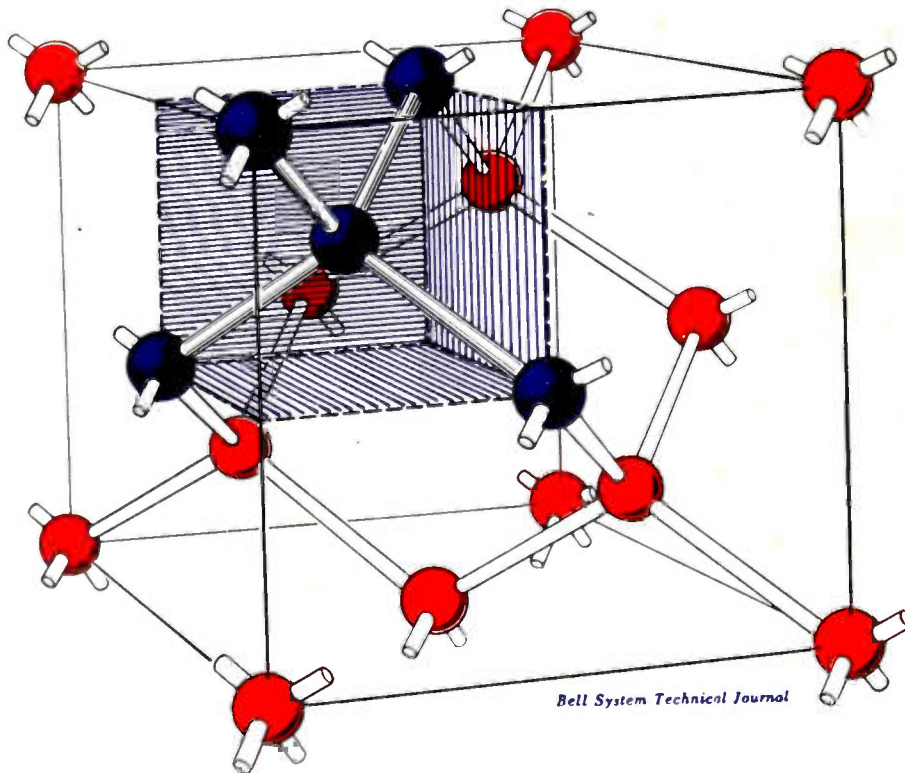# Proceedings

of the I·R·E

TRANSISTOR ISSUE

Atomic Structure of Germanium Crystal

Bell System Technical Journal

IN THIS ISSUE

Forty-Eight Transistor Papers on the Following Topics:

Physics of Transistors
Transistor Materials
Construction of Transistors

Operating Characteristics of Transistors
Transistor Applications
Transistor Circuitry

The Institute of Radio Engineers

Table of Contents will be found following page 64A

Styroflex Coaxial Cable

Now come to us for small or large quantities of Styroflex coaxial cable. We are ready to fill your needs with this new aluminum sheath cable as distributors for the producers — Phelps Dodge Copper Products Corporation.

We stock Styroflex cable for immediate shipment from our plant cut to length, supplied in coils or on reels, formed, fitted or otherwise fabricated to your requirements. We have developed a line of fittings for Styroflex cable soon to be available.

For a full description of Styroflex cable or its fittings send for technical literature which will be mailed at once. It includes data on losses throughout the VHF, UHF and Microwave ranges to 8000 megacycles. A price list is included.

COMMUNICATION PRODUCTS COMPANY•Inc
MARLBORO, NEW JERSEY   Telephone: FReehold 8-1880
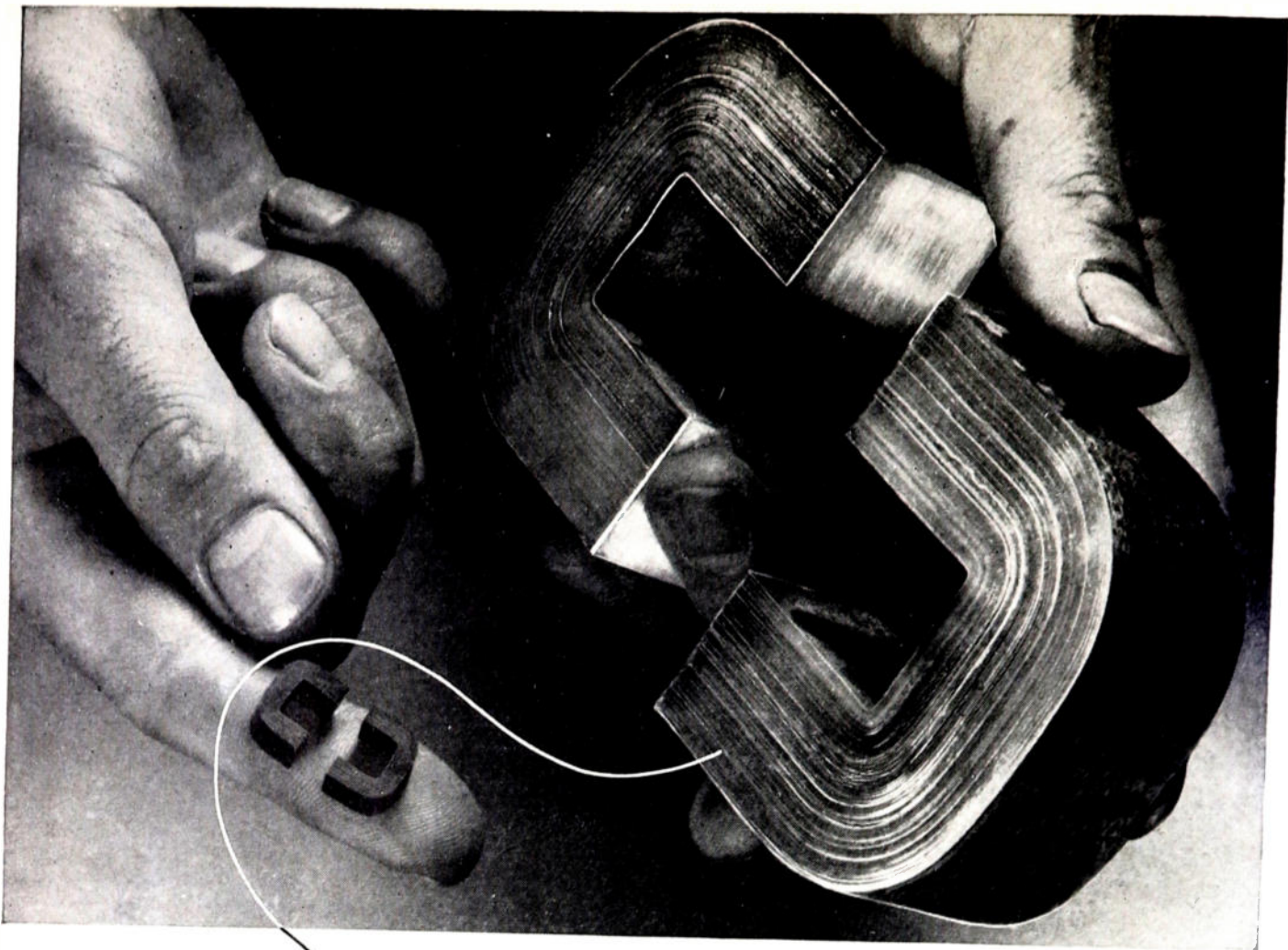
SILECTRON C-CORES... BIG or LITTLE

...*any quantity and any size*

Wound from precision rolled oriented silicon steel strip as thin as .00025"

For users operating on government schedules, Arnold is now producing C-Cores wound from ¼, ½, 1, 2, 4 and 12-mil Silectron strip. The ultra-thin oriented silicon steel strip is rolled to exacting tolerances in our own plant on precision cold-reducing equipment of the most modern type. Winding of cores, processing of butt joints, etc. are carefully controlled, assuring the lowest possible core losses, and freedom from short-circuiting of the laminations.

We can offer prompt delivery in production quantities—*and size is no object, from a fraction of an ounce to C-Cores of 200 pounds or more.* Rigid standard tests—and special electrical tests where required—give you assurance of the highest quality in all gauges. ● *Your inquiries are invited.*

