have been applied to a modified "fountain" soldering technique, to virtually eliminate "drossing" (tin and lead oxidation) and to reduce contamination to a minimum. A history is given of the development of this technique at Sanders, dating from the early days of the "Tinkertoy" program. An early sheet-aluminum stencil and "solder boat" dipping method is discussed, with relation to the long time-cycle and the trapped-gasses problem. Mentioned, is an improvement of bringing the solder to the work in solder cups, but which was mechanically grotesque and faced with the twin problem of thermal-loss and dross-formation. A one-shot pump modification is credited with being the forerunner of the true "fountain."

Presented as the final solution, is a positive-displacement pump that supplies solder through a baffled manifold to selected tube sizes, giving a steady flow of solder which "fountains" to a uniform height above the tubes. Advantages of the "fountain" technique

are listed, with emphasis being placed on accurate temperature control and lower soldering temperatures. A refinement is introduced in the form of an anti-oxident oil layer to control drossing, as used in the tin-ware industry in the "flow-brightening" of electro-deposited tin. Other precautions are outlined which reduce contamination.

Growing out of this soldering technique for printed wiring, a resistor-wafer soldering machine is described (for stacked-wafer modules using standard resistors). Also presented, is a "pile-driver" soldering machine for wafers using printed silver patterns and printed component parts.

The Chairman Keynotes the Session and Presents a Defense Profile—J. M. Bridges (p. 35)

Military Problems Imposed by Automation—W. I. Bull (p. 36)

The problems of logistics, maintenance and reliability, as imposed by automation on users of military electronics, are presented. It is reminded that the military is very careful to avoid telling industry how to build equipment, and therefore, in theory, automation should be only of military academic interest. On the other hand, with the cost of military electronics sometimes reaching staggering proportions, and with the economic status of the country as important as its military prowess, it is pointed out that the armed services have had a definite and active interest in automatic assembly equipment. The characteristics of naval shipboard equipment are enumerated as embracing MIL standard component parts, and a relatively small number of non-interchangeable units repairable on shipboard at the component-parts level. Describing the Bureau's policy as leaning toward a broad production base with small business participation, it is noted that machines for automatic production are costly.

Naval automation objectives compatible with these basic conditions are listed as including the use of conventional MIL component parts to build a wide variety of units, by means of low cost machinery which is capable of flexible programming. Emphasized is the need for selling the new concept to field personnel, and for expediting general specification changes in order to keep pace with technological progress. Automation is credited with a considerable contribution to reliability. It is suggested that, through the use of interchangeable subassemblies or modules, advantages are gained; yet it is indicated that the Navy would still want to maintain the ability (when the chips are down) to repair "the throw-away subassemblies." It is concluded that there should be no insurmountable difficulties in making changes to the present logistics and maintenance methods, in order to realize the full benefits of automatic assembly.

Development of Standards for Automation—A. W. Rogers (p. 39)

In implementing automation it is reminded that the industry is faced with frustrations and dissatisfactions for lack of adequate progress in the standardization of the many facets of mechanized assembly—patterns, parts, assembly, inspection. An advance look is given of an R & D document, to be promulgated by the Signal Corps, titled "Design Requirements for Printed Wiring Electronic Equipment." It is specified that modular units be employed, using printed wiring internally, also integrated through the use of printed wiring, and made adaptable to automation. Etched copper laminate is listed as preferred, line widths and spacings are specified, and the use of epoxy-type conformal coating is required unless the conductors are spaced 4 or 5 times the specified distance. It is presented that the diameter of component-parts leads should be within 0.015" of the mounting hole diameter, with clinching or flaring preferred, eyelets or other auxiliary devices to be avoided whenever possible. and through-plated holes to be paralleled with jumpers.

Other advance information shows temperature control at 520°F and 5 seconds, and rosin-base flux in accordance with MIL-F-14256 (Sig C). Modular hole dimensions are tabulated, the essential information being: 0.300" or 0.150" spacing for "in-line" leads; 0.100" or 0.050" for other leads, with the smaller dimensions acceptable for subminiaturization; and additionally, 0.025" permitted only for extreme subminiaturization. Modular dimensioning is described as the cornerstone of the projected overall automation system; through its use the hope is expressed that we can raise our overall volumetric packaging efficiency from 15% to 40 or 50%

The Problem of Compatibility Between Military Equipment Design and Commercial Automation Techniques—J. J. Lamb (p. 45)

The thesis is presented that standardization is the key to compatibility between military equipment design and commercial automation techniques. The definition of "automation," as authorized by the Electronic Industries Assn. (formerly RETMA) is given. The cognizant standardization organization, Panel "A" within EIA on Automation and Computers is described as consisting of seven main committees with most made up of several subcommittees. It is stated that technical representatives of non-member companies are invited to participate fully in the work; and coordination with IRE headquarters is noted.

The first national standard for the Baudot Code is mentioned, and A-3's proposed specification "Standard Dimensional System for Automation Requirements" (covering modular hole spacing) is reproduced in its entirety. A few of the "new attacks" on automation that Remington Rand has been developing, are pictured. A single drill head, a multiple drill head, and a component-parts insertion machine are shown, which operate on the "template programming" principle, using a pantograph-type mechanism. It is mentioned that the principle is adaptable to automatic programming if economically warranted. Reliability of both insertion and the product is discussed, and 99.6% is labeled as too low. An admonition is given to "Keep it simple," and emphasis is placed upon the importance of military feedback from operational units in the field.

Automating Small-Lot Electronic Production—W. A. Schneider (p. 50)

It is indicated that new developments of promise for automation of military electronics have yet to reach actual production. A plea is made for a self-instituted program of standardization and mechanism. The various types of semiautomatic, mechanized, and programmed varieties of automation machinery for electronic assembly, are described, and commented upon at some length. A tabulation of their numerous characteristics is contributed. Two graphs present short-run production rates of the eight basic types of machines, when used to assemble 24 or 48 component parts to printed-wiring boards; the number of units to be made are plotted from 0–100. Two other graphs amplify the lower portion and are plotted from 0–10 units. Considerable comment is included.

Indexes show cost, manpower and space requirements for batches of 30, 100, and 1000 units, and are tabulated against each of the eight varieties of automation. Suggestions are made for the adoption of mechanization. The importance of design modification, modular conversion, and standardization of component parts, are stressed. The results of a specific redesign are cited. Emphasis is placed upon the thesis, that before being submitted to production, the development of a mechanically-producible electronic system must have arrived at a more complete stage of design than one to be manually assembled. Several benefits to be obtained from automation, and some problems, are listed. It is concluded that military electronics development and production can be improved substantially by the action that is suggested.

The Component-Parts Bottleneck in a Peace-Time Production Profile—Louis Kahn (p. 60)

Although the title is addressed to peacetime conditions, it is pointed out that the possibility of mobilization for war effort, needs to be considered, at least as a background for decisions. It is indicated that mechanization in the component-parts industry is not new, having had substantial beginnings 20 years ago. The author states, although there are in operation ceramic capacitor assembly machines, capable of producing a minimum of 5000 completed capacitors per hour, that general mechanization of the component-parts industry is almost economically impossible. An analysis of production records shows that 75% of all orders are for less than 100 pieces, that 25 to 40 thousand different capacitors are manufactured each month and that 80 to 100 new designs or variations are created each week, at Aerovox.

It is stressed that the multiplicity of types and ratings makes it impossible to speed deliveries, maintain a sizable stock, or implement a general automation rogram. Materials are pictured as not being a bottleneck item on the standard parts level; but on the other hand, that special materials are badly needed to meet the extended environmental and operational conditions forced upon the industry by new weapon requirements. A step-up in the Military preferred parts activities is recommended as a move to drastically reduce the number of standard parts. Some concern is voiced regarding the controversial Aircraft Industry practice of commissioning small suppliers to design and build "specialty items." It is charged that the practice creates still more items purchased from sources where automation is even less practicable, and that reduces reliability through the sponsorship of a hand-assembly atmosphere.

Brain Pickings—"Storage Random Access (Human)" (p. 63)

News of the Chapters (p. 66)

Administrative Committee of the IRE Professional Group on Production Techniques (p. 72)

Standing and Working Committees of the PGPT Administrative Committee (p. 73)

Calendar of Coming Events (p. 74)

Abstracts and References

Compiled by the Radio Research Organization of the Department of Scientific and Industrial Research, London, England, and Published by Arrangement with that Department and the *Electronic and Radio Engineer*, incorporating *Wireless Engineer*, London, England

NOTE: The Institute of Radio Engineers does not have available copies of the publications mentioned in these pages, nor does it have reprints of the articles abstracted. Correspondence regarding these articles and requests for their procurement should be addressed to the individual publications, not to the IRE.

The number in heavy type at the upper left of each Abstract is its Universal Decimal Classification number and is not to be confused with the Decimal Classification used by the United States National Bureau of Standards. The number in heavy type at the top right is the serial number of the Abstract. DC numbers marked with a dagger (†) must be regarded as provisional.

The Index to the Abstracts and References published in the PROC. IRE from February, 1956 through January, 1957 is published by the PROC. IRE, May, 1957, Part II. It is also published by *Electronic and Radio Engineer*, incorporating *Wireless Engineer*, and included in the March, 1957 issue of that journal. Included with the Index is a selected list of journals scanned for abstracting with publishers' addresses.

ACOUSTICS AND AUDIO FREQUENCIES

534.121.2 1290
Variational Treatment of Arbitrarily Mass-Loaded Membranes—E. T. Kornhauser and D. B. Van Hulsteyn. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1204–1205; November, 1957.) See 921 of 1954 (Kornhauser and Mintzer).

534.133 1291
Vibrations of a Monoclinic Crystal Plate—F. G. Newman and R. D. Mindlin. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1206–1218; November, 1957.)

534.21 1292
Acoustical Radiation from a Point Source in the Presence of Two Media—D. I. Paul. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1102–1109; October, 1957.) Two semi-infinite isotropic media (porous or nonporous) are separated by a plane interface. Expressions for the resultant wave function are obtained by a method of steepest descents in a form applicable also to the analogous electromagnetic case. See also 3387 of 1947 (Rudnick).

534.21-8-14 1293
Ultrasonic Absorption by Steady Thermal Method—S. Parthasarathy and S. S. Mathur. (*Ann. Phys., Leipzig*, vol. 19, pp. 242–246; December 20, 1956. In English.) A new method of determining ultrasonic absorption coefficients in liquids is described.

534.21-8-14 1294
Ultrasonic Pulse Technique for Measuring Acoustic Losses and Velocities of Propagation in Liquids as a Function of Temperature and Hydrostatic Pressure—H. J. Meskimin. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1185–1192; November, 1957.)

534.212:534.121.2 1295
On the Transmission of a Spherical Sound Wave through a Stretched Membrane—G. L. Lamb, Jr. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1091–1095; October, 1957.)

534.22-14:546.212 1296
Speed of Sound in Water by a Direct Method—M. Greenspan and C. E. Tschiegg. (*J. Res. Natl. Bur. Stand.*, vol. 59, RP 2795, pp. 249–254; October, 1957.) The speed of sound in distilled water was measured over the temperature range 0°–100°C to within 1 part in 30,000. Details of the method and equipment are given and results are tabulated.

534.23 1297
Representation of the Field of an Acoustic Source as a Series of Multipole Fields—H. L. Oestreicher. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1219–1222; November, 1957.)

534.23 1298
Field of a Spatially Extended Moving Sound Source—H. L. Oestreicher. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1223–1232; November, 1957.)

534.231 1299
Oppositely Directed Plane Finite Waves—R. D. Fay. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1200–1203; November, 1957.) An adaptation of the analytical method described earlier (1300 of 1957) to determine the sound field of waves progressing in opposite directions.

534.232-14-8:534.133 1300
Relation between Efficiency of Quartz Transducers and Ultrasonic Absorption Coefficient of Liquids: Parts 1 and 2—S. Parthasarathy and P. P. Mahendroo. (*Z. Phys.*, vol. 147, pp. 573–581; February 6, 1957. In English.) Report on experimental investigations of absorption in organic liquids and discussion of the resulting transducer efficiency curves. See also 324 of 1957 (Parthasarathy and Narasimhan).

534.24 1301
Reflection on a Rough Surface from an Acoustic Point Source—M. A. Biot. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1193–1200; November, 1957.)

534.613 1302
Acoustic Torques and Forces on Disks—J. B. Keller. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1085–1090; October, 1957.)

534.78 1303
Acoustics and Physiology of Phonation—R. Husson. (*J. Phys. Radium*, vol. 18, Supplement to no. 3, *Phys. Appl.*, pp. 23A–35A; March, 1957.)

621.395.61+621.395.62 1304
Real-Power Damping of Electroacoustic Transducers—T. Hayasaka. (*Rep. Elec. Commun. Lab., Japan*, vol. 5, pp. 1–3; September, 1957.) Real-power damping is defined. Design methods using this are similar to those minimizing vector power damping.

621.395.623.7.012 1305
Panoramic Representation of the Sound Field—G. G. Sacerdote and C. Bordone-Sacerdote. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1165–1168; November, 1957.) Note of records obtained using a sonograph method of recording the sound field of a rotating loudspeaker fed by white noise. See 2950 of 1956.

621.395.625.3:621.397.5 1306
The Ampex Video Tape-Recording System—Snyder. (See 1568.)

ANTENNAS AND TRANSMISSION LINES

621.314.22:621.317.343.2 1307
Measurement of the Characteristic Impedance of a Coaxial Cable—D'Alton. (See 1496.)

621.372.2 1308
The Helical Line with a Coaxial Cylindrical Attenuating Layer—G. Landauer. (*Arch. Elek. Übertragung*, vol. 11, pp. 267–277; July, 1957.) The attenuated helical line is analyzed by considering a coaxial system formed by an obliquely conducting cylinder with infinitely thin walls and a cylindrical semiconducting outer shell. Attenuation and phase rotation are plotted as a function of the surface resistance of the attenuating cylinder. Two different surface resistances can produce the same attenuation but different phase velocities.

621.372.2:621.318.134 1309
Microwave Magnetic Field in Dielectric-Loaded Coaxial Line—B. J. Duncan, L. Swern, and K. Tomiyasu. (PROC. IRE, vol. 46, pp. 500–502; February, 1958.) Additional experimental work and theoretical considerations show that the probable mode configuration for a coaxial line half filled with a low-loss high-dielectric-constant material suggested in an earlier paper [2023 of 1957 (Duncan, *et al.*)] should be rotated by 180° relative to the dielectric.

621.372.8 1310
Ghost Modes in Imperfect Waveguides—

E. T. Jaynes. (PROC. IRE, vol. 46, pp. 416–418; February, 1958.) Microwave resonances exist which are nonradiating and thus have high Q, although the fields are not enclosed completely by metallic walls. Complicated resonance effects observed in waveguides operating close to the cutoff frequency of a propagation mode are explained. These spurious resonances are analogous to certain phenomena observed in imperfect crystals of solid-state materials.

621.372.8:537.226 1311
Shielded Dielectric Waveguide—H. Uchida, S. Nishida, and H. Shioya. (*Sci. Rep. Res. Inst. Tohoku Univ., Ser. B*, vol. 8, pp. 7–22; June, 1956.) A theoretical analysis of a shielded dielectric waveguide shows the existence of both symmetrical E and H waves. The dominant mode of the latter is easily propagated and has a low transmission loss. They both have similar cutoff frequencies but the E wave may be suppressed.

621.372.8:621.018.75 1312
Influence of Wall Losses on Pulse Propagation in Waveguides—R. Gajewski. (*J. Appl. Phys.*, vol. 29, pp. 22–24; January, 1958.) "Assuming that the resistivity of waveguide walls is not too large, the influence which energy losses have upon the shape of a pulse propagating along the waveguide is discussed. It is found that pulses are damped with a damping factor equal in the first approximation to that in a steady state."

621.372.829 1313
Propagation of Waves in Helical Waveguides—Chiao-Min Chu. (*J. Appl. Phys.*, vol. 29, pp. 88–99; January, 1958.) An analysis is developed for determining the effects of wire size and shape on the attenuation and harmonic fields of monofilar and multifilar helices. Calculated attenuation characteristics agree well with experimental data.

621.372.831 1314
Junction of Smooth Flared Waveguides—D. J. Leonard and J. L. Yen. (*J. Appl. Phys.*, vol. 28, pp. 1441–1448; December, 1957.) "Stevenson's general theory of em horns or flared waveguides [1841 of 1952], which is valid only for geometrical configurations whose cross sections are continuous along the direction of propagation together with their first and second derivatives, is generalized to include sudden jumps in the first derivatives of the cross sections."

621.372.85.011.21 1315
A Physical Interpretation of Impedance for Rectangular Waveguides—J. A. Lane. (*Proc. Phys. Soc.*, vol. 70, pp. 1173–1174; December 1, 1957.) Impedance measurements on narrow transverse films in the center of a rectangular waveguide indicate that the most appropriate definition of impedance for such obstacles is one based on total transmitted power and maximum transverse voltage.

621.372.852.3 1316
An Improved Microwave Attenuator for Military Use—F. L. Rose. (*Bell Lab. Rec.*, vol. 35, p. 418; October, 1957.)

621.372.852.323:621.318.134 1317
A Nonreciprocal Attenuator (Isolator) for 4 Gc/s—J. Deutsch and W. Haken. (*Frequenz*, vol. 11, pp. 217–220; July, 1957.) Description of a 3.8–4.2-kmc ferrite-type isolator; the reflection coefficient is about 1 per cent over the whole frequency range, reverse attenuation is greater than 13 db, and forward attenuation about 0.5 db.

621.396.67:621.396.11 1318
Interaction between Two Aerials—J. Robieux. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 793–796; August 12, 1957.) A general expression for the transmission of energy from one antenna to another is derived on the assumptions that 1) both antennas have directivity and 2) they are so far apart that the radiation of one is unaffected by the other. Applications of the expression are suggested for cases of tropospheric propagation, or radiation involving diffraction.