W&D 4363
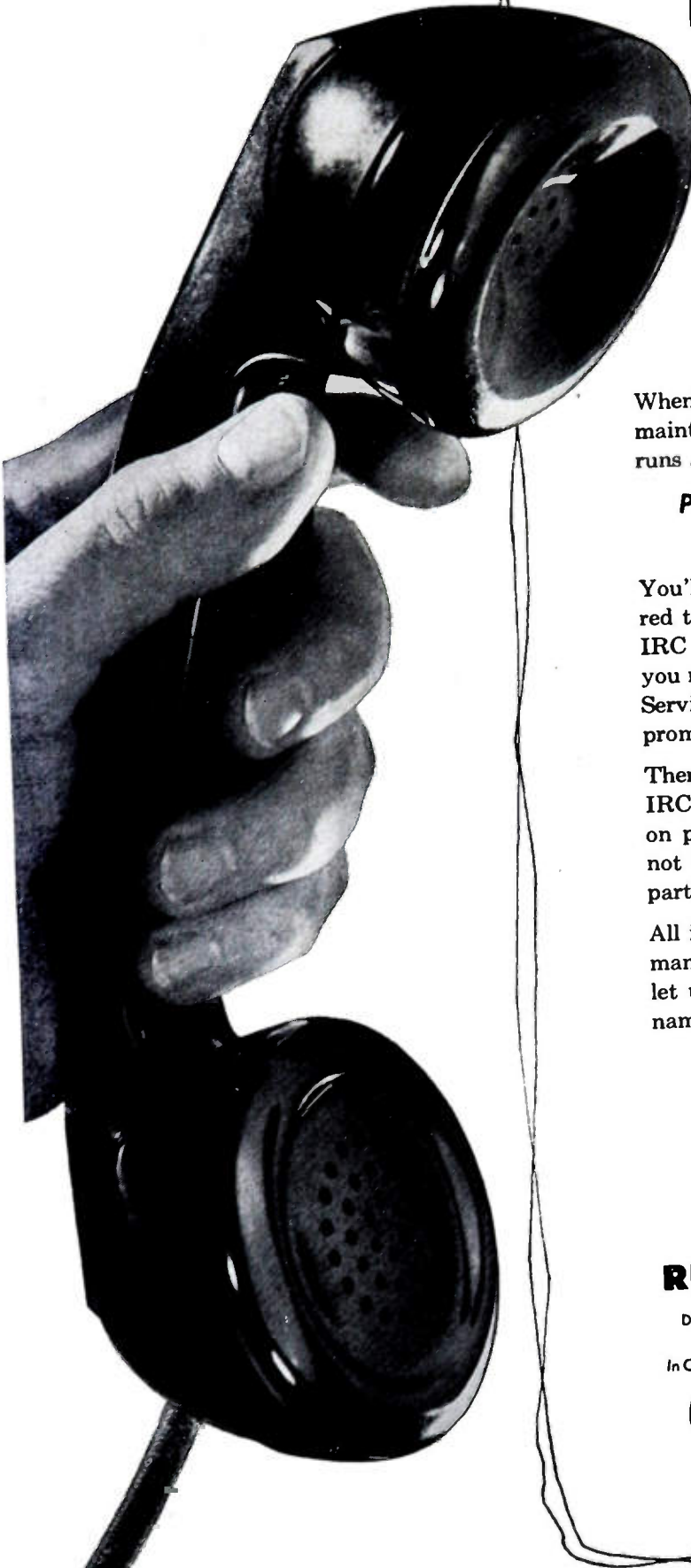
THE ARNOLD ENGINEERING COMPANY

SUBSIDIARY OF ALLEGHENY LUDLUM STEEL CORPORATION

General Office & Plant: Marengo, Illinois

MALLORY electrolytic capacitors

JAN C-62 types

For use in military electronic equipment, Mallory manufactures a line of electrolytic capacitors which will meet the requirements of Specification JAN C-62. Included in the Mallory line is the full selection of standard case styles, ratings and characteristics required under the specification.

Into these military-type capacitors go the same engineering know-how and production craftsmanship which have made Mallory capacitors the standard of quality in the radio and television industry.

Look to Mallory for all your capacitor needs . . . whether for military or for civilian applications.

New Folder Outlines JAN Capacitor Types

A new folder, available on request, condenses the information on type designations of all electrolytic capacitors covered by JAN C-62, to convenient, easy-to-read chart form. It's an ideal reference for everyone who specifies or uses electrolytic capacitors. Write to Mallory for your copy today.

P.R. MALLORY & CO. Inc.
MALLORY

SERVING INDUSTRY WITH THESE PRODUCTS:
Electromechanical—Resistors • Switches • Television Tuners • Vibrators
Electrochemical—Capacitors • Rectifiers • Mercury Dry Batteries
Metallurgical—Contacts • Special Metals and Ceramics • Welding Materials

P. R. MALLORY & CO., INC., INDIANAPOLIS 6, INDIANA

ENGINEERED TO MEET ELECTRONICS' TOUGHEST ASSIGNMENTS

Sylvania Premium-Performance Subminiatures meet requirements of U. S. Armed Services
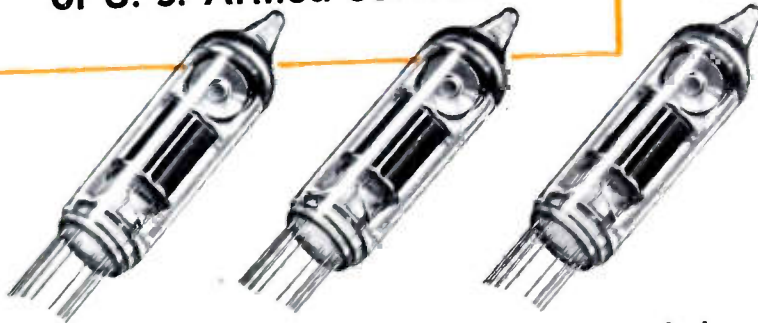
SYLVANIA proudly offers Premium-Performance Subminiature Tubes to meet your toughest application assignments in both military and civilian equipment.

These tubes are rugged plus! They're precision engineered and quality tested for maximum dependability under conditions of high shock, vibration, and temperature. The long life of these tubes is further assured by a special Sylvania "burn-in" process before testing.

Developed by Sylvania scientists especially for the U. S. Armed Services, these premium tubes are designed for a life expectancy (80% minimum, average life) of 5000 hours, under test conditions, in an ambient temperature of 30°C.

So, if you are looking for really rugged tubes with proved ability to fight off vibration and shock, be sure to specify Sylvania Premium-Performance Subminiatures. New illustrated folder gives ratings, characteristics, and application data. Write for your file-copy of this folder NOW. Address: Sylvania Electric Products Inc., Dept. R-2111, 1740 Broadway, New York 19, N. Y.

SEND FOR THIS FOLDER!

SYLVANIA

RADIO TUBES; TELEVISION PICTURE TUBES; ELECTRONIC PRODUCTS; ELECTRONIC TEST EQUIPMENT; FLUORESCENT TUBES, FIXTURES, SIGN TUBING, WIRING DEVICES; LIGHT BULBS; PHOTOLAMPS; TELEVISION SETS

TYPE 252, JAN-R-19, Type RA20

2 watt, 1¹⁷⁄₆₄″ diameter variable wirewound resistor. Also available with other special military features not covered by JAN-R-19. Attached Switch can be supplied.

Resistance	RA20, JAN Shaft Type SD		RA20 High Torque, JAN Shaft Type SD	
	CTS Part	JAN-R-19 TYPE	CTS Part	JAN-R-19 TYPE
50 ±10%	B8079	RA20A1SD500AK	X3496	RA20A2SD500AK
100 ±10%	W6929	RA20A1SD101AK	L9388	RA20A2SD101AK
250 ±10%	X3497	RA20A1SD251AK	M9879	RA20A2SD251AK
500 ±10%	W6931	RA20A1SD501AK	X3498	RA20A2SD501AK
1000 ±10%	W6932	RA20A1SD102AK	X3499	RA20A2SD102AK
1500 ±10%	W6933	RA20A1SD152AK	M9809	RA20A2SD152AK
2500 ±10%	W6934	RA20A1SD252AK	L9103	RA20A2SD252AK
5000 ±10%	W6935	RA20A1SD502AK	L9104	RA20A2SD502AK
10,000 ±10%	W6936	RA20A1SD103AK	H8979	RA20A2SD103AK

TYPE 25, JAN-R-19, Type RA30 (May also be used as Type RA25)

4 watt, 1¹⁷⁄₃₂″ diameter variable wirewound resistor. Also available with other special military features not covered by JAN-R-19. Attached Switch can be supplied.

Resistance	RA30, JAN Shaft Type SO		RA30 High Torque, JAN Shaft Type SO	
	CTS Part	JAN-R-19 TYPE	CTS Part	JAN-R-19 TYPE
50 ±10%	X3502	RA30A1SD500AK	W2837	RA30A2SD500AK
100 ±10%	X3503	RA30A1SD101AK	X3504	RA30A2SD101AK
250 ±10%	X3505	RA30A1SD251AK	X3506	RA30A2SD251AK
500 ±10%	X3507	RA30A1SD501AK	M7566	RA30A2SD501AK
1000 ±10%	X3508	RA30A1SD102AK	S2444	RA30A2SD102AK
1500 ±10%	X3509	RA30A1SD152AK	X3510	RA30A2SD152AK
2500 ±10%	X3511	RA30A1SD252AK	S2736	RA30A2SD252AK
5000 ±10%	Q1409	RA30A1SD502AK	X3512	RA30A2SD502AK
10,000 ±10%	X3513	RA30A1SD103AK	R1561	RA30A2SD103AK
15,000 ±10%	X3514	RA30A1SD153AK	L9107	RA30A2SD153AK

TYPE 65

½ watt 70° C, ¾″ diameter __miniaturized__ variable composition resistor.

Resistance	CTS Part CTS Shaft Type RE	CTS Part Locking Bushing CTS Shaft Type LT-2
250±10%	X3516	X3530
500±10%	X3517	X3531
1000±10%	X3518	X3532
2500±10%	X3519	X3533
5000±10%	X3520	X3534
10,000±10%	X3521	X3535
25,000±10%	X3522	X3536
50,000±10%	X3523	X3537
100,000±10%	X3524	X3538
250,000±10%	X3525	X3539
500,000±10%	X3526	X3540
1 Meg±20%	X3527	X3541
2.5 Meg±25%	X3528	X3542

TYPE 95, JAN-R-94, Type RV4

2 watt 70°C, 1⅛″ diameter variable composition resistor. Also available with other special military features not covered by JAN-R-94. Attached Switch can be supplied.

Resistance	JAN-R-94 TYPE RV4 JAN Shaft Type SD	JAN-R-94 TYPE RV4 JAN Shaft Type RJ	CTS Part Non-JAN Locking Bushing CTS Shaft Type LT-1
100±10%	RV4ATSD101A	RV4ATRJ101A	W3160
250±10%	RV4ATSD251A	RV4ATRJ251A	W3161
500±10%	RV4ATSD501A	RV4ATRJ501A	W3162
1000±10%	RV4ATSD102A	RV4ATRJ102A	W3166
2500±10%	RV4ATSD252A	RV4ATRJ252A	W3163
5000±10%	RV4ATSD502A	RV4ATRJ502A	W3164
10,000±10%	RV4ATSD103A	RV4ATRJ103A	W3167
25,000±10%	RV4ATSD253A	RV4ATRJ253A	W3168
50,000±10%	RV4ATSD503A	RV4ATRJ503A	W3169
100,000±10%	RV4ATSD104A	RV4ATRJ104A	W3170
250,000±10%	RV4ATSD254A	RV4ATRJ254A	W3171
500,000±10%	RV4ATSD504A	RV4ATRJ504A	W3172
1 Meg±20%	RV4ATSD105B	RV4ATRJ105B	W3173
2.5 Meg±20%	RV4ATSD255B	RV4ATRJ255B	W3165
5 Meg±20%	RV4ATSD505B	RV4ATRJ505B	W3159

TYPE 45, JAN-R-94, Type RV2

¼ watt, 15/16″ diameter variable composition resistor. Also available with other special military features not covered by JAN-R-94. Attached Switch can be supplied.