621.396.67.012.12 1319
Graphical Solution of the Radiation from Aerial Systems Composed of Two Active Elements—V. Caha. (*Slab. Obz., Praha*, vol. 18, pp. 144–149; March, 1957.) A theoretical analysis of the radiation diagrams of antenna systems with two active elements. Details of a direct geometric procedure for drawing these diagrams is given.

621.396.67.029.6+621.372.8](091) 1320
Microwave Antenna and Waveguide Techniques before 1900—J. F. Ramsay. (PROC. IRE, vol. 46, pp. 405–415; February, 1958.) In the years between 1888 when Hertz initially demonstrated radio waves and 1900 when Marconi established radio communication, experimenters developed microwave devices and techniques anticipating much of present-day practice. A historical review is given of antennas and waveguides of the period with some reference to associated microwave techniques.

621.396.674.3–415 1321
Rolled Triangular-Sheet Antennas—J. R. McDougal, S. Adachi, and Y. Mushiake. (*Sci. Rep. Res. Inst. Tohoku Univ., Ser. B*, vol. 8, pp. 125–132; December, 1956.) A dipole made up of two spirally rolled right-angled triangular metal sheets has a power pattern and input impedance suitable for wide-band application in the full-wavelength region. In the half-wavelength region the dipole approximates to an ordinary cylindrical dipole.

621.396.676 1322
Capacity Feed for the Mobile Whip—J. M. Osborne. (*Short Wave Mag.*, vol. 15, pp. 408–410; October, 1957.) Details of a capacitive feed system for a whip antenna for mobile use.

621.396.677 1323
Simple Two-Band Cubical Quad—C. Teale. (*Short Wave Mag.*, vol. 15, pp. 406–407; October, 1957.) Design and constructional details of an inexpensive directional antenna array of small dimensions for 10 and 15 mλ with a gain of 8–10 db.

621.396.677.3 1324
More about the Minibeam—G. A. Bird. (*RSGB Bull.*, vol. 33, pp. 168–172; October, 1957.) Additional constructional and operational details of an antenna array for 10, 15, and 20 m λ of a design described earlier (*ibid.*, vol. 31, pp. 355–358; February, 1956).

621.396.677.7/.8 1325
The Design of Horn-Parabola Aerials—H. Laub. (*Frequenz*, vol. 11, pp. 201–207; July, 1957.) The dependence of the antenna dimensions on the angle of illumination is investigated. See also 2628 of 1956 (Laub and Stohr).

621.396.677.71 1326
Pattern of a Flush-Mounted Microwave Antenna—J. R. Wait. (*J. Res. Natl. Bur. Stand.*, vol. 59, RP 2796, pp. 255–259; October, 1957.) "The numerical results for the far-zone radiation from an axial slot on a circular cylinder of perfect conductivity and infinite length are discussed. It is shown that the results for large-diameter cylinders can be expressed in a universal form that is suitable for pattern calculations for arrays of slots on a gently curved surface." See also 355 of 1957 (Wait and Kates).

621.396.677.71 1327
Radiation from Slots on Dielectric-Clad and Corrugated Cylinders—J. R. Wait and A. M. Conda. (*J. Res. Natl. Bur. Stand.*, vol. 59, RP 2802, pp. 307–316; November, 1957.) "An approximate formula is derived for the radiation pattern of an axially slotted cylinder with a thin dielectric coating. The accuracy of the formula is shown to be sufficient for practical purposes. Using a similar method, the pattern function for a slot on a corrugated cylinder is derived. Extensive numerical results are presented for both dielectric-clad and corrugated cylinders."

621.396.677.8 1328
On Focusing Electromagnetic Radiations—R. W. Bickmore. (*Canad. J. Phys.*, vol. 35, pp. 1292–1298; November, 1957.) "The transmission of electromagnetic energy between two apertures is examined as a function of their sizes, separation, excitation functions, and surface shapes with Fresnel approximations made throughout. Relations are derived which show when it is advantageous to focus the apertures by curving about a spherical surface."

621.396.677.8 1329
Fraunhofer Pattern Measurement in the Fresnel Region—R. W. Bickmore. (*Canad. J. Phys.*, vol. 35, pp. 1299–1308; November, 1957.) The Fraunhofer pattern of an aperture may be obtained by probing the field at a radius of $0.1(l^2/\lambda)$ instead of $2l^2/\lambda$, provided the aperture can be moulded to a radius of curvature less than $2l^2/\lambda$. There is less inherent phase error in the main-lobe region than in measurements on a plane aperture at $2l^2/\lambda$.

621.396.677.81 1330
Reflection of a Plane Wave at a Wire Mesh in the Case of Normal Polarization—V. G. Yampol'skiĭ. (*Radiotekhnika, Moscow*, vol. 11, pp. 33–37; November, 1956.) The reflection of a plane em wave at a wire mesh consisting of equidistant circular conductors is considered for the case in which the field intensity vector of the incident wave is perpendicular to the axes of the conductors.

621.396.677.81:621.396.932.1 1331
Current Distribution on Vertical Cylindrical Reflectors—G. Ziehm. (*Frequenz*, vol. 11, pp. 233–243; August, 1957.) Problems arising in ship df equipment due to the effects of parasitic currents are discussed. The magnitude and distribution of vertical currents are determined for various types of antenna and are shown graphically. Suggestions for minimizing errors are based on these investigations.

621.396.677.833.2:621.396.65 1332
Antennas for 4000-Mc/s Radio Links—R. L. Corke and J. Hooper. (*P.O. Elec. Eng. J.*, vol. 50, part 3, pp. 178–185; October, 1957.) "After a brief discussion of normal radio propagation losses a simple explanation is given of the way microwave antennas function. A particular form of the paraboloidal reflector antenna is then described in some detail."

AUTOMATIC COMPUTERS

681.142 1333
A Unique Approach to Computer Versatility—L. S. Michels. (*Electronic Ind. Tele-Tech*, vol. 16, pp. 72–75, 142; October, 1957.) The need for a computer which can perform specialized computations as well as general data processing is met by the integration of a general-purpose computer with a digital differential analyzer. The design involves a shared storage system and special input/output facilities.

681.142 1334
Description of a Large Electronic Computer—A. P. Speiser. (*Bull. schweiz. elektrotech. Ver.*, vol. 48, pp. 1013–1016; November 9, 1957.) Description of the IBM Type-704 computer which comprises storage units of magnetic tape, core, and drum type.

681.142 1335
Cycle and Delay Time Considerations in a Real-Time Digital Computer—H. Freeman. (*Commun. and Electronics*, pp. 588–593; November, 1957.) Considerable improvement in a computer program can be obtained by careful selection of cycle and delay times. The criterion of information-handling efficiency provides a basis for comparing the quality of various programs.

681.142 1336
Recording and Read-Out Circuits for Binary Numbers in a Magnetic-Drum Storage System—L. Dadda. (*Ricerca Sci.*, vol. 27, pp. 2403–2425; August, 1957.) Details of the storage system used in the computer Type CRC 102A at the Milan Polytechnic. See also 3758 of 1957.

681.142 1337
A Swiss Analogue Computer—E. Jucker. (*Bull. schweiz. elektrotech. Ver.*, vol. 48, pp. 1017–1020; November 9, 1957.) An electromechanical system is described in which quantities are represented by ac voltage amplitudes and by the angle of rotation of shafts in a servosystem.

681.142:537.312.62 1338
Trapped-Flux Superconducting Memory—J. W. Crowe. (*IBM J. Res. Devel.*, vol. 1, pp. 294–303; October, 1957.) "A memory cell based on trapped flux in superconductors has been built and tested. The cell is constructed entirely by vacuum evaporation of thin films and can be selected by coincident current or by other techniques, with drive-current requirements less than 150 ma. The short transition time of the trapped-flux cell indicates its possible use in high-speed memories. The superconductive film memory does not exhibit the problems of 'delta noise' in core memories resulting from the difference in half-select pulse outputs."

681.142:537.312.62 1339
An Analysis of a Persistent-Supercurrent Memory Cell—R. L. Garwin. (*IBM. J. Res., Dev.*, vol. 1, pp. 304–308; October, 1957.) A theoretical model of a storage cell based on thin films of superconductors is discussed. An experimental device built to resemble the model was found to have the predicted behavior.

681.142:621.385.832 1340
A Binary-Weighted Current Decoder—E. J. Smura. (*IBM. J. Res. Dev.*, vol. 1, pp. 356–362; October, 1957.) A method for driving cathode-ray tubes from digital equipment is described, in which the deflection yoke is fed directly from binary-weighted constant-current sources. Design considerations are outlined and a comparison is made with other methods.

681.142:621.396.828 1341
Radio-Interference Control as Applied to Business Machines—Sarley and Hendery. (See 1542.)

CIRCUITS AND CIRCUIT ELEMENTS

621.3.011.21:621.314.7:621.372.52 1342
Negative Impedances, Transistors and Feedback Circuits and Their Interrelations—T. Scheler and H. W. Becke. (*Frequenz*, vol. 11, pp. 207–217, and 250–259; July and August, 1957.) The use of transistors in negative-impedance feedback circuits is investigated and the influence of transistor characteristics on the network parameters is determined theoretically and by measurement. Some practical difficulties are discussed.

621.3.049.7 1343
Mechanics of Electronics—W. D. Cussins. (*Wireless World*, vol. 64, pp. 133–137; March, 1958.) Some suggestions for layout and mechanical design of equipment.

621.3.09 1344
The Influence of the Coefficients of the Transfer Function of a Transmission System on the Output Characteristics as a Function of Time—R. Hofmann and W. Walcher. (*Arch. elek. Übertragung*, vol. 11, pp. 321–324; August, 1957.) The input considered is in the form of a step function.

621.314.22.029.55:621.318.134 1345
On the Use of Ferrites in Wide-Band H.F. Transformers—(*Point to Point Telecommun.*, vol. 1, pp. 22–25; June, 1957.) Design problems encountered using ferrite cores in HF power transformers are discussed, including methods of heat dissipation. Characteristics of two commercially available designs are given.

621.318.57:621.375.132.3 1346
Electronic Switch doubles as Cathode Follower—R. Benjamin. (*Electronics*, vol. 31, pp. 81–83; January 17, 1958.) "Basic two-way electronic switch may be expanded to multiway unit by adding input selector circuits, or may be used as a precision cathode follower by eliminating the selector. Circuit has near-infinite input impedance and near-zero output impedance. Comparator compensation permits accuracy of 0.1 per cent over ±100 volts."

621.318.57:621.396.669 1347
An Electronic Transmitter-Receiver Antenna Switch—E. Arvonio. (*QST*, vol. 41, pp. 32–33; October, 1957.) Description of a twin-triode circuit for instantaneous break-in operation.

621.319.45 1348
Miniaturized Tantalum Solid Electrolytic Capacitors—F. S. Power. (*Bell Lab. Rec.*, vol. 35, pp. 419–422; October, 1957.) The replacement of an aqueous electrolyte by an oxide film results in an improvement in the low-temperature characteristics and an extension of the frequency range. These capacitors also have the advantage of storing more charge per unit volume than previous types.

621.372.001.1 1349
Theory of Electrical Networks with Nonlinear Elements—J. J. Schäffer. (*Arch. Elektrotech.*, vol. 43, pp. 151–168; May 20, 1957.) Mathematical treatment making use of almost periodic functions.

621.372.41:621.318.424 1350
The Occurrence of Abnormal States in Certain Ferroresonant Circuits—M. Panet. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 834–837; August 19, 1957.) Two identical parallel resonance circuits incorporating coils with iron cores are connected in series. Conditions are examined under which an applied alternating voltage is not equally distributed between the circuits. See also 2986 of 1956 (Skalnik).

621.372.412 1351
The Problem of Increasing the Frequency Stability of Crystal Oscillators by means of Compensation—G. Becker. (*Arch. elek. Übertragung*, vol. 11, pp. 289–294; July, 1957.) Herzog's method (2734 of 1952) of improving the Pierce oscillator by including a compensating resistance is investigated. The dependence of the frequency oscillator on the circuit parameters, with or without compensation, is calculated. The advantages of compensation are offset by serious drawbacks especially in crystal-clock applications.

621.372.5:512.831 1352
The Calculation of Linear Circuits—W. Klein. (*Arch. elek. Übertragung*, vol. 11, pp. 341–347; August, 1957.) Circuit components, which may include tubes, transistors and transformers, are set up in the form of a multipole admittance matrix; from this five determinants are obtained.

621.372.5:621.316.82 1353
Theory of Networks of Linearly Variable Resistances—H. Levenstein. (PROC. IRE, vol. 46, pp. 486–493; February, 1958.) A network of fixed resistances and linear rheostats driven by a common shaft can replace nonlinear elements. Such networks are amenable to good design techniques and can be described in a mathematical form analogous to that of fixed RL networks.

621.372.5.018.783:621.375.4.029.4 1354
Theoretical and Experimental Investigation of Distortion in Low-Frequency Junction-Transistor Quadripoles—G. A. Spescha and M. J. O. Strutt. (*Arch. elek. Übertragung*, vol. 11, pp. 307–320; August, 1957.) Expressions are derived for calculating distortion by differentiating the quasilinear hybrid parameters of the quadripole. Methods are described of measuring distortion in junction-transistor quadripoles in earthed-base, earthed-emitter, and earthed-collector circuits. Calculated and measured results are compared for the earthed-base and for the earthed-emitter circuits; close agreement is found. The application of the results of the investigation to the design of transistor amplifiers is discussed. See also 2392 of 1957 (Meyer).

621.372.54:621.372.412 1355
High-Frequency Crystal Filter Design Techniques and Applications—D. I. Kosowsky. (PROC. IRE, vol. 46, pp. 419–429; February, 1958.) Commercially available HF crystal filters exhibit performance characteristics previously attainable only at lower frequencies. Filters normally consist of from two to eight quartz resonators in a lattice or bridge network and are characterized by high Q and stability under extreme environmental conditions. Normal values of Q for filter components in the 1–40-mc range are from 50,000 to 200,000; in certain filters, element Q's exceed 1,000,000. By eliminating the need for multiple frequency conversions, the small rugged filters are useful in many AM, SSB, and FM receivers, as well as single-sideband generators.

621.372.54:621.396.828 1356
Low-Pass Filters for Mobile Use—W. Rudolph. (*QST*, vol. 41, pp. 24–25; October, 1957.) Constructional details are given of simple compact filters for the suppression of radiation at television frequencies from mobile transmitters operating at frequencies below 30 mc and in the 50–54-mc range.

621.372.54.029.62/.63:621.372.2 1357
Transmission-Line Low-Pass Filters—F. Charman. (*Electronic Radio Eng.*, vol. 35, pp. 103–111; March, 1958.) The networks consist of sections of line with suitably constructed lumped shunt capacitance. Design curves are given, and practical models are illustrated.

621.372.543.2 1358
Design of Three-Resonator Dissipative Band-Pass Filters having Minimum Insertion Loss—M. Dishal, B. Sellers, J. J. Taub, and B. F. Bogner. (PROC. IRE, vol. 46, pp. 498–499; Feburary, 1958.) Comment on 2370 of 1957 and authors' reply.

621.372.543.2(083.57) 1359
Band-Pass Filter Design Technique—D. R. J. White. (*Electronics*, vol. 31, pp. 79–81; January 3, 1958.) "Universal curves provide design information for Butterworth and Tchebycheff stagger-tune filter networks for band-pass amplifiers. Required number of stages, center frequency, cutoff frequencies, and stage gain requirements can be determined."

621.373.1:538.632 1360
Galvanomagnetic Oscillators—V. N. Bogomolov. (*Zh. tekh. Fiz.*, vol. 27, pp. 663–674; April, 1957.) The type of Hall-effect oscillator described has an efficiency of 37.5 per cent.

621.373.42 1361
An Ultrastable Keyed V.F.O.—J. M. Shulman. (*QST*, vol. 41, pp. 34–39; October, 1957.) High stability is achieved by constructing the variable tuned circuit, which has a high Q and is of rigid construction, as a single unit separate from the remainder of the oscillator. Clickless keying is obtained by a relatively small change in screen grid potential.

621.373.421.11 1362
Self-Oscillatory Systems with Two Degrees of Freedom at Multiple Frequencies—G. M. Utkin. (*Radiotekhnika, Moscow*, vol. 11, pp. 66–76; October, 1956.) Mathematical analysis of the operation of an oscillator using two coupled tuned circuits resonating at frequencies one of which is an approximate multiple of the other. Expressions are derived for the oscillator frequencies and their instability as a function of the instability of the tuned circuits and the supply voltages.

621.373.43 1363
Idealized Treatment of Relaxation Oscillator Circuits—M. Drăgănescu. (*Bul. Inst. polit. Bucureşti.*, vol. 18, pp. 231–244; January /June, 1956.) Topological analysis of idealized circuits with the introduction of critical-point notation. A correct solution can only be obtained if the system oscillates.

621.373.52 1364
Transistor Oscillator supplies Stable Signal—L. H. Dulberger. (*Electronics*, vol. 31, p. 43; January 31, 1958.) "Colpitts circuit, employing one germanium transistor and one Zener diode, operates from a laboratory regulated power supply to maintain a sine-wave voltage of precise amplitude."

621.373.52:621.373.431.1 1365
Determination of the Frequency of a Transistor Multivibrator between 4 and 4000 c/s—M. Bichara. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 896–898; August 26, 1957.) Experimental and theoretical results supplement those of McDuffie (682 of 1953) for the pulse duration and repetition rate of an astable transistor multivibrator circuit.

621.374.3 1366
Pulse Height Selector with Constant Analysis Time—M. Spighel and L. Pénege. (*J. Phys. Radium*, vol. 18, Supplement to no. 3, *Phys. Appl.*, pp. 19A–22A; March, 1957.) A single-channel pulse height selector with an analysis time constant to within 2×10^{-8}s independent of the pulse height is described.

621.374.3:621.314.7 1367
Unusual Transistor Circuits—P. L. Burton and J. Willis. (*Wireless World*, vol. 64, pp. 107–110; March, 1958.) The main features of design and a physical explanation of the action are given for voltage catching, transistor pump, emitter-squared follower, and transistor gating circuits.

621.374.3.029.65 1368
Millimicrosecond Pulses in the Millimetre-Wave Region—C. A. Burrus. (*Rev. Sci. Instr.*, vol. 28, pp. 1062–1065; December, 1957.) A technique is described for the generation, amplification and detection of pulses in the 5–6-mm region. The pulses are 3 mμs wide at the base, with a peak power of a few milliwatts.

621.374.32 1369
Forty-Megacycle Scaler—M. Nakamura. (*Rev. Sci. Instr.*, vol. 28, pp. 1015–1020; December, 1957.) The scaler has a scale of eight and uses a fast flip-flop as a basic element. The latter can be triggered at a rate higher than 50 mc and has a double-pulse resolution of 20 mμs.

621.374.32 1370
Decade Decimal Counter speeds Printed Read-Out—R. W. Wolfe. (*Electronics*, vol. 31, pp. 88–90; January 17, 1958.) "High-speed circuit uses magnetron beam-switching tubes to sample, store, and provide multi-output functions without stopping the original count or losing input information during readout."

621.375.1.012.6 1371
Amplifier Low-Frequency Compensation—J. E. Flood and J. E. Halder. (*Electronic Radio Eng.*, vol. 35, pp. 92–100; March, 1958.) General expressions are deduced for the indicial response (*i.e.*, response to unit step input) and gain/frequency and phase/frequency response, leading to conditions for maximal flatness. Compensation is then considered, together with the effects of negative feedback. Multistage amplifiers are also discussed.

621.375.1.029.4:621.376.54 1372
Pulse Method for the Amplification of A.F. Oscillations—V. V. Malanov. (*Radiotekhnika, Moscow*, vol. 11, pp. 38–46; October, 1956.) The advantages and efficiency of a pwm method are discussed.

621.375.132.018.756 1373
Dynamic Range of Negative-Feedback Pulse Amplifiers—V. Pauker. (*Bul. Inst. polit. Bucureşti*, vol. 18, pp. 245–252; January/June 1956.) Analysis of the limiting conditions for grid current in two- or three-stage amplifiers.

621.375.2.029.3 1374
More Transformerless Amplifiers—(*Wireless World*, vol. 64, pp. 145–146; March, 1958.) A review of recent work published since an earlier article (1030 of 1957). Single-input series-connected and Petersen-Sinclair output stages (1250 of 1952) are discussed.

621.375.4.029.33 1375
Practical Circuits of Video Amplifiers using Junction Transistors—T. M. Agakhanyan and V. A. Volkov. (*Radiotekhnika, Moscow*, vol. 11, pp. 38–44; November, 1956.) Practical circuits with two types of Russian transistor are described. Complex feedback is used, in order to reduce the distortion of the leading edges of the pulses.

621.375.9:538.569.4.029.63:621.396.822 1376
Noise Temperature Measurement on a Solid-State Maser—A. L. McWhorter, J. W. Meyer, and P. D. Strum. (*Phys. Rev.*, vol. 108, pp. 1642–1644; December 15, 1957.) Noise measurements have been made on a three-level, 2800-mc maser operating at 1.25°K, with sufficient accuracy to establish that its noise temperature does not exceed 20°K.

621.375.9:621.385.029.64:537.533 1377
A Parametric Electron-Beam Amplifier—Bridges. (See 1608.)

621.376.32 1378
The Optimum Design of a Tunable Multiplier for Frequency-Modulated Oscillations—H. Schönfelder. (*Frequenz*, vol. 11, pp. 244–249; August, 1957.) The design of tunable frequency multipliers, containing simple one- or two-stage filters in the individual multiplier stages, is investigated. 70-db adjacent-channel attenuation and 3 per cent distortion can be achieved if the multiplication per stage does not exceed 5:1. The fundamental frequency should not be less than about 1 mc.

621.376.332 1379
Combined Limiter and Discriminator—J. W. Head and C. G. Mayo. (*Electronic Radio Eng.*, vol. 35, pp. 85–88; March, 1958.) The limiter incorporates a third-harmonic rejector in series with the diode. Distortion less than 0.1 per cent of any harmonic is easily obtained, and stray capacitances are small, leading to high sensitivity and output.

621.376.5 1380
Pulse Detector with an Inductance-Capacitance Filter—E. L. Gerenrot. (*Radiotekhnika, Moscow*, vol. 11, pp. 30–37; October, 1956.) An ideal detector with an LCR filter is considered taking into account the internal impedance of the current source. A method is proposed for calculating the voltage at the load, when detecting pulses of arbitrary shape, and the necessary formulas are derived.

GENERAL PHYSICS

530.112:530.12:531.18 1381
The Ether and the Special Theory of Relativity—A. Datzeff. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 827–829 and 891–894; August 19 and 26, 1957.) An ether which is immobile in its immediate surroundings but capable of a movement of translation is suggested. It is hoped that this hypothesis will provide a physical explanation of electrodynamic and other phenomena and will not be inconsistent with the results of the special theory of relativity.

535.56–15 1382
Polarimetry in the Infrared—R. Duverney and A. M. Vergnoux. (*J. Phys. Radium*, vol. 18, pp. 526–536; August/September, 1957.)

537.12 1383
Connection between Electron Temperature Determination by means of the Langmuir Probe Method and the Two-Probe Method—V. I. Tverdokhlebov. (*Zh. Tekh. Fiz.*, vol. 27, pp. 753–755; April, 1957.)

537.122:539.152 1384
The Lamb Shift—(*Electronic Radio Eng.*, vol. 35, pp. 52–55 and 89–91; February and March, 1958.) An elementary account of present-day knowledge concerning the electron, and a discussion of Lamb and Retherford's experiment (see *e.g.*, *Rep. Progr. Phys.*, vol. 14, pp. 19–63; 1951) in relation to the anomalous magnetic moment of the electron.

537.226:537.311.3:530.17 1385
Use of Complex Conductivity in the Representation of Dielectric Phenomena—F. A. Grant. (*J. Appl. Phys.*, vol. 29, pp. 76–80; January, 1958.) A representation for complex conductivity is proposed in which a simple Debye mechanism results in a semicircular locus. A nonzero value of dc conductivity does not cause this plot to lose its unique shape in the low-frequency range.

537.312.62 1386
Theory of Superconductivity—J. Bardeen, L. N. Cooper, and J. R. Schrieffer. (*Phys. Rev.*, vol. 108, pp. 1175–1204; December 1, 1957.)

537.321:[621.314.63+537.311.4 1387
Temperature Rise of Solid Junctions under Pulse Load—E. J. Diebold. (*Commun. and Electronics*, no. 33, pp. 593–598; November, 1957.) The magnitude and duration of current pulses are correlated with temperature rise in the junction. Use of successive approximations assuming that various parts of a body 1) are not yet subjected to a temperature increase or 2) have already reached a temperature differential corresponding to the steady state, reduces the problem to a simple case of heat diffusion.

537.533:621.385.029.6 1388
Charged Particles in a Nonuniform Radio-Frequency Field—H. A. H. Boot and R. B. R. S. Harvie. (*Nature, London*, vol. 180, p. 1187; November 30, 1957.) Under certain conditions particles of either sign will experience an acceleration towards the position of least field strength. Such an acceleration is demonstrated in a 10-cm slotted magnetron having no magnetic field and a cold cathode.

538.221 1389
Theory of Ferromagnetic Anistropy—W. J. Carr, Jr. (*Phys. Rev.*, vol. 108, pp. 1158–1163; December 1, 1957.) By means of a virial theorem and perturbation theory, the anisotropy energy of a ferromagnetic crystal is expressed in terms of the Coulomb energy alone. This is approximated by a multipole expansion and the anisotropy constants are given in terms of electric multipole moments and crystalline-potential constants. The multipole moments which arise from the orbital angular momentum induced by spin-orbit coupling have been estimated from the known values of angular momentum.

538.221:548.0 1390
Effect of Crystalline Electric Fields on Ferromagnetic Anisotropy—W. P. Wolf. (*Phys. Rev.*, vol. 108, pp. 1152–1157; December 1, 1957.) The effect of the electrostatic crystalline field has been considered for a magnetic crystal in which the ions are strongly coupled by ferromagnetic exchange. On the basis of the one-ion approximation that the exchange can be represented by a Weiss molecular field, expressions for the anisotropy constants have been derived. The treatment assumes that the magnetic electrons can be considered as localized on the individual ions, and thus applies primarily to nonmetallic substances such as ferrites.

538.244:538.221 1391
The Time Sequence and Amplitude Distribution of Barkhausen Jumps—J. Kost. (*Z. Phys.*, vol. 147; pp. 520–530; February 6, 1957.) Report and discussion of results obtained in investigations of the magnetization of Fe-Ni alloy wire specimens to determine the dependence of Barkhausen-jump characteristics on field strength and specimen shape.

538.566 1392
Some New Aspects of the Reflection of Electromagnetic Waves on a Rough Surface—M. A. Biot. (*J. Appl. Phys.*, vol. 28, pp. 1455–1463; December, 1957.) The roughness is represented by electromagnetically interacting hemispherical bosses whose radii and mutual distances are small relative to the wavelength. The effects of such a surface on both vertically and horizontally polarized radiations are considered as functions of the angle of incidence.

538.566:535.42 1393
Apparatus for the Experimental Study of the Diffraction of Centimetre Waves—J. Mével. (*J. Phys. Radium*, vol. 18, Supplement to no. 3, *Phys. Appl.*, pp. 45A–53A; March, 1957.) Apparatus operating at 1.25 cm λ is described for determining the phase and intensity at any point of the em field. Two versions are presented: one for the investigation of diffraction near the axis and the other for the distant diffraction field. The effect of diffracting bodies less than λ/100 in size can be detected.

538.566:535.42 1394
Diffraction of 3.2-cm Electromagnetic Waves by Dielectric Rods: Part 3—Lucite 1½-in-Diameter Semicylinder, Fields Very Close to Surface—C. E. Jordan and A. B. McLay. (*Canad. J. Phys.*, vol. 35, pp. 1253–1264; November, 1957.) The diffraction field is examined very close to the surface of a lucite semicylindrical rod with its plane surface at various orientations relative to the axis of propagation of the incident wave. Parts 1 and 2: 93 of 1957 (Subbarao and McLay).

538.569.4 1395
The Microwave Spectrum of TlI and $BiCl_3$ in the 3-cm and 1.5-cm Bands—H. Happ. (*Z. Phys.*, vol. 147, pp. 567–572; February 6, 1957.)

538.569.4:537.525 1396
The Characteristic Frequencies of Negative Molecular Ions of Oxygens between 3 and 13 Mc/s, Observed in Ionized-Air Discharge Tubes—T. V. Ionescu and O. C. Gheorghiu. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 898–901; August 26, 1957.) Absorption frequencies are determined from the current variations in a coil round a discharge tube containing pure air, supplemented in some cases by O_2 at pressures of 10^{-2}–10^{-5} mm Hg. Absorption frequencies above 5 mc give rise to strong absorption at pressures below 10^{-3} mm Hg.

538.569.4.029.63:621.375.9:621.396.822 1397
Noise Temperature Measurement on a Solid-State Maser—McWhorter, Meyer, and Strum. (See 1376.)

GEOPHYSICAL AND EXTRATERRESTRIAL PHENOMENA

523.164.32 1398
Radio Pictures of the Sun—W. N. Christiansen, D. S. Mathewson, and J. L. Pawsey. (*Nature, London*, vol. 180, pp. 944–946; November 9, 1957.) A radioheliograph erected at Sydney, Australia, combines the principles of the multi-element or grating interferometer and the Mills cross. It consists of two rows of parabolic antennas of 19-foot diameter arranged in the form of a cross, each row being 1240 feet long. Diagrams of the lower corona have been prepared from observations of 21-cm-λ radiation.

523.164.32 1399
Investigation of Scintillation of the Sun Observed at a Wavelength of 3.2 cm—I. Kazès. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 636–639; August 5, 1957.) Results of observations including simultaneous measurements at 3.2 and 34 cm λ indicate that scintillations are of atmospheric origin and are related to zenithal height, solar activity and wavelength. The apparent diameter of the sun also fluctuates.

523.164.32 1400
Investigation of the Scintillation of the Sun Observed using Several Aerials on a Wavelength of 3.2 cm—I. Kazès and J. L. Steinberg. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 782–785; August 12, 1957.) Results of measurements of scintillation made with three receivers at variable distances show that the average dimension of the ground shadows is 170 m; their velocities are compared with those of winds at the altitude of the tropopause.

523.164.32:523.75 1401
Polarization of Solar Radio Outbursts—K. Akabane and T. Hatanaka. (*Nature, London*, vol. 180, pp. 1062–1063; November, 1957.) Changes in the polarization of RF emission at 9 kmc were recorded at Tokyo on July 3, 1957 during a solar flare.

523.3:621.396.9 1402
The Scattering of Radio Waves by the Moon—J. V. Evans. (*Proc. Phys. Soc.*, vol. 70, pp. 1105–1112; December 1, 1957.) Measurements of the rapid fading of echoes at a frequency of 120 mc show that the effective scattering region has a radius of about one third of the lunar radius and is located at the center of the visible disk. Range measurements indicate that the echo is returned from the front edge, with a power reflection coefficient of 0.1. See also 753 of 1957 (Browne, *et al.*).

523.74:538.12 1403
Sweet's Mechanism for Merging Magnetic Fields in Conducting Fluids—E. N. Parker. (*J. Geophys. Res.*, vol. 62, pp. 509–520; December, 1957.) The mechanism is investigated in a semiquantitative manner. If two oppositely directed sunspot fields, having scales of about 10^4 km, were brought together, Sweet's mechanism would allow them to merge in about two weeks. The mechanism gives a means of altering quickly the configuration of magnetic fields in ionized gases and converting magnetic energy into kinetic energy of the fluid.

523.75:551.594.6 1404
The Solar Flare of 23rd February 1956: its Cosmic and Geophysical Effects—R. Bureau and A. Dauvillier. (*J. Phys. Radium*, vol. 18, pp. 512–517; August/September, 1957.) Records of the intensity of cosmic rays and of atmospherics show that a sudden large increase in the former at 0345 UT was accompanied by a sudden decrease in VLF atmospherics, particularly as observed at Bagneux at 11 km λ.

550.38:525.24 1405
The Earth and its Magnetic Field—G. H. A. Cole. (*Sci. Progr.*, vol. 45, pp. 628–645; October, 1957.) Review of modern theory which attempts to explain the presence and properties of the observed magnetic field in terms fully compatible with the structure of the earth determined independently. Fifty references.

550.385:523.75 1406
The Effect of Solar Flares on the Geomagnetic Field—R. Pratap. (*J. Geophys. Res.*, vol. 62, pp. 581–588; December, 1957.) "The dynamo equation is solved for a conductivity produced by solar-flare ultraviolet radiations from the sun. The crochet amplitudes in horizontal field components are then computed and compared with observed results. It is found that fair agreement exists between the theoretical and experimental values only if the seat of the crochet current system is within a few kilometers of the current system producing the quiet-day solar variation."

550.385.4:551.511 1407
Sudden Commencements of Magnetic Storms and Atmospheric Dynamo Action—T. Obayashi and J. A. Jacobs. (*J. Geophys. Res.*, vol. 62, pp. 589–616; December, 1957.) Statistical analysis of world-wide magnetic data discloses an appreciable diurnal change in the amplitude of SC's. The average electric current system for the disturbance diurnal variation shows marked concentrations in polar regions and suggests that the system exists within the earth's atmosphere while the current system of the storm-time variation would appear more likely to be of extraterrestrial origin. A dynamo theory is applied to explain the atmospheric part of magnetic disturbances and a consistent wind system is obtained to produce current systems for magnetically quiet and disturbed days. This system contains diurnal and semi-diurnal terms and its order of magnitude

agrees with recent ionospheric wind observations.

550.389.2:551.510.535 1408
The Geophysical Year and the Ionosphere—R. R. Lejay. (*J. Phys. Radium*, vol. 18, pp. 481–489; August/September, 1957.) Discussion of the organization and program of the IGY with particular reference to ionospheric measurements and techniques.

550.389.2:629.19 1409
Radio Observations of the Russian Earth Satellite—(*Nature, London*, vol. 180, pp. 879–883; November 2, 1957.) Observations were made at 20, 40, and 80 mc at the Mullard Radio Observatory, Cambridge, using interferometer and Doppler techniques. Details of the calculation of the orbit are given and possible new methods of finding the distribution of electron density with height are suggested.

550.389.2:629.19 1410
Observations on the Orbit of the First Russian Earth Satellite—(*Nature, London*, vol. 180, pp. 937–941; November 9, 1957.) Report of calculations carried out at the Royal Aircraft Establishment, Farnborough, Hants, to determine orbit constants from interferometer and Doppler measurements. Observations of signal amplitude and fading indicate that the ray path of long-distance transmission from the satellite to a receiving station is complicated.

550.389.2:629.19 1411
Radar Observations of the First Russian Earth Satellite and Carrier Rocket—(*Nature, London*, vol. 180, pp. 941–942; November 9, 1957.) A note of instruments used and observations made at Jodrell Bank at frequencies of 120 and 36 mc with pulse durations of 2 ms and 150 μs, respectively.

550.389.2:629.19 1412
Further Radio Observations of the First Satellite—I. R. King, G. C. McVittie, G. W. Swenson, Jr., and S. P. Wyatt, Jr: N. Lea. (*Nature, London*, vol. 180, p. 943; November 9, 1957.) Reports of interferometer measurements at 40 mc made at Urbana, Ill., and Doppler measurements at 40 mc made at Chelmsford.

551.51 1413
Charge Transfer Reactions—S. N. Ghosh and W. F. Sheridan. (*Indian J. Phys.*, vol. 31, pp. 337–352; July, 1957.) An improved method for determining charge-transfer cross sections is applied to symmetric and unsymmetric reactions for a large number of gases and gives higher values than originally reported. A comparison of the rates of collisional process in the D, E, and F regions shows that the charge transfer rate is much longer than any other collisional process in the upper atmosphere. See 3365 of 1956 (Ghosh).