Resistance	RV2, JAN Shaft Type SD CTS Part	JAN-R-94 TYPE	CTS Part Non-JAN Locking Bushing CTS Shaft Type LT-1
100±10%	A5876	RV2ATSD101A	A5922
250±10%	A5877	RV2ATSD251A	A5923
500±10%	A5878	RV2ATSD501A	A5924
1000±10%	A5879	RV2ATSD102A	A5925
2500±10%	A5880	RV2ATSD252A	A5926
5000±10%	A5881	RV2ATSD502A	A5927
10,000±10%	A5882	RV2ATSD103A	A5928
25,000±10%	A5883	RV2ATSD253A	A5929
50,000±10%	A5884	RV2ATSD503A	A5930
100,000±10%	A5885	RV2ATSD104A	A5931
250,000±10%	A5886	RV2ATSD254A	A5932
500,000±10%	A5887	RV2ATSD504A	A5933
1 Meg±20%	A5888	RV2ATSD105B	A5934
2.5 Meg±20%	A5889	RV2ATSD255B	A5935

TYPE 35, JAN-R-94, Type RV3

½ watt, 1⅛″ diameter variable composition resistor. Also available with other special military features not covered by JAN-R-94. Attached Switch can be supplied.

Resistance	RV3, JAN Shaft Type SD CTS Part	JAN-R-94 TYPE	CTS Part Non-JAN Locking Bushing CTS Shaft Type LT-1
100±10%	A5861	RV3ATSD101A	A5907
250±10%	A5862	RV3ATSD251A	A5908
500±10%	A5863	RV3ATSD501A	A5909
1000±10%	A5864	RV3ATSD102A	A5910
2500±10%	A5865	RV3ATSD252A	A5911
5000±10%	A5866	RV3ATSD502A	A5912
10,000±10%	A5867	RV3ATSD103A	A5913
25,000±10%	A5868	RV3ATSD253A	A5914
50,000±10%	A5869	RV3ATSD503A	A5915
100,000±10%	A5870	RV3ATSD104A	A5916
250,000±10%	A5871	RV3ATSD254A	A5917
500,000±10%	A5872	RV3ATSD504A	A5918
1 Meg±20%	A5873	RV3ATSD105B	A5919
2.5 Meg±20%	A5874	RV3ATSD255B	A5920
5 Meg±20%	A5875	RV3ATSD505B	A5921

JAN SHAFT TYPE SD

250 ±.001
SCREW DRIVER SLOT
.047 ±.005 WIDE x .063 ±.015/.000 DEEP
3/8
7/8
⅜-32P-NEF-2 THD.
MOUNTING HARDWARE ASSEMBLED
MOUNTING NUT 7/16 HEX x 1/32
LOCK WASHER #1920A

JAN SHAFT TYPE RJ
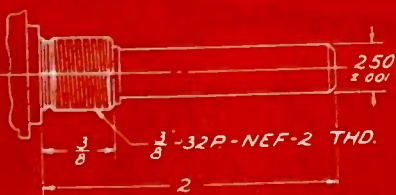
250 ±.001
3/8
2
⅜-32P-NEF-2 THD.
MOUNTING HARDWARE ASSEMBLED
MOUNTING NUT 9/16 HEX. x 1/32
LOCK WASHER #1920A

CTS SHAFT TYPE LT-1
LOCKING BUSHING

250 ±.001
SCREW DRIVER SLOT
.047 ±.005 WIDE x .063 ±.015/.000 DEEP
1/2
5/8
⅜-32P-NEF-2 THD.
MOUNTING HARDWARE ASSEMBLED
MOUNTING NUT 7/16 HEX. x 1/32
LOCK NUT ½ HEX. x 1/32
LOCK WASHER #1920A

J. B. THARPE
National Sales Mgr.
Clifton, New Jersey

H. BLDOMBERG
Central District Mgr.
Chicago

L. PETT
Western District Mgr.
Dallas

L. C. RADFORD, JR.
Eastern District Mgr.
Washington, D. C.

F. O'CONNELL
Baltimore

G. SCOTT
New York City

J. W. MORRISEY
New York City

T. W. KIRKSEY
Dallas

T. MOSELEY
Dallas

W. I. McCORD
Chicago

B. J. KLINDWORTH
Chicago

LOS ANGELES

J. BECKER
Chicago

C. BELL
Atlanta-Charlotte

R. MYERS
Los Angeles

WM. C. O'BRIEN
Los Angeles

*In addition to these outstanding Representatives located in your area,
DuMont maintains a staff of highly experienced Sales Engineers.
These men, backed by years of field experience, work in conjunction
with your Representative and provide him with the necessary informa-
tion and service that may be required to fulfill your particular needs.*

maximum e.r.p.

Let's get our circuits straight

PRINTED CIRCUITS ARE NOT PRINTED ELECTRONIC CIRCUITS

PRINTED ELECTRONIC CIRCUITS are complete or partial circuits in truly miniature sizes — furnished *complete* with conductors, resistors, capacitors and brought out to convenient, permanently anchored mechanical leads. Centralab, the originators of Printed *Electronic* Circuits, makes the world's most complete line — from single resistor plates to complete speech amplifiers.

A PRINTED CIRCUIT is a conductive pattern of an electric circuit, *but* provides conductors only. Don't be misled. A Printed Circuit is *not* a Printed *Electronic* Circuit. There is a place for both in electronic design. Many times they can be used together in the same circuit. But *don't* expect Printed Circuits to do the job that can be provided only by Printed *Electronic* Circuits.

CENTRALAB now offers
smaller sizes in PRINTED
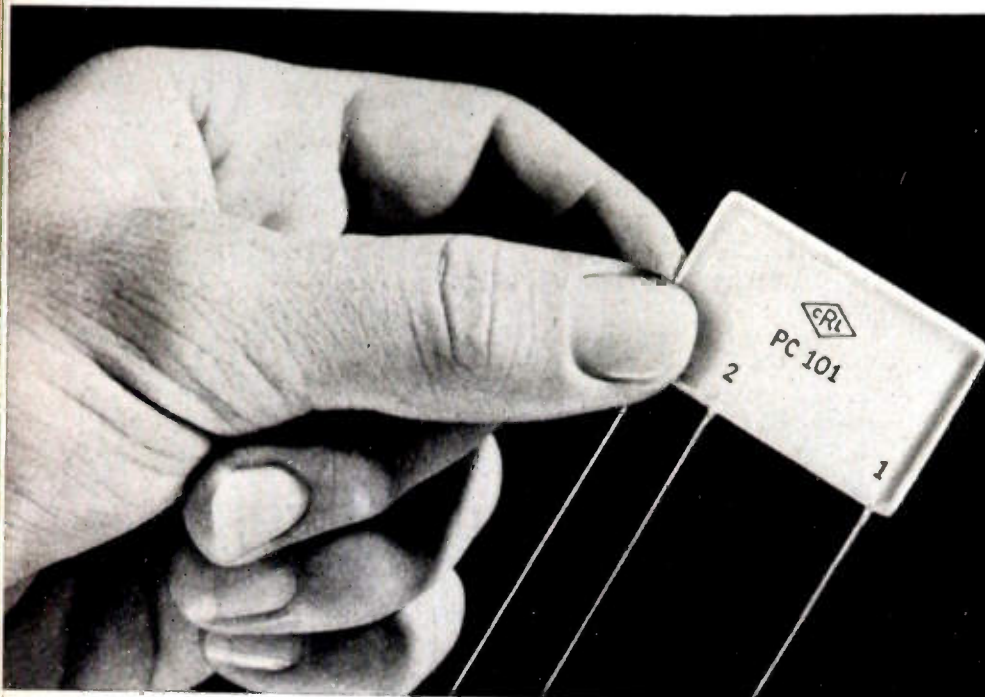
Compare the size of the former Vertical Integrator, (shown actual size at left) with Centralab's new smaller design (actual size, below). Only ⅔ as much space is needed by this new miniaturized Printed Electronic Circuit.
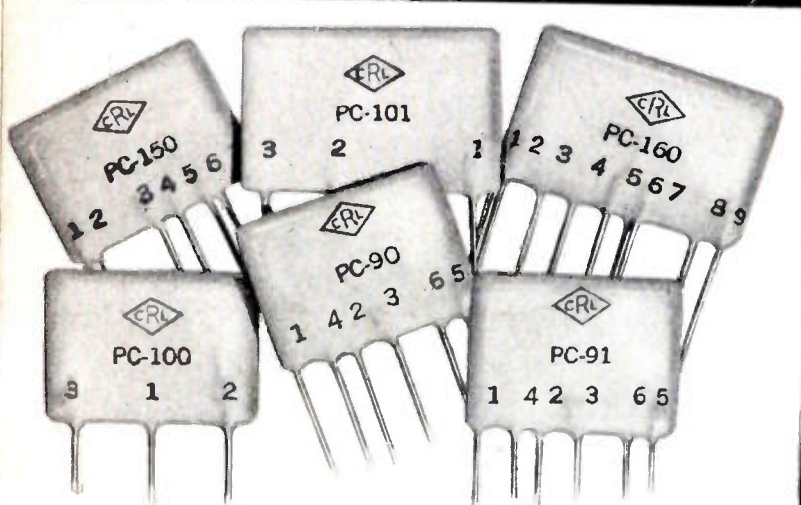
Don't overlook the savings achieved by new, reduced prices on this "bargain group" of PEC's. Check coupon for bulletins. Pentodes (Bulletin 42-128), Vertical Integrators (Bulletin 42-126), Audets (Bulletin 42-129), Pendets (Bulletin 42-149).
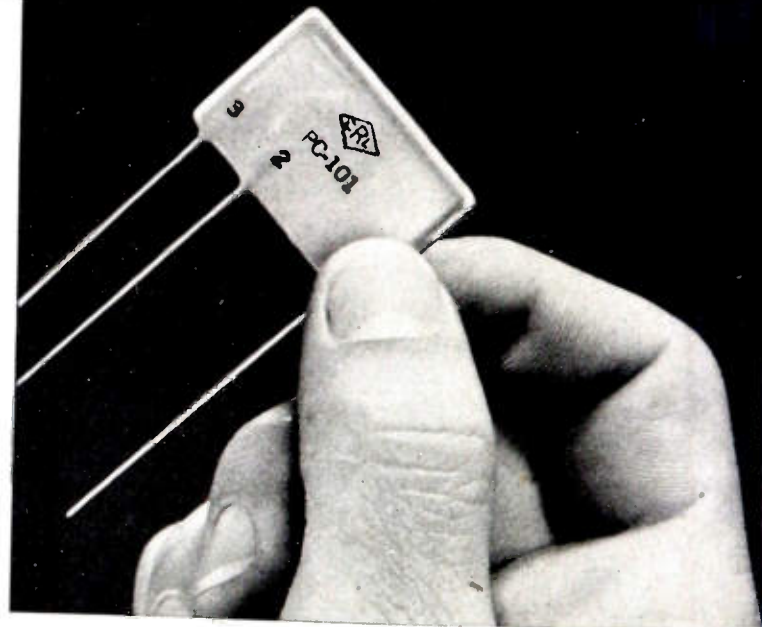
Now — Centralab gives you even more versatility . . . still greater savings in electronic design. Yes, the prices of several Printed *Electronic* Circuits have been reduced. What's more, these components have been miniaturized to still smaller sizes. We've achieved maximum compactness plus top performance . . . at a new low price.