551.510.53:539.12.08 1414
An Appraisal of Photon Counter Measurements of Upper-Atmosphere Parameters—G. J. Simmons. (*J. Geophys. Res.*, vol. 62, pp. 565–571; December, 1957.) A model atmosphere has been computed from measurements of density and O_2 concentration using rocket-borne photon counters near 100 km. Inconsistencies have appeared and it is concluded that the density at 100 km is higher than that found using photon counters.

551.510.535 1415
Calculation of Group Indices and Group Heights at Low Frequencies—J. J. Gibbons and B. R. Rao. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 151–162; 1957.) A formula is derived from which the "group index" (designated $\mu'=\mu+f\partial\mu/\partial f$) for normal incidence, including collision effects, can be computed easily as a function of the real and imaginary parts of the refractive index for frequencies below 1 mc. A set of curves can then be prepared from which the group height h' to be expected from a given electron-density profile can be quickly determined by graphical methods.

551.510.535 1416
The Movements of Sporadic-E-Layer Clouds—G. L. Goodwin. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 177–186; 1957.) The drift speed of isolated clouds was measured by observing the beat between a ground-wave signal from a 1.55-mc CW transmitter and the sky wave reflected from the moving cloud. Comparison of these observations, made at night, with other evidence suggests that scattering is greater than in the day. A method for determining the direction of drift is suggested.

551.510.535 1417
Relation between Virtual and Actual Heights in the Ionosphere—G. A. M. King. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 209–222; 1957.) A mathematical discussion of the various methods of determining the distribution of electron density with height in the ionosphere and a treatment of the comparison and integral methods. An annotated bibliography with thirty-nine references is given.

551.510.535 1418
The Coefficient of Diffusion of Ions in the F_2 Regions—V. C. A. Ferraro. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 296–298; 1957.) The application of Sutherland's molecular model to the F_2 region gives too large a coefficient of diffusion if recent rocket estimates of F_2-layer molecular density are correct. If Langevin's formula for the diffusion coefficient is used instead, the discrepancy is less but, for an F_2 region in which diffusion is inappreciable, calculated values of the coefficient are still about 10 times too large. Possible influencing factors are discussed.

551.510.535:523.164 1419
The Origin of the Ionospheric Irregularities Responsible for Radio-Star Scintillations and Spread F: Part 1—M. Dagg. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 133–138; 1957.) "The present state of knowledge about the irregularities responsible for radio-star scintillations is summarized, and the existing theories of the origin of these irregularities are discussed. All of the suggestions are shown to be inadequate to explain the observed features of scintillations and spread F. It is shown that any ionizing agent from outside the earth's atmosphere is unlikely to be responsible for the ionospheric irregularities that cause radio-star scintillations, and that the mechanism for their production must be sought in the terrestrial atmosphere."

551.510.535:523.164 1420
The Origin of the Ionospheric Irregularities Responsible for Radio-Star Scintillations and Spread F: Part 2—M. Dagg. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 139–150; 1957.) "A theory is presented which attributes the occurrence of ionospheric irregularities in the F region to turbulent wind motion in the dynamo region at a height of 110–150 km. The resulting turbulent component of the electric potential field produced is communicated to the F region, as suggested by Martyn [*The Physics of the Ionosphere*, pp. 163–165; 1955], where magneto-electric forces then cause the ionization to form eddies. It is suggested that the absence of daytime scintillations is due to the inhibition of turbulent flow by large temperature gradients during the day." The theory is then shown to explain the major features of radio-star scintillations, together with the long-term correlation of scintillation amplitude with magnetic activity and the variation of spread F and scintillations at different parts of the earth over the sunspot cycle.

551.510.535:523.164 1421
Diurnal Absorption in the D Region—J. W. Warwick and H. Zirin. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 187–191; 1957.) An analysis of the diurnal variation of cosmic noise at 18 mc from which are derived the D-region electron density and an exponential approximation to the vertical distribution of nitric oxide.

551.510.535:523.164.83 1422
The Electron Content of the Ionosphere—J. V. Evans. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 259–271; 1957.) Report of an investigation noted earlier (765 of 1957) using radio echoes from the moon to determine the electron content. The apparatus and the methods of analysis are described. Results for seven periods between October, 1955 and September, 1956 are given. An almost constant ratio (2:1) is found between observed electron content and that expected in a parabolic layer, There is some evidence for tidal effects.

551.510.535:523.746:621.396.11 1423
A Critical Discussion about Special Ionospheric Characteristics—R. Eyfrig. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 163–176; 1957.) Twelve-month running means of ($M3000$) F_2 and R are compared for 37 locations and linear relations between the two are apparent for limited regions. Deviations from linearity may be genuine at some locations and due to magnetic effects, but may arise from observational errors and inconsistencies at others. Present data are inadequate for examination of M factors on a world-wide basis.

551.510.535:523.746:621.396.11 1424
The Sunspot Cycle and Radio Communications—Millington. (See 1529.)

551.510.535:621.396.11 1425
Some Problems of the Physics of the Ionosphere: Part 1—Fluctuation of the Electron Density and Scattering of Radio Waves—V. L. Al'pert. (*Uspekhi. fiz. Nauk*, vol. 61, pp. 423–450; March, 1957.) Electron density, usw scattering, and turbulence of the ionosphere are discussed and experimental results are analysed. Nineteen references.

551.510.535:621.396.812.3 1426
Curvature-Induced Error in the Analysis of Fading Records—N. J. Rumsey. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 255–258; 1957.) "Curvature of lines of maximum amplitude in a radio field-strength pattern drifting across an array of three receivers can introduce an error into the estimate of the drift velocity, but the error is expected to be large only infrequently. The mean error is smaller for an array of receivers at the corners of an equilateral triangle than for one at the corners of a right-angled triangle."

551.594.221:621.396.96 1427
Radar Observations of Lightning on 1.5 Metres—J. L. Pawsey. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 289–290; 1957.) Details are given of observations made in 1943. The echo cross section was approximately 40 m², the visually estimated echo duration approximately ½ second and the horizontal extent several miles. The associated atmospherics are described.

551.594.5+551.593 1428
Measurements of the Absolute Intensity of the Aurora and Night Airglow in the 0.9–2.0-μ Region—A. W. Harrison and A. V. Jones. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 192–199; 1957.)

551.594.5:621.396.96 1429
Radio Reflections from Aurorae: Part 3—K. Bullough, T. W. Davidson, T. R. Kaiser, and C. D. Watkins. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 237–254; 1957.) A further study of radio echoes at 72 mc following earlier work by Bullough and Kaiser (2622 of 1955). Good correlation is found between the daily frequency distribution of echoes and the mean daily variation in magnetic disturbance at Eskdalemuir. Results and conclusions conflict with the Chapman-Ferraro-Martyn theory.

551.594.5:621.396.96 1430
U.H.F. Radar Observations of Aurora—S. J. Fricker, R. P. Ingalls, M. L. Stone, and S. C. Wang. (*J. Geophys. Res.*, vol. 62, pp. 527–546; December, 1957.) Radar echoes have been obtained on a frequency of 412.85 mc at South Dartmouth, Mass. (41.5° N, 71° W). The diurnal and seasonal variations in the occurrence of echoes is discussed and also the conditions required to obtain echoes.

651.594.6:621.396.11.029.4 1431
The Waveforms of Atmospherics and the Propagation of Very-Low-Frequency Radio Waves—Chapman. (See 1530.)

551.594.9:523.5:621.396.11.029.55 1432
V.H.F. Radar Echoes Associated with Atmospheric Phenomena—G. C. Rumi. (*J. Geophys. Res.*, vol. 62, pp. 547–564; Dec., 1957.) During observations with radar equipment at 27.85 mc, echoes were obtained which could not be attributed either to auroras or to meteors. The characteristics of these echoes suggest that they may be due to "upward discharges" from the troposphere to the ionosphere which may be triggered off by meteors.

LOCATION AND AIDS TO NAVIGATION

621.396.933 1433
The Latest Developments in Radio Navigation in Aeronautics—E. Roessler. (*Electrotech. Z., Edn B*, vol. 9, pp. 335–339; August 21, 1957.) Modern aids to navigation are described and attempts to develop more universal systerms are discussed. Thirty-three references.

621.396.933.2 1434
Bearing Memory Improves Direction Finder—R. E. Anderson. (*Electronics*, vol. 31, pp. 44–48; January 31, 1958.) A high-frequency direction-finding equipment employing Doppler principles is described. The received wave is frequency-modulated by scanning round a fixed circular antenna array at 42 cps. Gaps between the pulses of a coded transmission are filled in by the use of a recording drum rotating at the scanning frequency.

621.396.96:621.396.822:621.317.7 1435
Monitor displays Radar Noise Figures—L. Young. (*Electronics*, vol. 31, pp. 49–51; January 31, 1958.) The receiver noise is compared against a monitor pulse, by passing them through a logarithmic receiver followed by a difference amplifier. The noise figure is displayed directly on a meter calibrated in db.

621.396.967 1436
An Integrated Airborne Radar—J. H. H. Grover. (*Brit. Commun. Electronics*, vol. 4, pp. 628–632; October, 1957.) The airborne landing and approach aid described overcomes inadequacies of present ground-based landing aids.

621.396.969.3 1437
Measurement of the Radar Cross-Section of a Man—F. V. Schultz, R. C. Burgener, and S. King. (PROC. IRE, vol. 46, pp. 476–481; February, 1958.) Doppler measurements were made at five frequencies from 410 mc to 9375 mc using both vertical and horizontal polarization from various angles. The side view presented the smallest radar target and the back view the greatest. The difference between polarizations was greatest at the lowest frequency with horizontal polarization giving the smaller value.

MATERIALS AND SUBSIDIARY TECHNIQUES

535.215.5:537.312.5 1438
The Dependence of Photoelectric Currents on Electric Field Strength—F. Stöckmann. (*Z. Phys.*, vol. 147, pp. 544–566; February 6, 1957.) Photoconductors can be classified in four groups according to conduction mechanism. The differences in their photo-electric characteristics are analyzed as a function of field strength, and type of excitation, and the influence of the electrode contacts on photo-currents is discussed.

535.37 1439
Luminescence—H. G. Jenkins. (*J. Telev. Soc.*, vol. 8, pp. 261–272; July/September, 1957.) After describing various forms of the phenomenon, the nature of luminescence and a qualitative theory are discussed. The characteristics of some important inorganic phosphors are tabulated and the methods of preparation are described. Some uses of these materials in electric discharge lamps, both low- and high-pressure types, is given with particular reference to tube loading and luminous efficiency. The application of electroluminescence in light amplifiers is described, and possible applications in the fields of television, astronomy, and X-ray diagnosis are indicated.

535.37 1440
The Effect of H_2^+ Ions on the Luminescence Properties of Phosphors—W. Martin. (*Z. Phys.*, vol. 147, pp. 582–592; February 6, 1957.) The effects of ion bombardment on various phosphors are discussed with reference to experiments with an ion source of 45-kev maximum energy. From the glow curves obtained it is seen that electron traps are formed. The damage to the phosphors can be greatly reduced by subsequent irradiation by electrons.

535.37:53.082.5 1441
Instrument to Measure Fluorescence Lifetimes in the Millimicrosecond Region—S. S. Brody. (*Rev. Sci. Instr.*, vol. 28, pp. 1021–1026; December, 1957.) Fluorescence decay is displayed on a high-speed oscilloscope; the observed fluorescence is corrected for the response of the hydrogen-discharge flash lamp and the detection equipment. Measured lifetimes are given for various photosynthetic pigments, some organic dyes and chlorophyll fluorescence in algae.

537.227/.228:061.03 1442
First Conference on Ferroelectricity (Leningrad, 19th–24th June 1956)—(*Izv. Ak. Nauk S.S.S.R., Ser. fiz.*, vol. 21, pp. 295–472; March, 1957.) Texts are given of the following 26 papers presented at the conference. Others were noted in 2785 of 1957.

Growth and Investigation of Ferroelectric Single Crystals—N. S. Novosil'tsev, A. L. Khodakov, M. L. Sholokhovich, E. G. Fesenko, and O. P. Kramarov (pp. 295–304).

Some Properties of Single Crystals of $PbTiO_3$ and Single Crystals of Solid Solutions of (Ba, Pb) TiO_3—E. G. Fesenko, O. P. Kramarov, A. L., Khodakov, and M. L. Sholokhovich (pp. 305–310).

Structure of the OH band in Crystals containing Hydrogen Bonds—A. I. Stekhanov (pp. 311–321).

Shape of the Potential Curve of Hydrogen Bonds in Certain Crystals—A. N. Lazarev (pp. 322–328).

Peculiarities of Ferroelectric Properties of Crystals of Rochelle Salt Exposed to Radioactivity—V. A. Yurin (pp. 329–333).

Some Properties of Dielectric Hysteresis of Rochelle Salt—I. Y. Eisner (pp. 343–339).

Microscopic Theories of the Ferroelectric Properties of Barium Titanate—R. E. Pasynkov (pp. 340–351).

Dynamics of Ions and Electrostatic Energy of Ferroelectrics—V. K. Kozlovskiĭ (pp. 352–358).

Theory of Orientational Ordering of Molecular Crystals—V. I. Klyachkin (pp. 359–367).

Theory of Hysteresis Phenomena in $BaTiO_3$—L. P. Kholodenko (pp. 368–373).

Relation between Dielectric Piezoelectric and Elastic Properties of Polycrystalline Ceramics and Single Crystals—S. V. Bogdanov, B. M. Vul, and A. M. Timonin (pp. 374–378).

Nonlinear Properties of Ferroelectrics—B. M. Vul (pp. 379–381).

Behaviour of Certain Ferroelectrics in Strong Electric Fields—V. A. Bokov (pp. 382–389).

Piezoelectric Properties of a Ferroelectric Ceramic of Barium Titanate with Certain Impurities—S. V. Bogdanov (pp. 390–393).

Reserve Piezoelectric Effect of Polycrystalline $BaTiO_3$ in Static-Type Measurements—G. M. Kovalenko (pp. 394–396).

Method of Measuring the Piezoelectric Coefficient α_{31} using the Radial Vibrations of a Disk—S. V. Bogdanov and A. M. Tumonin (pp. 397–398.

Pyroelectric Effect and Piezoelectric Effect in Polycrystalline Barium Titanate—M. S. Kosman and Z. A. Shamro (pp. 399–401).

Permittivity of Niobates and Tantalate of Divalent Metals—V. A. Isupov (pp. 402–410).

Investigation of Antiferroelectric Properties of Certain Solid Solutions containing Lead Zirconate—N. N. Krainik (pp. 411–422).

Study of Solid Solutions (Ba, Pb)(Ti, Sn)O_3 possessing Ferroelectric Properties—I. E. Myl'nikova (pp. 423–432).

Further Information on the $PbTiO_3$-$SrTiO_3$ Solid-Solution System—A. G. Boganov and R. A. Khomutetskaya (pp. 433–438).

Dependence of Permittivity of Polycrystalline Ferroelectrics on the Duration Application of a Mechanical Load and a Constant Electric Field—M. S. Lur'e (pp. 439–443).

Study of Ultrasonic Radiators made from Crystals of Rochelle Salt—K. A. Minaeva (pp. 444–449).

Stabilized Piezoelectric Ceramic Materials—R. E. Pasynkov and V. V. Vinogradov (pp. 450–454).

Application of Ferroelectrics to Frequency Multipliers—D. M. Kazarnovskiĭ and V. P. Sidorenko (pp. 455–465).

Apparatus for Exploratory Investigations in Ferroelectric Regions in Small Samples—I. S. Rez (pp. 466–472).

537.227:546.431.824–31 1443
The Effect of an Electric Field on the Transitions of Barium Titanate—M. E. Drougard and E. J. Huibregtse. (*IBM J. Res. Dev.*, vol. 1, pp. 318–329; October, 1957.) A review is presented of the effects of electric fields on the ferroelectric phase transitions of $BaTiO_3$ at 120°C and 5°C. The double hysteresis loop observed at the Curie point and the triple hysteresis loop and dielectric constant measured at the 5°C transition are examined in the light of Devonshire's thermodynamic theory of ferroelectricity in $BaTiO_3$. The data and various published experimental results are shown to agree with calculations based on the Devonshire free-energy function. The discrepancy between the coercive fields predicted by the theory and those actually observed is discussed.

537.311.3 1444
Screening of Coulomb Fields by Free Charges in Metals and Semiconductors—I. Z. Fisher. (*Zh. Tekh. Fiz.*, vol. 27, pp. 638–650; April, 1957.)

537.311.33 1445
On the Theory of the Thermal Capture of Electrons in Semiconductors—G. Richayzen. (*Proc. Roy. Soc. A*, vol. 241, pp. 480–494; September 10, 1957.) A new formula is obtained for the thermal capture rate in semiconductors when E_1/k is very much greater than the Debye temperature, E_1 being the ionizing potential of the bound electrons. It is found that the ratio of the probability of the impurity being ionized to that of a free electron being captured is $\exp(-E_1/kT)$. Applying this theory to a model of an impurity in a continuum, the capture rate is strongly dependent on the radius of the bound electron, the electron-lattice coupling, and the temperature; which agrees qualitatively with experiment.

537.311.33 1446
Effect of Neutral Impurities on Mobility in Nondegenerate Semiconductors—M. S. Sodha and P. C. Eastman. (*Phys. Rev.*, vol. 108, pp. 1373–1375; December 10, 1957.) The theoretical variation of mobility with electric field is studied, taking into account scattering by lattice vibrations and by both ionized and neutral impurities.

537.311.33 1447
Mobility of Carriers in Nondegenerate Semiconductors at Low Electric Fields—M. S. Sodha. (*Phys. Rev.*, vol. 108, pp. 1375–1376; December 15, 1957). "By using the velocity distribution of carriers in the presence of an electric field E, due to Yamashita and Watanabe [see 3093 of 1956 (Yamashita)] it is shown that the mobility μ is given by $\mu=\mu_0(1+\beta E^2)$ for low fields."

537.311.33:535.34 1448
Intensity of Optical Absorption by Excitations—R. J. Elliott. (*Phys. Rev.*, vol. 108, pp. 1384–1389; December 15, 1957.) The intensity of optical absorption close to the edge in semiconductors is examined using band theory together with the effective-mass approximation for the excitons. The cases of both direct transitions, which occur when the band extrema on either side of the forbidden gap are at the same K, and indirect transitions involving phonons are considered. The experimental results in Cu_2O and Ge are in good qualitative agreement with direct forbidden and indirect transitions, respectively.

537.311.33:535.34-1 1449
Oscillatory Magneto-absorption in Semiconductors—S. Zwerdling, B. Lax, and L. M. Roth. (*Phys. Rev.*, vol. 108, pp. 1402–1408; December 15, 1957.) Infrared magneto-absorption has been investigated in thin samples ($\sim 10\mu$) of Ge, InAs, and InSb in magnetic fields up to 27 kG. Accurate determinations of the energy gaps and effective masses have been made. The anisotropy of the magneto-absorption effect was measured for Ge and InSb.

537.311.33:[546.28+546.289 1450
Drift Mobility Measurements—M. Green. (*J. Appl. Phys.*, vol. 28, pp. 1473–1478; December, 1957.) Measurements on Ge and Si have been made by observing the transit time of a pulse of carriers over the distance between the point of photo-injection and a point of detection. The detector point may be a reverse-biased junction, a constriction in the cross-sectional area, or a point-contact electrode.

537.311.33:[546.28+546.289 1451
The Direct Observations of Dislocations in Germanium and Silicon—G. A. Geach, B. A. Irving, and R. Phillips. (*Research, London*, vol. 10, pp. 411–412; October, 1957.)

537.311.33:[546.28+546.289 1452
Optical Measurement of Film Growth on Silicon and Germanium Surfaces in Room Air—R. J. Archer. (*J. Electrochem. Soc.*, vol. 104, pp. 619–622; October, 1957.) "The thickness and growth kinetics of oxide films on polished Si and Ge exposed to room air after having been rinsed in hydrofluoric acid were obtained by measuring the ellipticity of reflected polarized light. Film growth obeys the Elovich equation."

537.311.33:546.28 1453
Effect of Oxygen on Etch-Pit Formation in Silicon—R. A. Logan and A. J. Peters. (*J. Appl. Phys.*, vol. 28, pp. 1419–1423; December, 1957.) The rate of etching is smaller the larger is the concentration of dissolved O_2 in the crystal. Etch-pit formation is correspondingly impeded. Observations have been made on virgin and heat-treated crystals. See also 3906 of 1957.

537.311.33:546.28 1454
Lifetime in Pulled Silicon Crystals—C. A. Bittmann and G. Bemski. (*J. Appl. Phys.*, vol. 28, pp. 1423–1426; December, 1957.) "Lifetime data of 46 pulled Si crystals are interpreted in terms of the Shockley-Read recombination theory. The data are consistent with a single recombination level and a constant concentration of recombination centers independent of the resistivity of the crystals."

537.311.33:546.28 1455
Effect of Heat Treatment upon the Electrical Properties of Silicon Crystals—C. S. Fuller and R. A. Logan. (*J. Appl. Phys.*, vol. 28, pp. 1427–1436; December, 1957.) Studies have been made of the process by which donors are introduced into Si by heating in the temperature range 300–500°C and are caused to disappear on heating at higher temperature. It is shown that O_2 is the impurity from which the donors are formed. See also 2808 of 1957 (Kaiser).

537.311.33:546.28 1456
Double-Acceptor Behavior of Zinc in Silicon—R. O. Carlson. (*Phys. Rev.*, vol. 108, pp. 1390–1393; December 15, 1957.) Evidence is presented to show that, under proper doping conditions, Zn acts as a double-acceptor impurity with levels at 0.31 ev from the valence band and 0.55 ev from the conduction band.

537.311.33:546.28 1457
Properties of Silicon Doped with Iron or Copper—C. B. Collins and R. O. Carlson. (*Phys. Rev.*, vol. 108, pp. 1409–1414; December 15, 1957.) It was found that Fe introduced a donor level into Si at 0.40 ev from the valence band. The electrically active solubility of Fe was 1.5×10^{16} cm^{-3} at 1200°C; the distribution coefficient was 8×10^{-6}. Studies of lifetime indicated a larger capture cross section for electrons than for holes. Cu introduced a donor level at 0.24 ev and an acceptor level at 0.49 ev, both as measured from the valence band. The maximum electrical activity in quenched samples was 5×10^{14} cm^{-3} out of a total concentration of 10^{18} cm^{-3} at 1200°C.

537.311.33:546.28 1458
Conductivity Mobilities of Electrons and Holes in Heavily Doped Silicon—G. Backenstoss. (*Phys. Rev.*, vol. 108, pp. 1416–1419; December 15, 1957.) The samples used had impurity concentrations up to 6×10^{19} cm^{-3} and 6×10^{18} cm^{-3} for n- and p-type Si, respectively. The conductivity mobilities were calculated by considering the percentage of ionized impurities. A comparison with the existing theory of impurity scattering yielded better agreement for n-type than for p-type Si.

537.311.33:546.28 1459
Thermal Generation of Recombination Centres in Silicon—B. Ross and J. R. Madigan. (*Phys. Rev.*, vol. 108, pp. 1428–1433; December 15, 1957.) The measurement of minority-carrier lifetime vs bulk resistivity in diffused silicon p-n junctions showed that Hall-Shockley-Read statistics are obeyed. A trap level of approximately 0.1 ev above the valence band in both conductivity types of silicon was deduced from lifetime vs temperature data on junctions which were annealed, quenched, and reannealed.

537.311.33:546.28:537.533 1460
Field Emission from Silicon—L. A. D'Asaro. (*J. Appl. Phys.*, vol. 29, pp. 33–34; January, 1958.) A description of field-emission patterns obtained using a single crystal of p-type Si. Stable patterns were obtained only after the point was distorted by heating in the presence of an electric field.

537.311.33:546.28:537.533 1461
Electron Emission from Avalanche Breakdown in Silicon—J. A. Burton. (*Phys. Rev.*, vol. 108, pp. 1342–1343; December 1, 1957.) A A new kind of electron emission has been observed, which arises from the energetic electrons produced in avalanche breakdown in a Si p-n junction at room temperature, when the Si work function is lowered by adsorbed Cs.

537.311.33:[546.28+546.289]:537.534.9 1462
Positive-Ion Bombardment of Germanium and Silicon—S. P. Wolsky. (*Phys. Rev.*, vol. 108, pp. 1131–1136; December 1, 1957.) A very sensitive vacuum microbalance has been used to study the sputtering of Ge and Si by argon ion bombardment. Current densities of 1 to 12 μa/cm^2 were used.

537.311.33:546.289 1463
Fine Structure in the Absorption-Edge Spectrum of Ge—G. G. Macfarlane, T. P. McLean, J. E. Quarrington, and V. Roberts. (*Phys. Rev.*, vol. 108, pp. 1377–1383; December 15, 1957.) Measurements of the absorption spectrum of Ge, made with high resolution, near the main absorption edge, at various temperatures between 4.2°K and 291°K, have revealed a fine structure on the long-wavelength side of this edge. This structure has been analyzed and can be interpreted in terms of indirect transitions involving phonons with energies corresponding to temperatures of 90°K and 320°K. The initial energy dependence of the components of the absorption coefficient associated with the 320°K phonons is interpreted as being due to the formation of excitons with a binding energy of 0.005 ev.

537.311.33:546.289:538.22 1464
Surface Paramagnetism of Germanium Films—V. L. Sandler. (*Phys. Rev.*, vol. 108, p. 1642; December 15, 1957.) Experiments suggest that a clean Ge surface is paramagnetic and that there is roughly one unpaired electron per Ge surface atom.

537.311.33:546.682.86 1465
Width of the Forbidden Band in InSb—V. V. Galavanov. (*Zh. Tekh. Fiz.*, vol. 27, pp. 651–655; April, 1957.) The relation is considered of temperature to conductivity and Hall constant in semiconductors with a large electron/hole mobility ratio. At $T=0$°K the width of the forbidden band in p-type samples appears nearly twice as large as that of n-type specimens. Tables and graphs of results are given.

537.311.33:546.682.86 1466
Indirect Transitions at the Centre of the Brillouin Zone with Application to InSb, and a Possible New Effect—W. P. Dumke. (*Phys. Rev.*, vol. 108, pp. 1419–1425; December 15, 1957.) The theory of indirect optical transitions is extended to the case where both the valence band and the conduction band extrema occur at the center of the Brillouin zone. The experi-

mental evidence on InSb is reviewed and found to be consistent with degenerate valence bands at the center of the Brillouin zone. A new effect is predicted involving the modulation of the indirect absorption constant by the selective excitation of the long-wavelength optical modes.

537.311.33:546.682.86:535.215 1467
Lifetime of Excess Carriers in InSb—P. T. Landsberg. (*Proc. Phys. Soc.*, vol. 70, pp. 1175–1176; December 1, 1957.) A discussion of the results of measurements giving a value of the ratio of bulk to radiative lifetime at room temperature. See also 2825 of 1957 (Moss).

537.311.33:546.682.86:538.63 1468
Galvanomagnetic Effects in *n*-Type Indium Antimonide—H. P. R. Frederikse and W. R. Hosler. (*Phys. Rev.*, vol. 108, pp. 1136–1145; December 1, 1957.) The magnetic-field dependence of the magnetoresistive effects and the Hall coefficient have been investigated at 78°K and at liquid-He temperatures. Results at very low magnetic-field strength are in agreement with the assumption of an isotropic conduction band. Quantization of the electron orbits causes deviations from the conventional theory at large field strengths. Oscillations in the magnetoresistance observed at 4.2°K and lower are attributed to these quantum effects. An effective mass value of 0.01 m_0 is obtained from the field and temperature dependence of the amplitude of the oscillations. The magnitude of the magnetoresistive effects appears to depend considerably on the geometry and inhomogeneity of the sample.

537.311.33:546.682.86:538.63 1469
Galvanomagnetic Effects in *p*-Type Indium Antimonide—H. P. R. Frederikse and W. R. Hosler. (*Phys. Rev.*, vol. 108, pp. 1146–1151; December 1, 1957.) The conductivity and Hall effect have been measured as a function of magnetic field strength and of temperature. Hall coefficient and magnetoresistance characteristics at 78°K are consistent with a valance band which consists of two bands with different effective masses and has degenerate maxima at the center of the Brillouin zone. Negative magnetoresistance has been observed at liquid-He temperatures; the effect is positive for very pure samples when the magnetic-field strength exceeds $\sim 10^3$ G.

537.311.33:546.873.241 1470
Optical and Electrical Investigation on Bismuth Telluride Bi_2Te_3—J. Lagrenaudie. (*J. Phys. Radium*, vol. 18, Supplement to no. 3 *Phys. Appl.*, pp. 39A–40A; March, 1957.)

537.311.33:546.873.241:537.32 1471
Electrical and Thermal Properties of Bi_2Te_3—C. B. Satterthwaite and R. W. Ure, Jr. (*Phys. Rev.*, vol. 108, pp. 1164–1170; December 1, 1957.) The phase diagram for Bi-Te in the region Bi_2Te_3 has been investigated. The Hall mobilities parallel to cleavage planes vary as $T^{-1.5}$ for holes and $T^{-2.7}$ for electrons. Room temperature values are $\mu_p = 420$, $\mu_n = 270$ cm²/v sec. The energy gap is 0.2 ev. The lattice conductivity is $5.10 \times 10^{-3}/T$ W/deg cm.

537.311.33:621.317.733 1472
Experimental Determination of the Mean Lifetime of Minority Carriers in Semiconductors—V. Andresciani and C. Della Pergola. (*Ricerca Sci.*, vol. 27, pp. 2663–2673; September, 1957.) A light-spot method using a Wheatstone bridge circuit is described. Results are compared with those obtained by means of a Many bridge (2129 of 1954) and are found to be adequate for practical purposes.

537.311.33:621.357.8 1473
Control of Optimum Conditions for Electrolytic Polishing of Semiconductors—I. Epelboin and M. Froment. (*J. Phys. Radium*, vol. 18, Supplement to no. 3, *Phys. Appl.*, pp. 60A–61A; March, 1957.) Polished Ge and Si semiconductors are examined by an interference contrast method using two polarized waves. Four micrographs are shown obtained with Ge samples polished in a mixture of 500 cm³ glycerine, 50 cm³ water, 50 cm³ ethyl alcohol.

537.533.8 1474
Effect of Reflected Electrons in Secondary Electron Emission—L. N. Dobretsov and T. L. Matskevich. (*Zh. Tekh. Fiz.*, vol. 27, pp. 734–744; April, 1957.) Specimens of mica, LiF, SiO_2, Ca_2CO_3, *etc.* were tested by means of a single-pulse circuit. The coefficients of secondary emission are tabulated and graphs shown.

538.22 1475
Theory of Antiferromagnetic-Ferromagnetic Transitions in Dilute Magnetic Alloys and in the Rare Earths—G. W. Pratt, Jr. (*Phys. Rev.*, vol. 108, pp. 1233–1242; December 1, 1957.) Both the molecular-field theory and a cluster theory due to Oguchi. (*Prog. Theor. Phys.*, vol. 13, pp. 148–150; February, 1955.) are applied. Each leads to a magnetic transition and both predict that for all such transitions the ferromagnetic state must have a lower free energy at 0°K than the antiferromagnetic state. A comparison is made between the present theory and recent experimental results for Cu-Mn alloys.

538.221 1476
Some Magnetothermal Relations for Ferromagnetics—R. W. Teale and G. Rowlands. (*Proc. Phys. Soc.*, vol. 70, pp. 1123–1134; December 1, 1957.) An expression is derived which relates magnetic and thermal changes for reversible rotation of the intrinsic magnetization or reversible domain wall motion. Experiments are proposed to differentiate between the possible energy changes which accompany these processes.

538.221 1477
Criterion for Ferromagnetism from Observations of Magnetic Isotherms—A. Arrott. (*Phys. Rev.*, vol. 108, pp. 1394–1396; December 15, 1957.) "A criterion is proposed for determining the onset of ferromagnetism in a material as its temperature is lowered from a region in which the linearity of its magnetic moment vs field isotherm gives an indication of paramagnetism."

538.221 1478
Specific Heat of a Ferromagnetic Substance Above the Curie Point—C. Domb and M. F. Sykes. (*Phys. Rev.*, vol. 108, pp. 1415–1416; December 15, 1957.) "High-temperature expansions for the specific heat of the Ising model are compared for the triangular and fcc lattices. It is concluded that the 'tail' of the specific heat curve above the Curie point is much smaller for the fcc lattice, but is steeper in the immediate neighborhood of the Curie point."

538.221:621.318.124+621.318.134 1479
Convention on Ferrites—(*Proc. IEE*, Pt. B, vol. 104, Supplement no. 6, pp. 267–398; 1957.) The following papers were included among those read at the IEE Convention held in London October 29–November 2, 1956. See also 1202 of April.

Microwave Introductory Session

A Survey of the Theory and Applications of Ferrites at Microwave Frequencies—P. J. B. Clarricoats, A. G. Hayes, and A. F. Harvey pp. 267–282).

Discussion (pp. 283–285).

Microwave Theory and Measurements

Some Properties of Circular Waveguides containing Ferrites—P. J. B. Clarricoats (pp. 286–295).

The Measurement of Complex Permittivity and Complex Tensor Permeability of Ferrite Materials at Microwave Frequencies—I. G. MacBean (pp. 296–306).

Theory of the Measurement of the Elements of the Permeability Tensor of a Ferrite by means of a Resonant Cavity—R. A. Waldron (pp. 307–315).

Nonreciprocal Network Theory applied to Ferrite Microwave Devices—H. J. Carlin (pp. 316–319).

Discussion (pp. 320–323).

Microwave Measurements and Properties

Microwave Faraday Effect and Conductivity in Nickel Ferrites and Ferrite-Aluminates—R. Derry and M. S. Wills (pp. 324–330).

Some Properties and Applications of Ferrites at 3-cm Wavelength—S. Boronski (pp. 331–337).

Measurement of Ferrite Properties in a Rectangular Cavity at 10000 Mc/s—J. Roberts and C. M. Srivastava (pp. 338–341).

The Quantum Theory of Spontaneous Magnetization of Ferrites at Low Temperatures—E. I. Kondorskiĭ (p. 342). See also 3930 of 1957 (Kondorskiĭ *et al.*).

Discussion (pp. 343–345).

Microwave Apparatus

Ferrite Structures for Millimetre Wavelengths—A. F. Harvey (pp. 346–354).

Applications of Ferrites at 3-cm Wavelength—R. M. Godfrey, B. L. Humphreys, P. E. V. Allin, and G. Mott (pp. 355–361).

Development of a Rotation Isolator for 6-cm Wavelength—P. E. Ljung (pp. 362–363).

A Resonance Absorption Isolator in Microstrip for 4 Gc/s—L. Lewin (pp. 364–365).

Notes on Microwave Ferromagnetics Research—A. G. Fox (pp. 371–378).

Some Measurements and Applications of the Microwave Properties of a Magnesium-Manganese Ferrite in the 8–9-mm Waveband—E. Laverick and A. Rivett-Carnac (pp. 379–382).

The 45°-Faraday-Rotation Ferrite Isolator—A. L. Morris (pp. 383–387).

Some Applications of Ferrites to Microwave Directional Couplers, Switches, and Cavity Filters—E. Strumwasser (pp. 388–394).

Discussion (pp. 366–370, 395–398).

538.221:621.318.134 1480
Preparation of Polycrystalline Ferrimagnetic Garnet Materials for Microwave Applications—W. D. Wolf and G. P. Rodrigue. (*J. Appl. Phys.*, vol. 29, pp. 105–108; January, 1958.) A simple chemical method is described for preparing high-density, polycrystalline, magnetic garnets having the formula $3M_2O_3 \cdot 5Fe_2O_3$, where M is a rare earth from Sm to Lu, or yttrium. Results of measurements of ferrimagnetic resonance line width and dielectric loss tangent are given.