If your designs specify the capacities fulfilled by Pentodes, Vertical Integrators, Audets, or Pendets — look forward to savings ranging from 0.1 to 7 cents per unit.

Actually, these miniature components have always saved you money in time and labor. Now, for the first time, their *first cost is less* than that of the components they replace.

Add up these savings — lower first cost . . . less production time and labor . . . reduced purchasing and inventory requirements. No wonder volume users find they can save thousands of dollars with Centralab Printed *Electronic* Circuits.

even greater savings,
ELECTRONIC CIRCUITS

Save time and money ... space and weight with these PEC's

NEWS and NEW PRODUCTS

November 1952

Duplex Signaling Unit

The Hammerlund Mfg. Co., Inc., 460 West 34 St., New York 1, N. Y., has announced the introduction of a new Duplex Signaling Unit consisting of a tone transmitter and frequence selective receiver designed to operate over wire lines, telephone or power line carrier, and radio or microwave communications circuits. The units may be used to transmit and receive signaling, dialing, telemetering, teleprinting, supervisory controls or other information.
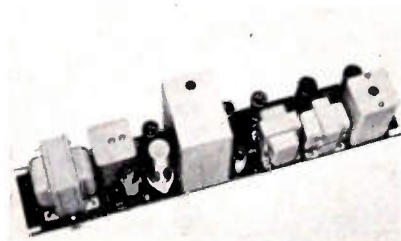
The equipment is designed for use by pipeline and power companies, military and governmental agencies, railroads, airports, and other groups requiring remote on-off switching, continuous indication of operating conditions, and automatic detection of wire line or power source failures along the systems.

These DSU units have been designed to fulfill the need resulting from the increased use of radio and microwave systems for functions other than simple communications by implementing these new operations at minimum cost and at the same time making maximum use of spectrum space.

Transmitters and receivers are available on the same, or different frequencies between 2,000 and 6,025 cps. Up to 33 channels can be used over a single circuit.

UHF Attenuator

Empire Devices, Inc., 38–25 Bell Blvd., Bayside 61, N. Y., present their new laboratory standard UHF Attenuator, Model AT-50, which is now in full production.

TWO TYPICAL ATTENUATOR PADS
(20DB & 60DB)

This is a resistive T-network of concentric line construction, frequency range

— dc to 4,000 mc. The vswr is better than 1.1 to 1 at all frequencies within the above range. Attenuation—standard: 3, 6, 10, 20, 40, 60 db. Other values available upon request. Accuracy to $\pm \frac{1}{2}$ db.

UHF Attenuator, Model AT-50 has choice of type "C" or type "H" connectors, (one male, one female). The rated power is 250 mw continuous, 500 w peak dissipation.

Empire Devices also has available a UHF Power Attenuator, Model AT-60 which has a power dissipation of 2 w continuous, 2 kw peak, and UHF Step Attenuator, Model AT-100.

Oscillator

A new portable oscillator, designed as a source of signal power for field use, has been announced by the Southwestern Industrial Electronics Co., 2831 Post Oak Road, P. O. Box 13058, Houston 19, Texas.

The Model "MB-1" Oscillator derives its operating power entirely from self-contained batteries and covers a frequency range of 2 to 20,000 cps in 4 decade ranges. An output voltage of 5.5 volts is delivered to a 2,000 ohm load, and the instrument has an internal impedance of approximately 400 ohms. Distortion is less than 1 per cent in the audio spectrum, and the dial is accurate to within 2 per cent of its indication. The unit is equipped with a switch which decreases battery gain when full output is not required, thereby increasing battery life to 100 hours or more for intermittent service. The batteries are housed in a separate compartment so that they can be replaced in the field without exposing the circuit to dust and weather. Faulty operation as a result of moisture condensates has been minimized by careful circuit design and by the use of a special range selector switch. The unit is housed in a splash proof welded aluminum cabinet—12×12×11 inches, and weighs 30 pounds with batteries.

Cycling Timer

The Add-A-Cycle Timer is announced by the Becker Equipment Co., 3020 North Cicero Ave., Chicago, Ill. Set-up time is reduced to a minimum through the use of plug-in adjustable timers supplied in six overlapping ranges from 1.5 to 12 seconds, to 15 to 120 minutes. Thus it is possible to have adjacent cycles with widely varying time limits.

The Add-A-Cycle is furnished as a basic unit having two plug-in positions. Any number of them may be cabled together and will cycle as a group so an unlimited number of time cps may be obtained. Normal rating is 10 amperes, 115V, 60 cps, but special ratings may be had upon request.

Wide-Range Motor Control

The 302-A, a wide-range speed and torque generator, is being produced by Industrial Control Co., Wyandanch, Long Island, N. Y. It consists of a special motor and control box operating from the 117 v 60 cps line. Motor speed can be varied from 100 to 10,000 rpm, with a maximum output torque of 1.0 oz. in.
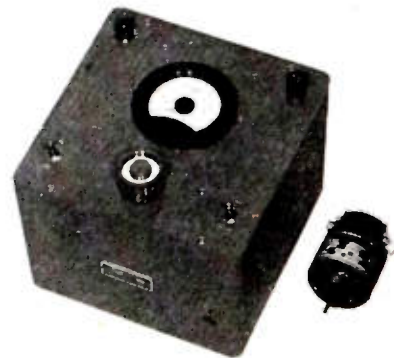
Matching gear boxes are available for any reduction ratio. Motor speed can be controlled remotely, either by a varying dc voltage or a shaft position. The speed is read directly on a front panel meter and provisions are made to measure the torque delivered to the load.

The 302-A finds application in automatic control systems, light machine tools, industrial process control, coil-winding equipment, servo directors, variable speed turntables and textile looms. With it, running torque measurements can be made on devices such as gear boxes and tachometers.

RELIATRON Tubes are backed by Westinghouse Reliability

Because of Westinghouse experience and the unlimited resources and facilities of its new Electronic Tube Division, it is now producing electronic tubes which are the finest ever made...Westinghouse RELIATRON Tubes.

TUBE RESEARCH AND DEVELOPMENT

Westinghouse tube leadership is based on the untiring efforts of its research staff. These men are now improving present tube types and developing new types for superior service and new applications, including UHF.

QUALITY CONTROL

RELIATRON Tube performance is assured by exacting quality control. Every step in the manufacture of RELIATRON Tubes—from raw materials to finished product—must meet standards which are the highest in the industry.

ENGINEERING AND SALES SERVICES

Whatever your tube problem, you will find Westinghouse electronic tube sales representatives and application engineers at your service. Sales and engineering offices are strategically located throughout the country to serve you.

ADVERTISING

Trade acceptance of Westinghouse RELIATRON Tubes will be aided by a nationwide advertising campaign second to none. Sales promotion programs for distributors and service dealers will be hard-hitting sales builders. Your product or service will profit from the fullest consumer acceptance.

DISTRIBUTORS, EQUIPMENT MANUFACTURERS, WRITE NOW

For complete information on the Westinghouse line of RELIATRON Receiving Tubes, Television Picture Tubes, and transmitting, industrial, and special purpose tubes, write or wire Westinghouse Electric Corporation, Dept. C-11, Elmira, New York. Or call your nearest Westinghouse Electronic Tube Division Sales office.
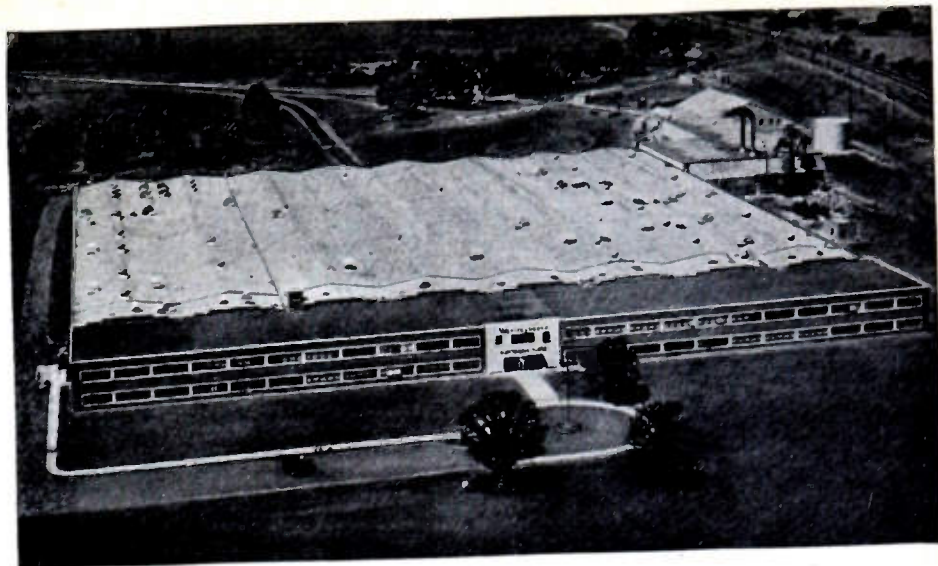
WESTINGHOUSE IN ELMIRA, NEW YORK

360,000 square feet of steel, glass and brick designed for one thing—to house the most efficient electronic tube production in the world. Here are produced Westinghouse RELIATRON television picture tubes, transmitting tubes, industrial tubes, special purpose tubes—all of unsurpassed quality. Here, too, is located the headquarters of the Westinghouse Electronic Tube Division with sales, engineering and production management ready to extend a warm welcome to you.