538.221:621.318.134 1481
Grandular Structures with Superficial Layer in Ceramics with Iron Oxide Base—J. Suchet. (*J. Phys. Radium*, vol. 18, Supplement to no. 3, *Phys. Appl.*, pp. 10A–18A; March, 1957.) In ceramics with Fe_2O_3 base particularly Mn-Zn ferrite there are two types of resistive surface layer formed by oxidation during the cooling process.

538.221:621.318.2 1482
Fine-Particle Magnets—T. O. Paine, L. I. Mendelsohn, and F. E. Luborsky. (*Elec. Eng., N. Y.*, vol. 76, pp. 851–857; October, 1957.) The basic factors which determine the behavior of permanent magnets made from fine particles are reviewed. Alignment and packing of single-domain particles are discussed and the properties of manganese-bismuth, barium

ferrite, and fine-particle iron as permanent-magnet materials are described. Thiry-six references.

538.221:621.385.833 1483
Study of the Magnetic Leakage Field in Cobalt by Means of a Divergent Electron Beam—M. Blackman and E. Grünbaum. (*Nature, London*, vol. 180, pp. 1189–1190; November 30, 1957.) An electron beam diverges from a small movable aperture of 6–25 μ in a direction parallel to the crystal edge and is deflected by the leakage field. Curves obtained at 20°–265°C are given. See 205 of 1957.

538.222:538.569.4 1484
Research on the Paramagnetism of Solids at the Institute of Physics of Palermo University—M. Santangelo. (*Ricerca Sci.*, vol. 27, pp. 2768–2772; September, 1957.) Microwave equipment for the measurement of magnetic resonance absorption is briefly described.

538.222:621.318.134 1485
Investigation of the Paramagnetism of the Ferrites $5Fe_2O_3,3M_2O_3$, where M = Gd, Dy, Er—R. Aléonard and J. C. Barbier. (*C.R. Acad. Sci., Paris*, vol. 245, pp. 831–834; August 19, 1957.) Investigation of the thermal variation of susceptibility above the Curie point of rare-earth ferrites of the garnet type.

621.315.615:537.528 1486
The Breakdown of Liquid Dielectrics and its Dependence on Oxidation of the Electrodes—R. Hancox. (*Brit. J. Appl. Phys.*, vol. 8, pp. 476–480; December, 1957.) Theoretical considerations and experiments with pre-treated transformer oil show that oxide layers on the electrodes increase the recorded breakdown strength of the dielectric.

MATHEMATICS

519.283 1487
Nonparametric Definition of the Representativeness of a Sample—with Tables—M. Sobel and M. J. Huyett. (*Bell. Sys. Tech. J.*, vol. 37, pp. 135–161; January, 1958.) The problem is to determine how large a random sample is needed to guarantee with pre-assigned probability that the sample will have a specified amount (or a specified degree) of representativeness of the true, unknown (cumulative) distribution F under study. The solution given is nonparametric (*i.e.*, distribution-free) so that the results obtained and the tables and graphs constructed are valid for any true underlying distribution.

MEASUREMENTS AND TEST GEAR

529.786+621.3.018.41(083.74) 1488
The "Time Centre" at the Milan Polytechnic—C. Mazzon. (*Ricerca Sci.*, vol. 27, pp. 2727–2747; September, 1957.) Time- and Frequency-standard equipment and control and recording apparatus are described. Details are given of a system for improving the accuracy of calibration on the basis of standard-frequency signal reception.

529.786:621.317.4 1489
A Small Quartz Clock with Transistor Drive—D. E. Cridlan and J. E. Thwaites. (*P.O. Elec. Eng. J.*, vol. 50, Pt 3, pp. 189–191; October, 1957.) A quartz tuning fork gives a frequency of 800 cps which is divided down in four identical binary stages to drive a 50-cps mains clock movement. Transistors are used in the oscillator drive circuit, the frequency dividers, and amplifier.

621.317.2:621.374.3 1490
A Versatile Pulse Pattern Generator—P. H. Cutler and L. R. Peters. (*Electronic Eng.* vol. 30, pp. 39–42; January, 1958.) "Two independent pulse pattern generators are used to produce any sequence of pulses of either polarity; either sequence contains up to ten pulses, and the pulse duration may be varied from 45 μs to 1 s. The two separate patterns may be combined to give a third pattern whose maximum length is ninety pulses."

621.317.3:621.314.7 1491
Measuring Transistor "Power Gain" at High Frequencies—W. N. Coffey. (*Electronic Ind. Tele-Tech.* vol. 16, pp. 66–68, 169; October 1957.) A method of measuring directly the common-emitter power gain of transistors in the 40–300-mc range when driven by a resistive generator.

621.317.3.029.63:621.396.822 1492
The Precise Measurement of Noise Temperatures of Two-Terminal Networks in the Decimetre Wavelength Range by a Substitution Method—L. Mollwo. (*Arch. elek. Übertragung*, vol. 11, pp. 295–306; July, 1957.) General formulas are derived for errors caused by impedance mismatch of network to measuring system. A substitution method for reducing the effect of these errors is described. Mismatch corrections are considered for the case of a saturated-diode noise generator.

621.317.329:621.316.8 1493
Practical Applications of a Resistance Network for Determining Plane Potential or Space-Charge Fields—G. Čremočsnik and M. J. O. Strutt. (*Arch. Electrotech.*, vol. 43, pp. 177–186; May 2, 1957.) The reduction of errors in using a resistance-network analog is discussed on the basis of theoretical investigations. See, *e.g.*, 974 of 1958.

621.317.331:621.316.721.078.3 1494
A Current Regulator to Facilitate Resistance Measurements at Low Temperature—M. W. Thompson. (*J. Sci. Instr.*, vol. 34, p. 515; December, 1957.)

621.317.341.3(083.57) 1495
A Chart for Return Loss Determination—L. Kitajewski. (*Electronic Eng.*, vol. 30, pp. 42–43; January, 1958.) Curves are shown from which the return loss or the reflection coefficient can be determined quickly from bridge measurements of impedances. The curves are applicable only to the case when the return loss is required between an unknown impedance and a characteristic impedance of $\angle 125\ 0°\Omega$.

621.317.343.2:621.314.22 1496
Measurement of the Characteristic Impedance of a Coaxial Cable—L. B. D'Alton. (*Electronic Eng.*, vol. 30, pp. 37–38; January, 1958.) " A method is described for measuring the characteristic impedance of a coaxial cable and the velocity of propagation, the only equipment required being a radio receiver, a signal generator and a calibrated capacitor."

621.317.361.018.756 1497
Errors in the Measurement of Pulse Frequency Spectra by means of Stagger-Tuned Resonant Circuits—G. Seeger and H. G. Stäblein. (*Arch. elek. Übertragung*, vol. 11, pp. 325–330; August, 1957.) The type and magnitude of the errors due to the losses in the resonant circuits are investigated for circuits of equal absolute and equal relative bandwidth. For a given permissible error, the maximum frequency spectrum is covered by circuits suitably staggered with regard to bandwidth.

621.317.39:531.771:621.387 1498
A Very-High-Speed Precision Tachometer—J. K. Goodwin. (*Electronic Eng.*, vol. 30, pp. 18–24; January, 1958.) Describes fully a portable instrument that counts up to 40,000 rev/min. It may also be used to count pulses up to 20,000 ps with an accuracy of +1 ps, or random pulses to a total of 999,999. See 1153 of 1954 (Bland and Cooper).

621.317.39:531.78 1499
Strain-Gauge Oscillator for Flight Testing—W. H. Foster. (*Electronics*, vol. 31, pp. 40–42; January 31, 1958.) The frequency of a small transistor oscillator is modulated in proportion to the stress or pressure applied to a strain gauge incorporated in a feedback loop.

621.317.39:536.53:621.316.825 1500
Temperature Measurement with Thermistors—J. C. Anderson. (*Electronic Radio Eng.*, vol. 35, pp. 80–84; March, 1958.) "Characteristics of thermistors are discussed, and three temperature-measuring devices described: an industrial thermometer for 0°–100°C using a Wheatstone bridge, a medical thermometer covering 85°–105°F incorporating a balanced transistor amplifier and a high-sensitivity device using a two-stage transistor amplifier. Some observations on thermistor stability are included."

621.317.39:537.312.9 1501
Use of Piezoresistive Materials in the Measurement of Displacement, Force and Torque—W. P. Mason and R. N. Thurston. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1096–1101; October, 1957.) "The use of piezoresistive materials as strain gauges and in the measurement of displacement, force, and torque is discussed generally. A torsional transducer which has been constructed from *n*-type Ge is described, and the experimentally obtained voltage/torque characteristic is given."

621.317.4:621.317.755 1502
***B-H* Tester measures Memory Core Parameters**—T. H. Bonn, R. D. Torrey, and F. Bernstein. (*Electronics*, vol. 31, pp. 76–80; January 17, 1958.) This instrument measures the hysteresis properties of small magnetic toroids. The flux, drive current, ratio of remanent flux to maximum flux, squareness ratio, and ratio of coercive force to maximum magnetizing force are read from window potentiometers as the *B-H* loop is presented on a cathode-ray tube.

621.317.4:621.372.413 1503
A Re-entrant Cavity for Magnetic Measurements—E. A. Faulkner. (*J. Sci. Instr.*, vol. 34, pp. 514–515; December, 1957.) An open-ended resonant cavity is used to measure magnetic permeability of small toroidal specimens in the range 250–3300 mc.

621.317.431 1504
New Technique for Measuring Rotational Hysteresis in Ferromagnetic Materials—J. M. Kelly, Jr. (*Rev. Sci. Instr.*, vol. 28, pp. 1038–1040; December, 1957.) A technique for use in the range 3–30 cps using a sample disk of material mounted on a top rotating in a magnetic field.

621.317.723:621.375.024 1505
A Feedback Electrometer Amplifier—D. Allenden. (*Electronic Eng.*, vol. 30, pp. 31–33; January, 1958.) The inherent feedback loop gain is used to stabilize the electrometer-tube filament supply. The zero drift at full-scale voltage sensitivity of 2 v is less than ± 1 part in 1000 over several hours.

621.317.733 1506
The Accuracy of Determination of Capacitance and Loss Factor by means of A.C. Bridges—H. Hoyer and W. Wiessner. (*Arch. Elektrotech.*, vol. 43, pp. 169–177; May 20, 1957.) Simple formulas are derived for determining the limits of accuracy and sensitivity of conventional capacitance bridges, taking account of matching between bridge and indicating device.

621.317.74:621.385.029.64 1507
S.H.F. Frequency Sweeper uses Backward-

Wave Tube—D. E. Wheeler and P. D. Lacy. (*Electronics*, vol. 31, pp. 76–78; January 3, 1958.) "Swept-frequency signal source using backward-wave oscillator tube offers sweep rates from 40 mc to 400 kmc in the microwave region between 8.2 and 12.4 kmc. Rapid wide-range evaluations of reflection, gain, and attenuation are possible, as well as permanent records of measured data on an ink recorder. Sweep width is continually adjustable from 3 mc to 4.2 kmc and unit may be modulated with AM or FM."

621.317.75:621.317.715:534.1 1508

Response Spectra by means of Oscillograph Galvanometers—R. W. Conrad and I. Vigness. (*J. Acoust. Soc. Amer.*, vol. 29, pp. 1110–1115; October, 1957.) The principles of design and calibration of a response or shock-spectrum analyzer are described. It comprises twelve galvanometer elements with natural frequencies in the range 10–2500 cps, each with an associated amplifier and damping network. Damping is adjustable between about 3 and 50 per cent of critical. An example is given of its operation in the analysis of tape-recorded accelerometer signals played back at different speeds.

621.317.755.001.6(091) 1509

The Development of the Cathode-Ray Oscillograph since 1923—P. Hochhäusler. (*Elektrotech. Z., Edn A*, vol. 78, pp. 514–521; August 1, 1957.) A survey with 36 references, mainly to German literature.

621.317.755.087.5 1510

Continuous Recording of Waveforms on Photographic Film—V. B. Hulme. (*Electronic Eng.*, vol. 30, pp. 10–14; January, 1958.) The limitations on the information capacity of the deflected-spot cathode-ray tube waveform recorder are overcome by using a ribbon-beam tube and variable-area recording.

621.317.79:538.632:537.311.33 1511

A Simple Specimen Holder and Apparatus for Measurement of Conductivity and Hall Voltage over a Temperature Range—A. A. Brooker, R. A. Clay, and A. S. Young. (*J. Sci. Instr.*, vol. 34, pp. 512–513; December, 1957.)

OTHER APPLICATIONS OF RADIO AND ELECTRONICS

537.533.35(091) 1512

25 Years of Electron Microscopy—E. Ruska. (*Electrotech. Z., Edn A*, vol. 78, pp. 531–543; August, 1957.) An outline of the scientific background and the history of the technique including the latest developments. Ninety-four references.

538.74:538.632 1513

The Hall-Effect Compass—I. M. Ross, E. W. Saker, and N. A. C. Thompson. (*J. Sci. Instr.*, vol. 34, pp. 479–484; December, 1957.) A description of the use of the Hall-effect voltage from a semiconductor in the earth's magnetic field as the basis for an electrical compass. Expressions are derived for the voltage and power outputs, and the sensitivity and the relative merits of various semiconductors are reviewed. A practical design of compass is described and illustrated.

551.508.8:621.316.825 1514

Response of Radiosonde Thermistors—F. I. Badgley. (*Rev. Sci. Instr.*, vol. 28, pp. 1079–1084; December, 1957.) The thermistors are estimated to have lag coefficients of about 25 s (10× the ground level value) at heights of about 70,000 feet. The radiation error at these heights could be several degrees.

621.317.755:003.35 1515

Generating Characters for Cathode-Ray Read-Out—K. E. Perry and E. J. Aho. (*Electronics*, vol. 31, pp. 72–75; January 3, 1958.) The necessary X and Y deflection voltages for representation of the Arabic numerals 0 to 7 on a cathode-ray tube screen are obtained by Fourier synthesis technique involving the combination of sine and cosine terms of the first five harmonics from a 30-kc fundamental frequency. Each character is traced in about 30 μs.

621.384.6 1516

Fields in Gap-Excited Rectangular Ducts—J. Van Bladel. (*J. Appl. Phys.*, vol. 28, pp. 1479–1483; December, 1957.) "The em field in a rectangular waveguide, cut in two by a plane perpendicular to its longitudinal axis, is investigated. A voltage applied across the two halves of the guide generates the field structure, which is analyzed for several values of the frequency (between zero and cutoff) and of the aspect ratio of the cross section. The problem is of interest for the design of particle accelerators."

621.384.611 1517

A New Particle Accelerator—Y. P. Varshni. (*Indian J. Phys.*, vol. 31, pp. 384–386; July, 1957.) In a cyclotron of conventional form it is not possible to accelerate ions beyond a certain limit on account of their relativistic increase in mass at high velocities and consequent departure from resonance. This difficulty may be overcome using a cyclotron with spiral-shaped "dees" called a spiratron.

621.385.833.001.6(091) 1518

The History of the Development of the Electron Microscope—D. Gabor. (*Elektrotech. Z. Edn A*, vol. 78, pp. 522–530; August 1, 1958.) An illustrated review with eighty-three references.

621.398 1519

Telecontrol—E. L. Gruenberg. (IRE TRANS. ON TELEMETRY AND REMOTE CONTROL, vol. TRC-3, pp. 5–8; May, 1957. Abstract, PROC. IRE, vol. 45, pp. 1168–1169; August, 1957.)

621.398:061.3 1520

Proceedings of the 1957 National Symposium on Telemetering—(IRE TRANS. ON TELEMETRY AND REMOTE CONTROL, vol. TRC-3, April, 1957.) The text is given of 36 papers, with abstracts of seven others, presented at a symposium held at Philadelphia, Pa, April 14th–16th. Abstracts of most of the papers are given in PROC. IRE, vol. 45, pp. 1166–1168; August, 1957. The following papers were included.

A Wide-Band Microwave Link for Telemetering—R. E. Glass (Sec. 1.2, 13 pp.).

The Transmission of Pulse Width Modulated Signals over Restricted Bandwidth Systems—H. J. Heffernan (Sec. 2.3, 4 pp.).

Extension of F.M./F.M. Capabilities—H. O. Jeske (Sec. 2.4, 10 pp.).

Telemetering System for the X-17 Missile—J. A. Cox (Sec. 2.5, 14 pp.).

Transistor Circuits Applied to Telemetering—J. H. Smith (Sec. 3.1 11 pp.).

Low-Level Transistorized Chopper Amplifier—H. F. Harris and T. E. Smith (Sec. 3.5, 8 pp.).

Progress Report on a Solid-State F.M./F.M. Telemetering System—E. Y. Politi (Sec. 3.6, 19 pp.).

A Ruggedized R.F. Power Amplifier for Use in the 200-Mc/s Telemetry Band—D. D. McRae (Sec. 6.1, 4 pp.).

A New Transistor Magnetic F.M./F.M. Subcarrier Discriminator—G. H. Barnes and R. M. Tillman (Sec. 6.3, 9 pp.).

A Transistorized Pulse-Width Keyer—J. A. Riedel, Jr, (Sec. 6.5, 6 pp.).

P.D.M. Bandwidth Requirements—F. E. Rock (Sec. 7.2, 6 pp.).

Noise and Bandwidth in P.D.M./F.M. Radio Telemetering—K. M. Uglow (Sec. 7.5, 10 pp.).

Design of All-Channel Ultra-stable F.M. Discriminator—S. Rigby (Sec. 8.2, 8 pp.).

P.D.M./P.A.M. Conversion System—R. L. Kuehn and W. L. Johnston (Sec. 8.5, 11 pp.).