WESTINGHOUSE IN BATH, NEW YORK

This Westinghouse Receiving Tube plant is another 220,000 square feet of modern production efficiency. It lies only a few miles from a major source of glass tube envelopes. It is served by all modern transportation media to assure prompt shipment of your requirements—only hours away from all principal markets. Here at Bath the most modern equipment is operated by the industry's leading craftsmen. From it are shipped the finest receiving tubes in the industry—Westinghouse RELIATRON Tubes.
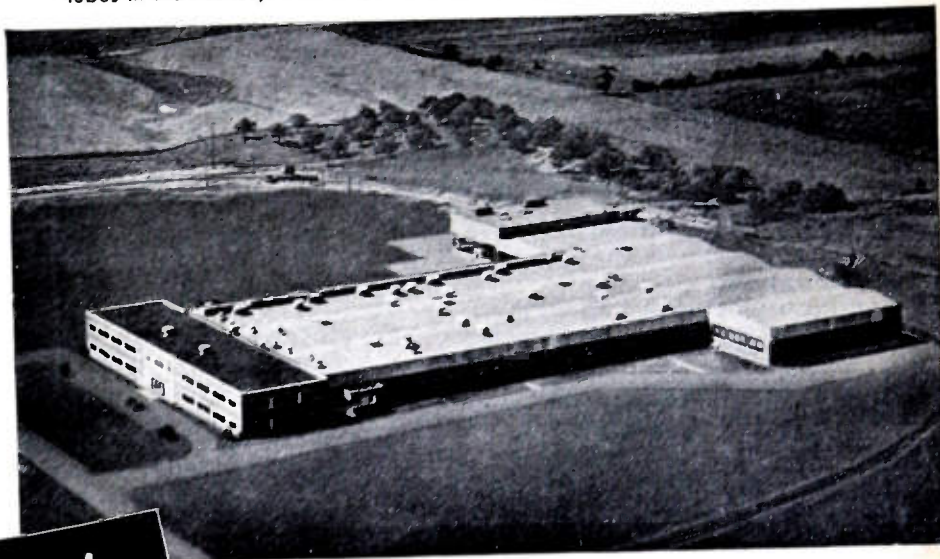
YOU CAN BE SURE...IF IT'S

Westinghouse

ELECTRONIC TUBE DIVISION

WESTINGHOUSE ELECTRIC CORPORATION, ELMIRA, N. Y.

ET-95001

An automatic heat treat machine. Production is about 3 times that possible with manual methods while quality is held within very close limits.

CRUCIBLE ALNICO MAGNETS

KEEP COSTS DOWN ... *through*

automatic production that gives quality control

Alnico magnets have been getting smaller and lighter, thanks to production techniques in use at Crucible. Automatic machinery cuts the possibility of human error to a minimum, so rejections are low. This helps to maintain stable price levels in the face of rising material and labor costs. At the same time, Crucible's rigid inspection standards and attention to quality have developed a magnet with the *highest gap flux per unit weight of any on the market.*

Today, Crucible can offer lighter, magnetically stronger Alnico magnets because of these automatic production techniques developed over the sixteen years that we have been producing the Alnico alloys. And behind our familiarity with permanent magnets lies more than 52 years' experience with specialty steelmaking. Let us advise you on your magnet problem.

50A

Units shown magnified approximately 2½ times

Make _sure_ of meeting government "specs"...
see C.T.C. for ceramic insulated components

You have to be 100% on-the-beam if your equipment is to withstand the conditions it must undergo in military service.

That's why manufacturers using electrical and electronic components turn to C.T.C. for their ceramic insulated units. Our long experience and constant dealing with government requirements have gained us a wide acceptance as an outstanding supplier to those working on U. S. contracts ... especially for the armed forces.

Whatever your needs in ceramic insulated terminals, feed-throughs or terminal boards you can depend on C.T.C. We meet the most exacting government standards for materials, tolerances, finishes, moisture prevention and anti-fungus treatment. Fin-

ishes on metal surfaces for instance, can be hot tinned, electro-tinned, cadmium plated, silver plated or gold plated to your requirements. All ceramic units in our standard line are grade L-5, silicone impregnated.

C.T.C. offers a consulting service at no extra charge to help solve your spe-

cial problems. For all specifications and prices, write to Cambridge Thermionic Corporation, 456 Concord Avenue, Cambridge 38, Mass. West Coast Manufacturers contact: E. V. Roberts, 5068 West Washington Boulevard, Los Angeles 16 and 988 Market Street, San Francisco, California.

CAMBRIDGE THERMIONIC CORPORATION

custom or standard ... the guaranteed components
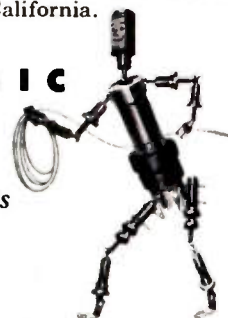
AVAILABLE NOW in a wide range of sizes!

GOVERNMENT APPROVED

MIL-W-5086

AIRCRAFT WIRE

AWG SIZES
24-22-20-18-16-14
12-10-8-6

ALSO
MIL-C-7078 (MIL-W-5086)
WITH SHIELD

COMPLETELY EXTRUDED INSULATIONS

Primary Insulation: Polyvinylchloride
Overall Insulation: Nylon Jacket

Write for samples and complete test data.

RHODE ISLAND INSULATED WIRE CO., Inc.

53 BURNHAM AVENUE   CRANSTON, RHODE ISLAND

National Sales Offices: 624 S. Michigan Blvd., Chicago, Ill.

PROCEEDINGS OF THE I.R.E.®

Published Monthly by

The Institute of Radio Engineers, Inc.

PROCEEDINGS OF THE I.R.E

PROCEEDINGS OF THE I.R.E.®

The Transistor Issue

The Board of Directors of The Institute of Radio Engineers, in its continued endeavors to meet more fully the needs of the membership, has established the precedent of authorizing on a nonscheduled basis special issues of the PROCEEDINGS OF THE I.R.E. Each such special issue is devoted, in timely fashion, to new and basically important portions of the domain of communications and electronics.

This present, greatly expanded issue of the PROCEEDINGS, the second such special issue to be published, is one assigned to a novel, significant, and rapidly developing field, namely, the broad subject of semiconductors and those of their properties and applications of primary interest to the membership of the Institute. Specifically, it is concerned also with electronic devices based on germanium, thus including the whole class of components known as "transistors."

Upon its emergence four years ago, the transistor promised to be one of the truly major advances, even in an art which has been accustomed to giant-stride progress. The fulfillment of this promise, however, depended on the difficult task of determining how to transform the transistor from a laboratory oddity to a practical device which could be manufactured uniformly and in quantity. Marked strides of recent months has furthered progress rapidly toward this goal, so that the threshold of activity has now been reached where the many capabilities of the transistor may be systematically explored, developed, and widely utilized.

The subject is a vast one, and under rapid development. There have, nevertheless, been gathered here a large selection of significant and instructive original papers covering both theory and practice, which it is believed will constitute a substantial and timely contribution to knowledge and expansion of this new and important field.

—The Editor.

Transistor Lineage

THE EXTENSIVE and rewarding research and development work already carried out in the transistor field fully justifies the dedication of this issue of the PROCEEDINGS OF THE I.R.E. to the electronic aspects of certain semiconductors. In a similar sense, this issue may be regarded as a tribute to those effective and tenacious workers and organizations many of whose accomplishments are described herein.

It seems timely and permissible also to endeavor partly to orient the transistor in certain of its historical and evolutionary respects.

Students of scientific advances often find that these largely issue from the operation of two factors, namely, progressive cyclic history, and inventive dependence on availability of instrumentalities.

Progressive cyclic history involves the recurrence of events or material arrangements in a form strongly suggestive of, or broadly based on earlier occurrences or assemblies. In brief, history repeats itself—but always with a difference. Examples of this cyclic phenomenon are modern microwave developments as a recent vastly improved version of the original Hertzian wave experiments. Radar obviously suggests the reflections of electromagnetic waves, both in the optical and radio fields as long ago discovered. Coherers and crystal detectors seem to be the remote and primitive ancestors of the semiconductive rectifiers and amplifiers of today. There are many other examples of the repetitive tendency which will occur to the readers.

It should be noted, however, that the recurrent or later invention is always an elaboration, improvement, or inspirational modification of earlier knowledge, and forms a substantial addition to mankind's store of information. For this reason, the inventive genius involved in the later developments often equals or even transcends that involved in the earlier and analogous case, even though historically the later work may sometimes be overshadowed by the pioneer aspect of the earlier inventive work.

The second factor mentioned in connection with scientific advances is one which is frequently overlooked. Inventive dependence on the availability of suitable instrumentalities is simply the obvious need for some physical materials or methods whereby an invention may actually be embodied in usable form. Inventions, in the effective sense, can usually be made only after the necessary factors for their physical creation are at hand or, by reasonable expectation, are soon to be at the inventor's disposal.

It would hardly have been possible, for example, to have invented the picture tubes of television before electron-beam excited phosphors were known, nor could television camera tubes have been produced before controllable photosensitive surfaces were available. Yet, great genius may be displayed in proposing nonrealizable and abortive invention. Nevertheless, mankind reserves its approval for the actual accomplishment. A similar comment is appropriate to the effect that invention, which too far precedes the availability of correct guiding principles, arrangements, or materials, may tend dangerously toward "science fiction."

Considering both these historical factors, the transistor is seen to have an honorable lineage, to be a worthy and advanced member of its family, and to have been born when electronic, chemical, and metallurgical knowledge in this field was ripe. Aside from its excellent ancestry, the transistor appears to have a bright future. Its compactness, light weight, long operating life, low power consumption, novel circuitry, and other factors suggest future accomplishments of value to mankind.

At this time, it is of course impossible to predict what further devices and principles will be evolved from the transistor at the next turn of the wheel of scientific and technical history, nor yet what new discoveries and instrumentalities will be needed to open the doors to such later evolution. One can be confident only that there are many remaining treasures in nature's storehouses awaiting the enterprise of newer pioneers of a later day.

—The Editor.

Transistors in Our Civilian Economy

J. W. McRAE

At relatively long intervals there appear on the technical and industrial horizons devices of such broad scope and major significance that they profoundly affect the fields of their use. One of these epochal developments is the transistor, which bids fair to take its place beside the electron tube as one of the foundation stones of future communications and electronics.

It is accordingly timely and suitable that certain of the probable future industrial uses and effects of the transistor should be here analyzed in a guest editorial by an engineer especially qualified for this task, and who is a Fellow and Director of the Institute, and a Vice President of Bell Telephone Laboratories.—*The Editor.*

IT IS NOW a little more than four years since transistors were first announced. The youngest members of the new family, junction transistors, are not yet two years old. What can be said, at this early stage in their history, about what transistors will mean to our civilian economy and to IRE members?

Vacuum tubes had been in existence about eight years when they had their first dramatic successes as amplifiers and oscillators. Thirty-seven years ago, in 1915, they made possible the first very long distance transmission of speech, by wire line across this continent and by radio across the Atlantic. No one could then have predicted the present situation. It would have been impossible to conceive of a production of close to half a billion vacuum tubes a year or an IRE membership exceeding thirty thousand.

The fact is that vacuum tubes not only performed their original functions as detectors of radio waves; they made possible entirely new techniques for use in transmitting information. The scope of existing communications services was greatly expanded and whole new industries were brought into existence.

More recently, vacuum tubes have been applied, not only to transmit information, but also to manipulate it and put it directly to use. Electronic computers now work out intricate problems in the laboratory and solve the equations necessary to aim and fire anti-aircraft guns. Through servomechanisms, they aim the guns and control other operations involving considerable power with finesse and precision. All of this is unified with the field of information transmission through the use of similar devices and techniques as well as by a common set of theoretical principles.

Are transistors to be merely competitors for these jobs now being done by vacuum tubes? Or can they, in their turn, further extend the field of electronics? It is most important to examine this second question. There will undoubtedly be many cases in which transistors will be able, by replacing vacuum tubes, to reduce the cost and increase the convenience or effectiveness of existing services. But the benefits will be greater if transistors can penetrate into portions of our present services where vacuum tubes are not yet widely used. It will be still better if some new area of electronics can so capitalize on transistors as to be able to expand significantly.

Two examples will serve to illustrate the fact that both possibilities have substance. The first is our local telephone system in which there are essentially no vacuum tubes. The only widely used amplification between the subscriber and the central office is provided by the carbon microphone or telephone transmitter. It supplies the necessary gain, is inexpensive and reliable, and conveniently requires only a small direct current to supply its power. Because of its reliability it can be located at many scattered points without incurring excessive maintenance costs. In the great majority of cases, its power requirements can be satisfied by transmission of power from the central office, so there is no necessity for relatively expensive local sources of dependable power.

Power requirements should be considered in terms of the power levels of the signals to be handled. A good many situations in communications require that the signals be brought to high power levels. But in this particular situation one microwatt of signal power produces readily audible sound in a telephone receiver. This is thousands of times smaller than the power required just to heat the cathode in one of the smallest commercial vacuum tubes. It is a million times smaller than the cathode heater power in most of the vacuum tubes designed for suitable performance, long life, and reliability in telephone service. These vacuum tubes also require comparable amounts of plate power, usually at voltages and currents difficult to supply over local telephone lines.

The transistor promises to combine long life and reliability with the ability to amplify very small signals at high efficiency. It should prove possible to supply the small amount of necessary power over the line. Transistors may be able therefore to extend electronics into this area; which the vacuum tube has been unable to do.

Undoubtedly there are other similar situations where the need is for the *transmission* of information; but consider now the second illustration, computers and other automata which *manipulate* information. Vacuum tubes made possible electronic analogue computers and in such computers have met with little competition from other devices. On the other hand, they have met stiff competition from electromechanical devices such as relays, in digital computers, and in other automatic machines or systems which work by digital methods. Automatic business machines and telephone switching systems are examples. With few exceptions, vacuum tubes have been used only when their great speed overcame their disadvantages in cost and reliability.

These applications generally involve the use of large numbers of individual devices such as relays, switches, or vacuum tubes. Questions of the reliability and the power consumption of each individual device therefore loom large. Here again the operations to be performed can for the most part be done at very low power levels. The power required to operate relays or heat the cathodes of vacuum tubes is many times larger than the needed power of signals to be handled in each device.

Transistors and their ancestors, germanium diodes, appear to make a natural team for such applications.

They ought to make it possible to close somewhat the great gap that now exists between our best electronic "brains" and the human brain, in regard to size and power consumption. At the same time, they promise better reliability and speed comparable to vacuum tubes. The future may well see a considerable expansion in the industrial use of complex automata. Perhaps it is significant that the first commercial use of transistors is expected to be in a part of the "brain" of the new telephone national toll dialing system.

These two illustrations demonstrate the possibilities for extensions in electronics as the result of the advent of transistors. However, it must not be assumed that the engineering job ahead will be easy. Transistors are not vacuum tubes, and many of our present circuit design practices, developed with vacuum tubes in mind, will not be applicable to transistors. It will often be necessary to analyze these new circuit problems in terms of those fundamental principles which have more general application.

Those directly concerned with the design and manufacture of transistors will get little assistance from the lore of vacuum tubes. They will have to turn to the fundamental physics of conduction in semiconductors. While this may be unfamiliar territory, they will find it well charted. Indeed, our present physical understanding of semiconductor devices is remarkably mature, very much better than our corresponding understanding of vacuum tubes in 1915. This gives us a good basis for confidence in future progress and it should help us to accelerate that progress.

Transistors and the Military

I. R. OBENCHAIN, Jr. and W. J. GALLOWAY

THE TRANSISTOR, and the unquestionable impact it may exert on electronic design as we know it today, have rapidly become major centers of interest. This impact will be important from two points of view: engineering and economics.

Having at least some bearing on both are questions of transistor application in large-scale national rearmament. For example: Exactly what are the military services doing about transistors? Can the Army use them for a wrist-watch radio? Can transistors substantially reduce the weight of electronic equipment in the Air Force's bombers? Will the Navy equip torpedoes with "transistorized" electronic brains?

Answers to the last questions necessarily stem from the results of service transistor programs, but the first is readily answered since military interest in this field has been close and continuous since 1948. Due to the difficulty of quantity production of transistors with uniform characteristics, immediate application to military equipments was impractical up to the spring of 1951. Evolution of the junction transistor in 1951, plus quantity production ability with uniform characteristics for both these new type low-noise units as well as the older point contact types, changed the previous somewhat restricted approach of the military services to an active and expanding program. This uniformity of performance is of course a necessity for military use, and consequently, the services are interested only in transistors that can be built to meet prescribed specifications.

Thereupon, the military services embarked upon substantial transistor programs in the summer and fall of 1951, which emphasized continued solid-state research, specific research to improve transistors, circuitry research for application of transistors, and the preparation of facilities for transistor production. The eventual impact of transistors on military electronic equipments was considered sufficient to form within the Defense Department's Research and Development Board an Ad Hoc Group on Transistors. This Ad Hoc Group was established under the Committee on Electronics as a high-level group to set broad policies for the co-ordination of transistor programs within the three services. Shortly afterwards a Subpanel on Semiconductor Devices was formed under the RDB's Panel on Electron Tubes. This latter group performs the normal subpanel functions of co-ordinating service projects and then passing them on to the Ad Hoc Group for final approval. On eventual dissolution of the Ad Hoc Group the Semiconductor Subpanel will take its normal position with respect to the Panel on Electron Tubes such as is now held by the other subpanels of that group—RF tubes, receiving tubes, and so on.

To speed active interest for continued research and development work on transistors and related devices (for instance, barrier rectifiers, and the like) including investigations of the use of materials other than germanium, the military services are presently supporting substantial contractual programs. Most of the contracts are joint service (equal support from the Army, Navy, and Air Force) in nature, with the services alternating as administrators for each new contract. The military are also aware of and intensely interested in the large amount of work being applied to this problem by industry without government support. A modest program in this general field is maintained on an internal basis within the various military research and development laboratories. It is hoped that this broad approach, especially the nongovernment supported work of private industry and educational institutions and the effort under special research and development contracts supported by military funds, will soon provide transistors with higher power ratings, higher frequency response, lower noise, and the ability to operate satisfactorily over wide ranges of temperature.

"Battle is the Payoff" is a slogan within the military. This thought is applied to the service transistor programs in determining where and how these new units may be used to the maximum advantage in military

The transistor promises to combine long life and reliability with the ability to amplify very small signals at high efficiency. It should prove possible to supply the small amount of necessary power over the line. Transistors may be able therefore to extend electronics into this area; which the vacuum tube has been unable to do.

Undoubtedly there are other similar situations where the need is for the *transmission* of information; but consider now the second illustration, computers and other automata which *manipulate* information. Vacuum tubes made possible electronic analogue computers and in such computers have met with little competition from other devices. On the other hand, they have met stiff competition from electromechanical devices such as relays, in digital computers, and in other automatic machines or systems which work by digital methods. Automatic business machines and telephone switching systems are examples. With few exceptions, vacuum tubes have been used only when their great speed overcame their disadvantages in cost and reliability.