621.398:621.376.3 1521

A Note on the Frequency Distribution of an F.M./F.M. Signal—P. B. Arnstein. (IRE TRANS. ON TELEMETRY AND REMOTE CONTROL, vol. TRC-3, pp. 13–16; May, 1957. Abstract, PROC. IRE, vol. 45, p. 1169; August, 1957.)

621.398:621.376.3 1522

Noise and Bandwidth in F.M./F.M. Radio Telemetering—K. M. Uglow. (IRE TRANS. ON TELEMETRY AND REMOTE CONTROL, vol. TRC-3, pp. 19–22; May, 1957. Abstract, PROC. IRE, vol. 45, p. 1169; August, 1957.)

621.398:621.376.56:621.314.7 1523

Transistorized Time Multiplexer for Telemetering—J. M. Sacks and E. R. Hill. (IRE TRANS. ON TELEMETRY AND REMOTE CONTROL, vol. TRC-3, pp. 26–30; May, 1957. Abstract, PROC. IRE, vol. 45, p. 1169; August, 1957.)

621.398:621.396.934:621.317.7 1524

Special Timing Techniques Employed on Guided-Missile Ranges—R. J. Garvey. (*Electronic Eng.*, vol. 30, pp. 2–9; January, 1958.) The timing and reference signals are derived from a 10-kc crystal oscillator driving a series of frequency-dividing circuits, and are transmitted to the measuring stations by underground telephone cables. One of the timing signals indicates time intervals from an arbitrary zero.

621.398:629.19 1525

Transistorized Memory monitors Earth Satellite—C. S. Warren, W. G. Rumble, and W. A. Helbig. (*Electronics*, vol. 31, pp. 66–70; January 7, 1958.) In the reception of telemetered data the circuits required to translate input information numerically and to present modified output information, use alloy-junction transistors as current drivers, gated-pulse amplifiers, voltage amplifiers, high-speed switches, and flip flops. Storage capacity is 6400 bits arranged as 256 characters of 25 bits each.

655.3.024:621.385.832 1526

The Scanatron: an Electronic Instrument for Preparing Colour Separations for Multicolour Printing—P. E. Klein. (*Elektronik*, vol. 6, pp. 238–239; August, 1957.) The equipment described comprises a single cathode-ray beam scanning tube and electronic circuitry for the production of corrected color-separation plates. The resolution is 200 lines per cm. See also 1563 of 1957 (Shapiro and Haynes).

PROPAGATION OF WAVES

621.396.11 1527

Spectrum of Turbulent Mixing—R. Bolgiano, Jr. (*Phys. Rev.*, vol. 100, p. 1348; December 1, 1957.) It is argued that the conclusions presented in a recent paper [2881 of 1957 (Wheelon)], that the spectral dependence of molecular fluctuations in the initial subrange should and do obey a k^{-3} law are incorrect. A $k^{-6/3}$ law is more appropriate.

621.396.11:551.510.535 1528

Some Problems of the Physics of the Ionosphere. Part 1—Fluctuation of the Electron Density and Scattering of Radio Waves—Al'pert. (See 1425.)

621.396.11:551.510.510.535:523.746 1529

The Sunspot Cycle and Radio Communications—G. Millington. (*Point to Point Telecommun.*, vol. 1, pp. 7–21; June, 1957.) The

characteristics of the ionosphere during a sunspot cycle and the consequent effects on HF radio communication circuits are discussed with particular attention to the problem of interference, at high sunspot numbers, to circuits operating in the lower end of the VHF band and to forward-scatter transmissions. Sunspot cycles since 1749 are examined but it is concluded that any precise predictions of future trends resulting from analysis of past behavior should be treated with great caution in planning future frequency allocations for radio communication.

621.396.11.029.4:551.594.6 1530

The Waveforms of Atmospherics and the Propagation of Very-Low-Frequency Radio Waves—J. Chapman. (*J. Atmos. Terr. Phys.*, vol. 11, pp. 223–236; 1957.) An analysis of waveform and frequency spectra observations, and comparison with theory in terms of "waveguide" propagation between earth and ionosphere. A new classification of atmospheric waveforms is used, which relates to frequency spectrum, day or night conditions, and source distance. Appreciable differences are indicated between the spectra of individual flashes at the source.

621.396.11.029.6:621.397.26 1531

Scatter Propagation and its Application to Television—J. A. Saxton. (*J. Telev. Soc.*, vol. 8, pp. 273–284; July/September, 1957.) The fundamental processes common to both ionospheric and tropospheric scatter propagation and the application of the mechanisms to point-to-point links are described. The dependence of performance on frequency, bandwidth, and antenna characteristics is given and it is shown that whereas VHF ionospheric scatter links are useful primarily for telegraphy over paths of 1,000–2,000 km, UHF tropospheric scatter links may possibly be used for relaying television over paths of a few hundred km. Thirty-three references.

621.396.11.029.62/.64 1532

Scatter Propagation—(*Wireless World*, vol. 64, pp. 124–125; March, 1958.) A digest of 23 papers on ionospheric and tropospheric forward-scatter propagation delivered at an IEE symposium held on January 28, 1958.

621.396.11.029.62:551.510.535 1533

Radio-Frequency and Scattering-Angle Dependence of Ionospheric Scatter Propagation at V.H.F.—R. Bolgiano, Jr. (*J. Geophys. Res.*, vol. 62, pp. 639–640; December, 1957.) Comment on 2576 of 1957 (Wheelon). The conclusion that the spectrum of mean square fluctuations produced by turbulent mixing of an established gradient is proportional, in the inertial subrange, to k^{-3} $\left(\text{where } k=\dfrac{4\pi}{\lambda \sin\theta/2}\right)$ is shown to be based on an inconsistent argument.

RECEPTION

621.376.233:621.314.7 1534

Transistor Regenerative Detectors—I. Gottlieb. (*QST*, vol. 41, pp. 30–31; October, 1957.) Details are given of the design and construction of a two-stage transistor receiver for the 80-m band. Its over-all performance is claimed to be better than that of an equivalent tube set.

621.376.33:621.396.82 1535

F.M. Demodulator Time-Constant Requirements for Interference Rejection—E. J. Baghdady. (PROC. IRE, vol. 46, pp. 432–440; February, 1958.) The upper bounds of permissible values of low-frequency time constants in the limiter and discriminator circuits operating under conditions of high-level interference are calculated. Severe restrictions are indicated which conflict with other fundamental requirements. Results show that these restrictions can be alleviated by simple schemes which enhance the capture performance of the FM receiver.

621.376.4:621.396.3 1536

Practical Elimination of "Reverse Operation" in an Amplitude-Phase Detector due to Pulse Interference—Y. S. Lezin. (*Radiotekhnika, Moscow*, vol. 11, pp. 45–52; November, 1956.) Change of polarity of the output voltage due to single-pulse or pulse-train interference can be eliminated by narrowing the pass band of the grid circuit of the detector. See 927 of 1957.

621.396.62:061.43(43) 1537

Novelties and Development Trends in the Construction of Receivers—C. Reuber. (*Elektrotech. Z., Edn B*, vol. 9, pp. 321–328; August 21, 1957.) Survey of the German Radio Show in Frankfurt with details of some of the circuit refinements and new ideas incorporated in the sound and television receivers exhibited there.

621.396.62:621.314.7 1538

Transistor Reflex Circuit trims Receiver Costs—E. Gottlieb. (*Electronics*, vol. 31, pp. 66–68; January 3, 1958.) These circuits, in which the second IF stage doubles as the AF amplifier, do not suffer from the disadvantages inherent in earlier tube versions. Further circuit simplifications may be effected by coupling an *n-p-n* reflex stage to a *p-n-p* output amplifier.

621.396.62:621.372.632 1539

Low-Noise Converter for I.G.Y. Propagation Study—L. F. Garrett. (*Electronics*, vol. 31, pp. 52–54; January 31, 1958.) A 400-mc converter, of 4-mc bandwidth, with gain of 36 db and noise figure of 2.5 db, is described in detail. Possible applications include scatter propagation and meteoric and auroral studies.

621.396.62.001.6 1540

Problems of Modern Radio-Receiver Development—W. Kausch. (*Elektrotech. Z., Edn B*, vol. 9, pp. 329–330; August 21, 1957.) Brief assessment of recent technical improvements and their effect on production techniques.

621.396.8:621.376.3 1541

Signal-to-Noise Ratios in Strong-Carrier F.M. Systems—H. Urkowitz. (*Commun. and Electronics*, pp. 599–602; November, 1957.) Ratios are derived without use of high-order probability density functions. General formulas are obtained by determining the signal-to-noise ratio with periodic modulation or with pulse modulation and numerical results are given for the special case of a flat symmetrical RF pass band. The effect of a mistuned frequency detector is taken into account.

621.396.828:681.142 1542

Radio-Interference Control as Applied to Business Machines—J. M. Sarley and R. J. Hendery. (*IBM J. Res. Dev.*, vol. 1, pp. 363–372; October, 1957.) The characteristics of the interference from machines is discussed. Some of the problems encountered in locating the various sources and of reducing the interference to a satisfactory level are described.

STATIONS AND COMMUNICATION SYSTEMS

621.391 1543

Double-Sampling Theorems in Continuous Signals—N. Honda. (*Sci. Rep. Res. Inst. Tohoku Univ., Ser. B*, vol. 8, pp. 23–34; June, 1956.) Sampling theorems are given for continuous signals whose frequency spectra are $(2k+1)$ bands in the frequency region $kf_0-f_m<f<kf_0+f_m$ $(k=-K, \cdots, 0, \cdots K$ and $f_0>2f_m$. It is proved that the signal is completely determined by $(2k+1)$ unique functions spaced $1/2f_m$ seconds apart. If $f_0/2f_m$ is an integer, the signal is completely determined by giving its $(2k+1)$ values spaced $1/(2k+1)$ f_0 seconds apart.

621.391 1544

The Sequential Error-Correcting Code—N. Honda. (*Sci. Rep. Res. Inst. Tohoku Univ., Ser. B*, vol. 8, pp. 113–124; December, 1956.) A practical sequential error-correcting code is proposed in which, if the number of errors is below a definite number, then the code may be corrected.

621.391 1545

Geometrical Proof of Shannon's Theorem—E. L. Blokh and A. A. Kharkevich. (*Radiotekhnika, Moscow*, vol. 11, pp. 5–16; November, 1956.) The attempt to provide a geometrical proof of Shannon's theorem made in two previous papers (259 of 1956 and *ibid.*, vol. 10, pp. 3–7; July, 1955.) led to results different from those of Shannon. It is shown that this discrepancy was due to the choice of an incorrect geometric model.

621.391:621.396.822 1546

Fluctuations of Random Noise Power—D. Slepian. (*Bell. Sys. Tech. J.*, vol. 37, pp. 163–184; January, 1958.) "The probability distribution of the power, y, of a sample of Gaussian noise of time duration T is considered. Some general theory is presented along with curves for the cumulative distribution and probability density of y for several different power spectra and values of T."

621.391.5.029.53:621.376.3 1547

Wireless Microphone uses F.M. Modulation—G. F. Montgomery. (*Electronics*, vol. 31, pp. 54–55; January 3, 1958.) A transistor circuit generating a low-power induction field at 460 kc. An FM system was chosen because of its inherent AVC action thus limiting the field strength during times of peak modulation.

621.396.1 1548

A Canadian Point of View on Radio-Frequency Spectrum Management—C. M. Brant. (*Commun. and Electronics*, pp. 455–461; September, 1957.) Future frequency allocations for the various services are discussed.

621.396.3:621.396.43:523.5 1549

V.H.F. Propagation by Ionized Meteor Trails—W. R. Vincent, R. T. Wolfram, B. M. Sifford, W. E. Jaye, and A. M. Peterson. (*Electronic. Ind Tele-Tech*, vol. 16, pp. 52–55, 146 and 84–88, 98; October and November, 1957. IRE WESCON CONVENTION RECORD, vol. 1, Pt. 1, pp. 263–282; 1957.) Data collected at a receiving station in California using signals from television transmitters gave information on the nature of the signals, their level, duration, fading and frequency of occurrence, and also on the optimum antenna direction. An experimental 820-mile link operating at 40 and 32 mc between Bozeman, Mont., and Palo Alto, Calif., is described with details of W/T and R/T equipment and the magnetic-tape storage systems.

621.396.4:621.315.212 1550

Trial Model of Super Wide-Band Repeater for Coaxial Cable—Y. Shigei. A. Ishii, G. Marubayashi, and T. Iwase. (*Rep. Elec. Commun. Lab., Japan*, vol. 5, pp. 1–3; August, 1957.) A capacity of 1) 2,700 telephone channels, or 2) one 4.3-mc television channel and 1,200 telephone channels, or 3) two 4.3-mc television channels, is aimed at. The characteristics are given of two new types of tube developed for this purpose. The system is briefly described.

621.396.41 1551
Design Factors for the Optimization of Multichannel Radio Systems—C. A. Parry. (*Commun. and Electronics*, pp. 606–620; November, 1957.) Satisfactory design needs evaluation of both nonlinear and thermal noise, optimum design being obtained with these noise powers equal. Essential performance factors required for this evaluation are considered. The effects of carrier power, channel spacing, speech-power distribution, and traffic activity on the signal-to-noise ratio are examined. Total reliability as a function of traffic fluctuations, propagation reliability, and carrier power is established. Theoretically optimum distributions as well as typical operating distributions are given. Standards are suggested both for systems which must conform to requirements of existing communication networks and for systems which exist independently. Over-all system testing by means of multiple tones is discussed and related basic requirements are outlined.

621.396.41:551.510.52(083.57) 1552
Tropo-scatter System Design Charts—L. P. Yeh. (*Electronics*, vol. 31, pp. 91–93; January 17, 1958.) In a frequency-division multiplex tropospheric scatter system, the unweighted signal-to-noise power ratio for a single-hop system may be obtained by consecutive use of the graphs provided.

621.396.43 1553
Microwave Radio Communication under Severe Winter Conditions on the Bonneville Power Administration System—E. M. Strong, E. J. Warchol, and S. E. Benson. (*Commun. and Electronics*, pp. 655–662; November, 1957.) The general layout of the system and terrain and weather conditions are described. Design factors are considered for protection of antennas and supporting structures from winds, snow, and ice.

621.396.44:621.314.22 1554
Contribution of Statistics to the Development Program of a Transformer for the L3 Carrier System—G. J. Levenbach. (*Bell. Sys. Tech. J.*, vol. 37, pp. 23–54; January, 1958.) Statistical methods played a significant part in the development program of the L3 system which can transmit a television signal over a distance of approximately 1000 miles and telephone signals approximately 4000 miles. Experiments were designed to assist in improving the manufacture of the input and output transformers of the amplifiers. Detailed analysis of a few of these experiments is presented.

621.396.65 1555
Wide-Band Microwave Transmission Systems—S. Fedida. (*Point to Point Telecommun.*, vol. 1, pp. 26–50; June, 1957.) The concept of an ideal wide-band channel is developed and the shortcomings of practical systems, due mainly to the type of modulation employed, are enumerated. The requirements of a particular system depend on the type of intelligence to be transmitted and the effect of such requirements on the system's performance is discussed. Variations in the transmission medium and their effect on the system are also examined. Typical practical equipment and techniques are described.

621.396.65:621.396.8 1556
Some Statistical Characteristics for Over-Reach Propagation and Co-frequency Interference in 2-Frequency-System Microwave Multi-relay Circuits—S. Ugai and T. Nomura. (*Rep. Elec. Commun. Lab., Japan*, vol. 5, pp. 10–15; July, 1957.) From tests made in 1951–1953, the variation of received signal power at at 4 kmc for the optical path (50 km) alone, approximates to a gamma distribution. The distribution of the received power in diffraction regions (100–230 km), however, was found (1953–1955) to be classifiable into three types. The interference between signals received optically and after diffraction depends on the receiving system used, but the order of the effect—the cumulative probability—is found from the correlation between the gamma distribution for the optical path and each of the three types of distribution for the diffracted signals. This information assists the design of relay routes to minimize interference.

621.396.65.029.64:621.376.3 1557
A 4000-Mc/s Radio System for the Transmission of Four Telephony Supergroups (240 Circuits)—*P.O. Elec. Engr. J.*, vol. 50, Pt. 2 and Pt. 3, pp. 106–119 and 150–158; July and October, 1957.)

Part 1 Outline Description of the System and Detailed Description of Radio and Intermediate-Frequency Equipment—R. L. Corke, E. V. Ephgrave, J. Hooper, and D. Wray. A description of equipment for FM transmission over a single-hop link of 55 miles.

Part 2 The Baseband and Supervisory Equipment—R. P. Froom, J. D. C. Madder, and C. G. Hilton. Preliminary over-all performance data for the link as a whole are given.

621.396.822.1:621.396.41 1558
Crosstalk Problems in Radio Relays with Double Modulation: P.P.M. and F.M.—E. Katz. (*Bul. Inst. Polit. Bucureşti*, vol. 18, pp. 331–352; January/June, 1956.) A special demodulation method is described which reduces crosstalk and effectively doubles the signal-to-noise ratio.

SUBSIDIARY APPARATUS

621.526:621.314.7 1559
Semiconductors Shrink Servo-System Size—H. L. Aronson and W. R. Lamb. (*Electronics*, vol. 31, pp. 69–71; January 3, 1958.) A velocity-type servosystem in which the network design is based on the constant-current driving source and low-impedance load conditions imposed by transistor operation. Output impedance is reduced considerably by using a double common-collector power stage with inverse feedback.

621.3.066.6:537.311.4 1560
The Effects of Dust and Force upon certain Very Light Electrical Contacts—A. J. Maddock. C. C. Fielding, J. H. Batchelor, and A. H. Jiggins. (*Brit. J. Appl. Phys.*, vol. 8, pp. 471–476; December, 1957.) A quantitative relation is established between dust concentration and contact failure. An examination of contact loading shows that contact resistance may rise rapidly at loadings less than 1 mg unless surface films can be permanently removed.