These applications generally involve the use of large numbers of individual devices such as relays, switches, or vacuum tubes. Questions of the reliability and the power consumption of each individual device therefore loom large. Here again the operations to be performed can for the most part be done at very low power levels. The power required to operate relays or heat the cathodes of vacuum tubes is many times larger than the needed power of signals to be handled in each device.

Transistors and their ancestors, germanium diodes, appear to make a natural team for such applications.

They ought to make it possible to close somewhat the great gap that now exists between our best electronic "brains" and the human brain, in regard to size and power consumption. At the same time, they promise better reliability and speed comparable to vacuum tubes. The future may well see a considerable expansion in the industrial use of complex automata. Perhaps it is significant that the first commercial use of transistors is expected to be in a part of the "brain" of the new telephone national toll dialing system.

These two illustrations demonstrate the possibilities for extensions in electronics as the result of the advent of transistors. However, it must not be assumed that the engineering job ahead will be easy. Transistors are not vacuum tubes, and many of our present circuit design practices, developed with vacuum tubes in mind, will not be applicable to transistors. It will often be necessary to analyze these new circuit problems in terms of those fundamental principles which have more general application.

Those directly concerned with the design and manufacture of transistors will get little assistance from the lore of vacuum tubes. They will have to turn to the fundamental physics of conduction in semiconductors. While this may be unfamiliar territory, they will find it well charted. Indeed, our present physical understanding of semiconductor devices is remarkably mature, very much better than our corresponding understanding of vacuum tubes in 1915. This gives us a good basis for confidence in future progress and it should help us to accelerate that progress.

Transistors and the Military

I. R. OBENCHAIN, JR. AND W. J. GALLOWAY

THE TRANSISTOR, and the unquestionable impact it may exert on electronic design as we know it today, have rapidly become major centers of interest. This impact will be important from two points of view: engineering and economics.

Having at least some bearing on both are questions of transistor application in large-scale national rearmament. For example: Exactly what are the military services doing about transistors? Can the Army use them for a wrist-watch radio? Can transistors substantially reduce the weight of electronic equipment in the Air Force's bombers? Will the Navy equip torpedoes with "transistorized" electronic brains?

Answers to the last questions necessarily stem from the results of service transistor programs, but the first is readily answered since military interest in this field has been close and continuous since 1948. Due to the difficulty of quantity production of transistors with uniform characteristics, immediate application to military equipments was impractical up to the spring of 1951. Evolution of the junction transistor in 1951, plus quantity production ability with uniform characteristics for both these new type low-noise units as well as the older point contact types, changed the previous somewhat restricted approach of the military services to an active and expanding program. This uniformity of performance is of course a necessity for military use, and consequently, the services are interested only in transistors that can be built to meet prescribed specifications.

Thereupon, the military services embarked upon substantial transistor programs in the summer and fall of 1951, which emphasized continued solid-state research, specific research to improve transistors, circuitry research for application of transistors, and the preparation of facilities for transistor production. The eventual impact of transistors on military electronic equipments was considered sufficient to form within the Defense Department's Research and Development Board an Ad Hoc Group on Transistors. This Ad Hoc Group was established under the Committee on Electronics as a high-level group to set broad policies for the co-ordination of transistor programs within the three services. Shortly afterwards a Subpanel on Semiconductor Devices was formed under the RDB's Panel on Electron Tubes. This latter group performs the normal subpanel functions of co-ordinating service projects and then passing them on to the Ad Hoc Group for final approval. On eventual dissolution of the Ad Hoc Group the Semiconductor Subpanel will take its normal position with respect to the Panel on Electron Tubes such as is now held by the other subpanels of that group—RF tubes, receiving tubes, and so on.

To speed active interest for continued research and development work on transistors and related devices (for instance, barrier rectifiers, and the like) including investigations of the use of materials other than germanium, the military services are presently supporting substantial contractual programs. Most of the contracts are joint service (equal support from the Army, Navy, and Air Force) in nature, with the services alternating as administrators for each new contract. The military are also aware of and intensely interested in the large amount of work being applied to this problem by industry without government support. A modest program in this general field is maintained on an internal basis within the various military research and development laboratories. It is hoped that this broad approach, especially the nongovernment supported work of private industry and educational institutions and the effort under special research and development contracts supported by military funds, will soon provide transistors with higher power ratings, higher frequency response, lower noise, and the ability to operate satisfactorily over wide ranges of temperature.

"Battle is the Payoff" is a slogan within the military. This thought is applied to the service transistor programs in determining where and how these new units may be used to the maximum advantage in military

equipments. The favorable characteristics of transistors, namely, no requirement for heater operation, high efficiency, resistance to shock, long life, and small size have already been widely publicized but are here re-emphasized as exceptionally desirable characteristics for active elements to be included in military equipments. The possible savings from a power standpoint alone are impressive when one remembers the considerable quantity of batteries required for hand-carried equipments during World War II—for example, the thirty-day supply of batteries for a certain type of amphibious communications company totaled forty-three tons. In general, the military services' transistor applications programs are aimed at improving existing types of equipments, making them lighter or more reliable, or providing new types of equipments which with vacuum tubes had not been considered feasible for military use. An example of the latter type of development would be small compact "transistorized" computers in which the use of several hundred vacuum tubes would have been impractical from a size and power consumption viewpoint. These programs are being conducted on a contractual as well as on an internal basis within government laboratories and will be greatly expanded during fiscal year 1953, as more transistors become available for the services to furnish to contractors working on application studies. Availability of transistors, however, is not the only problem; applications experience has confirmed the difficulties that circuitry engineers were expected to encounter in utilizing these new devices. There is a whole field of engineering circuitry data that is yet to be obtained for each of the general transistor types, and apparently the greatest results in the treatment of this problem can be obtained by the well-known methods of network analysis. Another typical problem associated with the application side of the program is how to make a small, economical battery source of constant current and long life. The large variation with temperature of the present type transistors' equivalent circuit parameters is also of particular difficulty, since so many military equipments require operation over extended temperature ranges.

Associated with the circuitry application programs, emphasis is being placed on the development of circuit packages wherein complete circuit functions are "potted" into plug-in units. Such items as oscillators, counting circuits, IF amplifiers, and so on, fall in this category. Investigations are also being conducted towards developing smaller passive components which will take advantage of the lower voltages utilized in transistor circuits to provide a minimum size component.

In more normal times the military services would embark on only a modest program of "transistorization" leaving the broad general problem of the maximum utilization of these devices to the ingenuity of our industry and research institutions. Now, however, in this period of international tension the services consider the possible benefits of transistors to military equipments as sufficient to warrant substantial programs in this field and to include concurrently not only research and development on the transistor itself and its application to equipments, but the planning and preparation of facilities for producing large quantities of these devices. The responsibility within the Department of Defense for planning these transistor production facilities has been delegated to the Army Signal Corps, with top policy guidance from the RDB Ad Hoc Group on Transistors. This program includes contracting for the establishment of factories, industrial mobilization study contracts, consideration of raw material availability, and the placement of supply orders. The concurrent establishment of production facilities and enlarged device research and development programs as well as increasing effort on applications faces one with the very real problem of what transistor types to prepare for in quantity.

The military are in fact now faced with the problem of preparing a filling feast without the unnecessary expense of a smorgasbord for very fussy guests while having only the slightest indication of their preference.

Fortunately, there is reason to believe that transistor factories can vary types of transistors produced without prohibitively costly modifications to capital equipment; however, there is still no solution to the use of stockpiles which could suddenly become outdated by the overnight discovery of a radical improvement.

It is hoped this provides at least some insight to the military's rather substantial interest in transistors. In present form, transistors will undoubtedly rather quickly fill a number of military requirements and will also provide devices for training of personnel engaged in circuitry studies. As they improve it is certain that additional applications will follow with increasing rapidity. With transistor pilot-production lines now starting to roll, the immediate need is to develop a new circuitry "know-how" which for electron tubes, took over a quarter century in its evolution toward current engineering practices. The military services will do everything within their means to compress the new circuitry and application studies into the smallest time interval possible, and in so doing will look largely to the ingenuity of the IRE membership for the assistance to realize the maximum utilization of new devices.

Transistor Electronics: Imperfections, Unipolar and Analog Transistors*

W. SHOCKLEY†

Summary—The electronic mechanisms that are of chief interest in transistor electronics are discussed from the point of view of solid-state physics. The important concepts of holes, electrons, donors, acceptors, and deathnium (recombination center for holes and electrons) are treated from a unified viewpoint as imperfections in a nearly perfect crystal. The behavior of an excess electron as a negative particle moving with random thermal motion and drifting in an electric field is described in detail. A hole is similar to an electron in all regards save sign of charge. Some fundamental experiments have been performed with transistor techniques and exhibit clearly the behavior of holes and electrons. The interactions of holes, electrons, donors, acceptors, and deathnium give rise to the properties of *p-n* junctions, *p-n* junction transistors, and Zener diodes. Point-contact transistors are not understood as well from a fundamental viewpoint. A new class of *unipolar* transistors is discussed. Of these, the *analog* transistor is described in terms of analogy to a vacuum tube.

1. Perfection and Imperfections in Crystals

TRANSISTOR ELECTRONICS exists because of the controlled presence of imperfections in otherwise nearly perfect crystals.[1] Many of the special words of the subject are the names of imperfections; hole, excess electron, and donor are examples.

In this article we shall describe these imperfections from the point of view of theory; again from the point of view of theory we shall show how the interplay of imperfections permits useful electronic devices to exist. We shall also indicate how the new experiments of transistor physics display the attributes of the imperfections and lead to measurements of their quantitative attributes.

Before discussing imperfections in a germanium crystal we must discuss perfection. Perfection is represented by an ideal crystal built on the plan of Fig. 1. Each atom has four neighbors, all at the same distance from it and all at equal distances from each other.[2] This arrangement satisfies perfectly the chemical affinity of the germanium atom.

For the purpose of transistor physics the germanium atom may be considered to consist of four valence electrons surrounding an inert core having a charge of plus four electronic units. The atom as a whole is thus neutral. Silicon and carbon atoms have this same structure and their crystals form in the same pattern as Fig. 1.