621.314.63:537.311.33 1561
Progress Report on Semiconductor Rectifiers—N. F. Bechtold. (*Electronic Ind. Tele-Tech*, vol. 16, pp. 70–71, 174; October, 1957.) "Findings indicate that silicon and germanium rectifiers are more desirable in high-temperature areas while selenium retains its superiority in high-current uses."

621.318.56 1562
Design of Magnetic Circuits for Miniature Relays—W. J. Richert. (*Electronic Ind. Tele-Tech*, vol. 16, pp. 56–57, 156; October, 1957.) The polarized type of relay with a rotary armature is suggested as the most suitable type for miniaturization.

621.352:541.135.6 1563
The Solion: an Electrochemical Transducer. (*Brit. Commun. Electronics*, vol. 4, p. 617; October, 1957.) The operation of this device is based on concentration polarization where an electric current passing through an electrochemical system is limited by diffusion and convection variables such as fluid flow. The solion consists of a small cylinder of plastic material containing a potassium-iodide solution in which two or more electrodes are immersed; polarization is effected by a 0.9-v battery. When the unit is stimulated by light, heat, sound, pressure, or movement the resultant hydraulic flow within the cell produces a current output. These reactions are reversible.

621.352.7 1564
Wax-Electrolyte Batteries—(*Tech. News. Bull. Natl. Bur. Stand.*, vol. 41, pp. 149–150; October, 1957.) Two types of 37.5-v punched-cell battery have been developed. They are 3 inches long and, respectively, 0.25 and 0.5 inch in diameter, with short-circuit current 0.03 and 0.3 μa. A cell is stamped from a three-layer sandwich consisting of a Zn sheet, a conductive vinyl film, and between the two layers a separator impregnated with a $ZnCl_2$ polyethylene glycol mixture.

621.396.662.6:621.374.3 1565
Self-Setting Servo Gate—E. R. Schmerling. (*Electronics*, vol. 31, p. 71; January 17, 1958.) "Simple circuit, used in ionospheric pulse experiments, picks out pulses transmitted at a fixed repetition frequency in the presence of random noise and improves reliability of synchronizing link by factor of 50."

TELEVISION AND PHOTOTELEGRAPHY

621.397.26:621.396.11.029 1566
Scatter Propagation and its Application to Television—Saxton. (See 1531.)

621.397.5:621.317](083.74) 1567
IRE Standards on Television: Measurement of Luminance Signal Levels, 1958—(PROC. IRE, vol. 46, pp. 482–486; February, 1958.) Standard 58 IRE 23. S1.

621.397.5:621.395.625.3 1568
The Ampex Video Tape-Recording System—R. H. Snyder. (*Brit. Commun. Electronics*, vol. 4, pp. 612–616; October, 1957.) See 4018 of 1957.

621.397.6:621.372.55 1569
A New Video Differential Phase-and-Gain Equalizer—J. H. Clark. (*Commun. and Electronics*, pp. 674–676; November, 1957.) The equalizer uses the nonlinear impedance characteristics of devices such as Ge or Si diodes to compensate undesirable nonlinearity.

621.397.61 1570
Highlight Equalizer Sharpens TV Pictures—M. V. Sullivan. (*Electronics*, vol. 31, pp. 72–74; January 7, 1958.) "Equalization of only the gray-to-white highlight region in the video signal provides better signal-to-noise ratio and improved definition over conventional aperture equalizers covering the full brightness range. Since most image-orthicon noise is in the lowlight region, the improvement stems from a division of the signal into two parts with only the relatively quiet highlight portion equalized for better tonal reproduction."

621.397.61:535.623 1571
A Modified Microwave System for Colour Television Transmission—F. F. McClatchie. (*Commun. and Electronics*, pp. 626–633; November, 1957.) NTSC color television signals

may be transmitted on portable microwave radio relay units in a routine manner if an adequate system is available and standardized operation and maintenance instructions are provided. Necessary modifications to existing equipment are described.

621.397.62:535.88 1572
Large-Scale Projection of the Television Image—H. Jensen. (*Elektrotech. Z., Edn B*, vol. 9, pp. 331–334; August 21, 1957.) Outline of methods proposed and adopted in practice, particularly those using a small high-intensity picture tube with an optical projection system.

621.397.621:621.385.832 1573
90° Scanning—R. H. C. Morgan and K. E. Martin. (*J. Telev. Soc.*, vol. 8, pp. 285–297; July/September, 1957.) The sensitivity and raster quality problems are discussed and two types of scanning coil assembly that reduce the energy required for full deflection are described. Improvements in line-timebase techniques give 90-degree scanning without an increase in the power required.

TRANSMISSION

621.376.32:621.396.41.029.6 1574
An Improved Frequency Modulator for Broad-Band Radio Relay Systems—I. A. Ravenscroft. (*P.O. Elec. Eng. J.*, vol. 50, pt. 3, pp. 186–188; October, 1957.) The linearity is improved by negative feedback so that the modulator may be used in 600-channel telephone circuits conforming to CCIR standards. A method of reducing distortion at high modulation frequencies (up to 2.5 mc) is discussed. See also 286 of 1956 (Ravenscroft and White).

621.396.61:621.314.7 1575
Simple Transistor Transmitter for 1.8 Mc/s—N. Waite. (*RSGB Bull.*, vol. 33, p. 177; October, 1957.) Circuit and constructional details of a single-stage cw transmitter deriving its power supply from a 4.5-volt battery.

621.396.712.029.62:621.523.8 1576
The Automatic 3-kW U.H.F. Transmitter "Hoher Bogen"—P. G. Zehnel and J. Brose. (*Elektronik*, vol. 6, pp. 240–242; August, 1957.) Two 250-watt transmitters, one being in reserve, and a 3-kw amplifying stage are controlled by means of the automatic equipment described.

TUBES AND THERMIONICS

621.314.63:621.318.57:546.28 1577
***p-n-p-n* Switching Diodes**—A. K. Jonscher. (*J. Electronics Control*, vol. 3, pp. 573–586; December, 1957.) "The paper describes silicon diodes of *p-n-p-n* structure which exhibit rapid switching from a high-resistance to a low-resistance state, similar to that described recently by Moll, *et al.* [3899 of 1956]. Two types of characteristics are shown, those switching on both the 'forward' and 'reverse' branches and those giving ordinary diode forward behavior and switching on the 'reverse' branch only. A theoretical treatment is given accounting for this behavior and discussing the nature of the conduction mechanism in the low-resistance state for both directions of current flow."

621.314.63:621.372.632 1578
Semiconductor Diodes Yield Converter Gain—A. Uhlir, Jr, and N. Bronstein. (*Bell Lab. Rec.*, vol. 35, p. 412; October, 1957.) Gains up to 6 db (with adequate bandwidth for some applications) have been achieved with gold-banded germanium diodes when converting from 75 mc to 6 kmc. In "down-conversion" stages using diffused silicon diodes gains up to 45 db have been obtained.

621.314.7 1579
Injection Coefficient and Forward Current /Voltage Characteristic of a Spherical Contact—Z. S. Gribnikov and K. B. Tolpygo. (*Zh. Tekh. Fiz.*, vol. 27, pp. 625–629; April, 1957.) The injection coefficient of a transistor *p-n* junction is examined, and the effect of increasing the radius of the metallic contact, and corresponding increase in donor concentration are discussed.

621.314.7:539.169 1580
Analysis of the Effect of Nuclear Radiation on Transistors—J. J. Loferski. (*J. Appl. Phys.*, vol. 29, pp. 35–40; January, 1958.) A description of the permanent changes produced in Ge transistors. Information on the effect of radiation on semiconductors is reviewed, and the results are combined with transistor theory. The amplification factor of transistors and the effect of the fast neutron field of a reactor are discussed in detail.

621.314.7:546.289 1581
Germanium Junction Transistors—H. Frank. (*Slab. Obz., Praha*, vol. 17, pp. 680–687; December, 1956.) The properties of Czech junction transistors with maximum power loss 20–50 mw and maximum operating frequency 500 kc are considered.

621.314.7:621.3.018.75 1582
The Passage of Single and Periodic Current Pulses through a Transistor—V. G. Kolotilova. (*Zh. Tekh. Fiz.*, vol. 27, pp. 630–637; April, 1957.) Two cases are considered: 1) transistors with earthed base and 2) transistors with earthed emitter or collector.

621.314.7:621.317.3 1583
Measurement of High-Frequency Equivalent-Circuit Parameters of Junction and Surface-Barrier Transistors—A. R. Molozzi, D. F. Page, and A. R. Boothroyd. (IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 120–125; April 5, 1957. Abstract, PROC. IRE, vol. 45, p. 1163; August, 1957.)

621.314.7:621.317.3 1584
Measuring Transistor "Power Gain" at High Frequencies—M. Coffey. (See 1491.)

621.314.7:621.396.822 1585
Transistor Noise in Low-Frequency Region—J. Schubert. (*Arch. elek. Übertragung*, vol. 11, pp. 331–340, 379–385, and 416–423; August–October, 1957.) The validity of an equivalent circuit based on earlier work [see, *e.g.*, 3393 of 1954 (van der Ziel)] is confirmed by measurements of noise characteristics on *p-n-p*-type transistors in the frequency range 86 cps–100 kc. The optimum operating conditions for the minimum noise figure are derived for the earthed-base, earthed-emitter, and earthed-collector circuits and the noise parameters of the equivalent noise quadripole of a transistor are given. Twenty-three references.

621.314.7:621.396.822 1586
Behavior of Noise Figure in Junction Transistors—W. N. Coffey. (PROC. IRE, vol. 46, pp. 495–496; February, 1958.) Noise-figure expressions are modified to include the effect of partial correlation of the emitter and collector noise generators.

621.314.7.002.2 1587
Alloy-Diffusion: a Process for Making Diffused-Base Junction Transistors—J. R. A. Beale. (*Proc. Phys. Soc.*, vol. 70, pp. 1087–1089; November, 1957.) In the process described the base width is determined by the difference in the diffusion depths of two impurities and is practically independent of the alloying depth. A reasonable substantial metal contact to the recrystallized emitter is automatically produced. Examples are quoted of the characteristics achieved in a group of transistors obtained using *p*-type Ge and a carrier metal, such as Pb, containing about 1 per cent Sb and 2 per cent Ga.

621.314.7+621.385].012.8 1588
High-Frequency Parameters of Transistors and Valves—J. Zawels. (*Electronic Eng.*, vol. 30, pp. 15–17; January, 1958.) The high-frequency equivalent circuits of junction transistors and tubes are developed; the exact open-circuit, short-circuit, and hybrid parameters are tabulated.

621.314.7.012.8 1589
Equivalent Circuit of a Transistor Operating at High Frequencies—I. I. Litvinov. (*Radiotekhnika, Moscow*, vol. 11, pp. 25–29; October, 1956.) Formulas are derived for calculating the parameters of the equivalent circuits which are valid over a wide frequency range. The accuracy of the formulas is confirmed by experiment.

621.383.2.032.217.2 1590
Photoelectric Emission of Oxide-Coated Cathodes—T. Yabumoto and S. Yamada. (*J. Phys. Soc. Japan*, vol. 12, p. 1163; October, 1957.) A new peak at 4.6 ev was found in the photo-electric emission of Ba-Sr oxide-coated cathodes exposed at room temperature to an oxygen atmosphere at a pressure of 10^{-5}–10^{-6} mm Hg for about five minutes.

621.383.27 1591
Spectral Response and Linearity of Photomultipliers—H. Edels and W. A. Gambling. (*Brit. J. Appl. Phys.*, vol. 8, pp. 481–482; December, 1957.) Spatial variations in spectral response are shown to change their form with wavelength. With suitably designed circuits, linearity of output with input intensity is achieved for output currents up to at least 10 μa.

621.333.27.032.21 1592
Characteristics of the Dark Current of Unfocused Secondary-Electron Multipliers with Cs_3Sb Photocathodes—F. Eckart. (*Ann. Phys., Lpz.*, vol. 19, pp. 133–144; December 29, 1956.) The temperature dependence of the thermionic emission of the photocathode was investigated in the temperature range from +20° to +50°C. The work function calculated from this is 1.2 ev which is 0.2 ev less than the photo-electric energy derived from the cutoff frequency of the external photo-effect. See also 930 of 1956.

621.383.4:535.371.07 1593
Solid-State Image Amplifiers—G. F. J. Garlick. (*J. Sci. Instr.*, vol. 34, pp. 473–479; December, 1957.) The large-area solid-state light amplifier is built up from sintered layers of photoconductor and phosphor, or from layers of microcrystalline material bonded in a suitable resin. A simple device uses a single evaporated phosphor layer. Image conversion may be possible in the infrared up to 6μ, in the ultraviolet and in the X-ray regions.

621.385.029.6 1594
Plug-In Reflex Klystrons for Microwaves—A. F. Pearce, K. H. Kreuchen, C. Baron, N. Houlding, and S. Ratcliffe. (*J. Electronics Control*, vol. 3, pp. 535–563; December, 1957.) Plug-in klystrons have the advantages of low replacement cost and improved consistency of characteristics. The design of the cavity is flexible and a smaller number of types is required to cover a wide frequency band. A discussion of the design of plug-in klystrons is included and details are given of the construction and characteristics of two British types:

CV2116 for the *S* band and CV2346 for the *X* band. Both these have useful performances over the range 2–12 kmc.

621.385.029.6 1595
Space-Charge Waves along Magnetically Focused Electron Beams—W. W. Rigrod: J. Labus. (PROC. IRE, vol. 46, pp. 358–360; January, 1958.) Comment on 2961 of 1957 and author's reply.

621.385.029.6 1596
Fabrication of Multicavity Magnetrons—A. Singh and N. C. Vaidya. (*J. Sci. Indus. Res.* vol. 16A, pp. 169–175; April, 1957.) Details of techniques and equipment used.

621.385.029.6 1597
Design and Performance of a High-Power Pulsed Magnetron—E. C. Okress, C. H. Gleason, R. A. White, and W. R. Hayter. (IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 161–171; April, 1957. Abstract, PROC. IRE, vol. 45, p. 1164; August, 1957.)

621.385.029.6 1598
Traveling-Wave-Tube Propagation Constants—G. R. Brewer and C. K. Birdsall. (IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 140–144; April, 1957. Abstract, PROC. IRE, vol. 45, pp. 1163–1164; August, 1957.)

621.385.029.6 1599
Effect of Magnetic Field on Coupled Helix Attenuators—M. H. Miller, B. Hershenov, and J. R. Black. (*J. Appl. Phys.*, vol. 28, pp. 1363–1364; November, 1957.) In a traveling-wave tube, a considerable variation of attenuation with magnetic field strength was observed, using a trifilar coupling helice wound from 0.002-inch Kanthal wire.

621.385.029.6 1600
Backward-Wave-Oscillator Starting Conditions—R. D. Weglein. (IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 177–179; April, 1957. Abstract, PROC. IRE, vol. 45, p. 1164; August, 1957.)

621.385.029.6 1601
Backward-Wave Oscillators for the 8000–18,000 Megacycle Band—H. R. Johnson and R. D. Weglein. (IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 180–184; April, 1957. Abstract, PROC. IRE, vol. 45, p. 1164; August, 1957.

621.385.029.6 1602
A High-Power Periodically Focused Traveling-Wave Tube—O. T. Purl, J. R. Anderson, and G. R. Brewer. (PROC. IRE, vol. 46, pp. 441–448; February, 1958.) The design considerations, construction, and performance of an *S*-band, pulsed, kilowatt-level traveling-wave tube, focused by means of periodic permanent magnets are described. The results of studies on the electron-beam defocusing effect of strong RF fields and on the effect of attenuator design on power buildup and saturation level are reported. Factors affecting the design of a high-power traveling-wave tube using periodic magnetic focusing are described in order to show the design constraints imposed by the periodic focusing system.

621.385.029.6:537.533 1603
Some Characteristics of a Cylindrical Electron Stream in Immersed Flow—G. R. Brewer. (IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 134–140; April, 1957. Abstract, PROC. IRE, vol. 45, p. 1163; August, 1957.)

621.385.029.6:537.533 1604
Use of Scanning Slits for Obtaining the Current Distribution in Electron Beams—K. J. Harker. (*J. Appl. Phys.*, vol. 28, pp. 1354–1357; November, 1957.) In axially symmetric electron beams, the current density is related to the current through a slit by an integral equation. This equation is solved and a technique for rapid numerical evaluation given.

621.385.029.6:621.372.2 1605
Analysis of Coupled-Structure Traveling-Wave Tubes—N. Rynn. (IRE TRANS. ON ELECTRON DEVICES, vol. ED-4, pp. 172–177; April, 1957. Abstract, PROC. IRE, vol. 45, p. 1164; August, 1957.)

621.385.029.64:537.533:621.375.9 1606
A Parametric Electron-Beam Amplifier—T. J. Bridges. (PROC. IRE, vol. 46, pp. 494–495; February, 1958.) An amplifier similar to the solid-state amplifier proposed by Suhl (3076 of 1957) can be devised using an electron beam instead of the ferrite. Such an amplifier has the possibility of very-low-noise operation since the fundamental limitations of noise performance of conventional microwave amplifiers do not apply. An amplifier has been built and initial experimental results are given.

621.385.029.65 1607
Experimental 8-MM Klystron Power Amplifiers—T. J. Bridges and H. J. Curnow. (PROC. IRE, vol. 46, pp. 430–432; February, 1958.) A cw output of 75 w was obtained with operation limited by the maximum current density of 1.5 a/cm² that could be drawn from sprayed-oxide cathodes.

621.385.832:621.397.621 1608
90° Scanning—Morgan and Martin. (See 1573.)

MISCELLANEOUS

061.6:621.3 1609
Electronics Research Laboratory—(*Engineer, London*, vol. 204, pp. 604–606; October 25, 1957.) Review of work carried out at the Mullard Research Laboratories, Surrey, England.

621.3(083.74)(083.86) 1610
Index to IRE Standards on Definitions of Terms 1942–1957—(PROC. IRE, vol. 46, pp. 449–476; February, 1958.) Standard 58 IRE 20.S1 containing an alphabetical list of approximately 3500 technical terms with the code number of the appropriate IRE standard and its date of publication.