[1] The subject of imperfections has recently been reviewed in a book, "Imperfections in Nearly Perfect Crystals," Editorial Committee, W. Shockley, Chairman; J. H. Hollomon, R. Maurer, F. Seitz; pub. by John Wiley and Sons, New York, N. Y.; 1952.
[2] The reader with an instinct for spatial visualization may enjoy thinking about the problem of surrounding one atom with several symmetrically placed neighbors. He will find that four is the largest number of neighbors that can surround one atom at equal distances and be equidistant from each other.

Fig. 1—Diamond structure.

The core of the germanium atom consists of the nucleus, whose charge is plus 32 electron units and 28 tightly bound electrons. These latter are closely held to the nucleus, as represented in Fig. 2, and do not enter into chemical reactions or transistor effects.
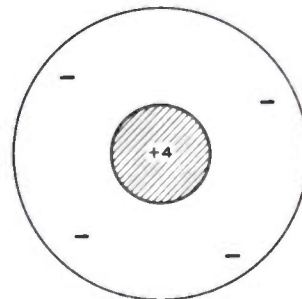
Fig. 2—The four valence electrons and positive core of a silicon or germanium atom.

The outer valence electrons are responsible for chemical binding. One of the principal building blocks of chemistry is the *electron-pair bond* or *covalent bond*. Such a bond may form when two atoms approach each other. In order to form the bond a pair of electrons, one from each atom, co-ordinate their motions so as to produce an accumulation of negative charge between the cores. This attracts the cores towards each other. Equilibrium is achieved when the attraction is balanced by electrostatic repulsion of the cores for each other.

The laws of physics which explain the behavior of the electron-pair bond are *quantum* laws and employ *wave-mechanics* to describe the motions of the electrons. These

laws explain why two electrons, rather than one or three, unite in forming the most stable bonds. We shall need to refer to this feature of the stability subsequently.

The germanium crystal gives each atom an ideal opportunity to form electron-pair bonds. Each atom is surrounded by four symmetrically placed neighbors. It forms one covalent bond with each of these. The electrons in each bond are shared equally by the two atoms at its ends, one electron coming from each atom.

Fig. 3 symbolically represents the bond structure of the crystal. In it the atoms lie in a plane rather than in space, but each atom is correctly surrounded by four neighbors. A crystal in which this bond structure is complete and every atom is at rest in its proper place, with no atoms or electrons missing or squeezed into wrong places, is in the *perfect* or reference state.

TABLE I
CLASSES OF IMPERFECTIONS

Energetic—photons (light), phonons (heat)
Electronic—(excess) electrons, holes
Atomic—disordered atoms (vacancies and interstitial atoms)
   chemical impurities (donors and acceptors)
   "deathnium"
   "traps"

The photon is the name for a quantum of light energy. This energy comes in packets of size

$$\text{energy} = h\nu, \tag{1}$$

where h is Planck's constant and ν is the frequency of the light. When light is absorbed, the energy is delivered in entire quanta.

Fig. 3—Representation of electron-pair bond structure in silicon or germanium.
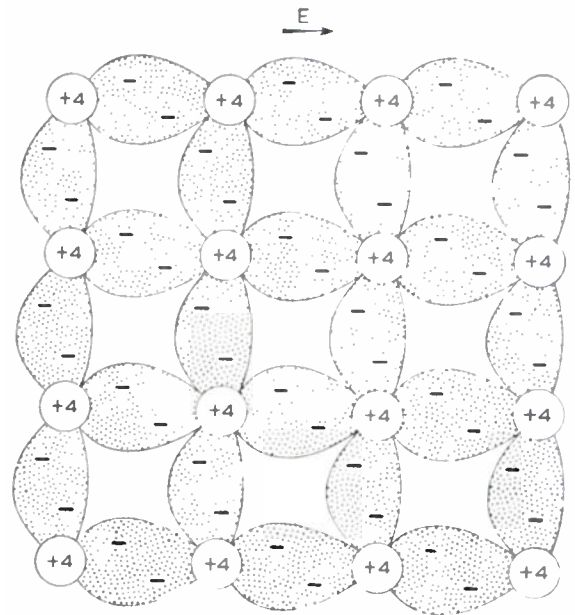
Fig. 4—Distortion produced by an electric field resulting in dielectric polarization.

The perfect state is relatively uninteresting from the transistor point of view. A germanium crystal in its perfect state would be an insulator. Its dielectric constant, however, would have the unusually high value of 16. The effect of an electric field in producing polarization is represented in Fig. 4. Since the electric field exerts forces in one direction upon the cores and in the opposite direction on the bonding electrons, the structure is distorted. There is evidently a relative shift to the right of positive charge in respect to negative charge. This displacement is linear and elastic in the electric field and disappears when the field is removed. (Fig. 4 shows an exaggerated case corresponding to a field of about 10^7 volts/cm, much more than can be applied without breakdown effects.)

On the basis of perfection as represented by Figs. 1 and 3, the imperfections of importance in transistor electronics can now be described. In Table I these imperfections are listed under three classifications.

As is implied by their names, the energetic imperfections involve specialized concepts from quantum theory.

Less generally familiar is the fact that mechanical vibrations of solids are also quantized. A beam of sound waves transmitted through a solid is composed of phonons just as a light beam is of photons. Just as a cavity is full of *black body radiation* under thermal equilibrium, so is a crystal full of a thermal equilibrium distribution of phonons. The phonons thus represent thermal vibrations and account for the specific heat of the solid. We shall be concerned with the phonons because of the way in which they interfere with electronic motion in the crystal. This is discussed more fully in the next section.

The other classes of imperfections involve particles. In the electronic class there are two possibilities, corresponding to adding and subtracting an electron. These are discussed in Section 2.

In the atomic class, it is also possible to add and subtract atoms of germanium from a germanium crystal. There are also other ways of disordering the crystal without introducing atoms of different kinds. We shall not deal with these here. The foreign atoms of chief interest are donors and acceptors, which come from columns of the periodic table adjacent to the germanium column.

Two classes of imperfection are shown described by their properties rather than by their nature. It is known that both the action of *deathnium* and that of *traps* can be accomplished by disorder. The same effects may be accomplished by impurities. These atomic imperfections are discussed in Section 3.

One important class of imperfection that has been omitted from Table I is the *grain boundary*. A grain boundary occurs when two crystals, each having a structure like Fig. 1, are tilted in respect to each other so that their axes are not parallel. Where these two crystal *grains* join, there is a region of atomic misfit called a "grain boundary." Grain boundaries have pronounced electrical effects in germanium, and are known in metallurgy to be loci for the precipitation of impurities.[3]

Grain boundaries are not considered in this article because they can be eliminated. As a result of stimulation from the desires of the transistor development group for highly uniform bulk material Teal and Little undertook a program of growing large single crystals of germanium.[4] These single crystals, weighing up to several hundred grams, have the same orientation throughout and no grain boundaries. For the last few years, practically all advances at Bell Telephone Laboratories in transistor electronics and transistor physics have been based on the availability of single-crystal material.

The importance of artificially grown single crystals in transistor work is not unique. The same situation holds true in all branches of solid-state physics; magnetism, plasticity of metals, piezoelectricity, and the study of order-disorder transformations in alloys are examples. In these cases, the availability of artificial single crystals has had a very marked stimulus upon the development of the science and the technology.

In this article emphasis is placed chiefly upon germanium with silicon also receiving some attention. This is appropriate in view of the current status of development. It should be mentioned, however, that these elements are not unique. Transistors have also been made using lead sulfide (galena)[5] and lead selenide.[6]

2. The Behavior of Holes and Electrons

A. Introduction

In this section we shall be concerned chiefly with the behavior of electronic imperfections. These imperfections play the essential role of current carriers in tran-

sistor electronics. A large part of this section is devoted to the interactions between electronic imperfections and the energetic imperfections mentioned in Section 1. It turns out to be convenient to include, at the same time, a discussion of deathnium, although it is not an electronic imperfection.

As most readers of this article probably know, the two important electronic imperfections are the *excess electron* and the *hole*. These are produced by adding and subtracting, respectively, one electron from the perfect situation. The behavior of the *excess electron*, or simply *electron* for short, can be understood relatively simply from an intuitive approach. This is not true for the hole. In fact, intuition leads to quite erroneous conclusions regarding the hole. In this section, the electron is, therefore, discussed first. Once the behavior of the electron has been described, that of the hole can be readily explained, even though the reasons for the behavior cannot. It is the author's opinion, based on several years' experience with these expositional problems, that the theoretical basis for the behavior of the hole cannot be simplified and presented on an intuitional level. Fortunately, the behavior itself is relatively simple and can be verified by direct experiment. The engineer who wishes to think in terms of holes may thus have confidence in the concepts as experimentally established engineering factors.

B. The Behavior of an Electron[7]

The simplest electronic imperfection is produced by adding one extra electron to the otherwise perfect crystal. From the electrical point of view this extra electron represents one negative unit of charge. The electric field set up by this charge is, of course, screened by the dielectric constant of the germanium, which has a value of 16. Thus the force which the electron will exert upon another charge is reduced by the effect of the medium upon its electric field. If we suppose that the energetic disturbances in the crystal are present in negligible proportions, then the behavior of the added electron is very simple. If it is added to the crystal with zero kinetic energy, it will stay substantially in the same place indefinitely. It is not bound to one place, however, as is an electron in a valence bond. The high stability of an electron-pair bond is restricted to two electrons. A third electron cannot fit into this structure and, consequently, the excess electron is free to move. If an electric field is applied, the excess electron will be accelerated by the resulting force. The acceleration is affected by the fluctuating electric fields due to the positively charged atomic cores and the negative valence bonds. For most purposes the effect of this potential, which varies periodically in space, can be lumped into one quantity, the *effective mass* of the electron. Under the influence of a given force the electron in the crystal will have an acceleration different from that which it would have in free space under the influence of the same force.

[3] G. L. Pearson, "Electrical properties of crystal grain boundaries in germanium," *Phys. Rev.*, vol. 76, p. 459; 1949.
N. H. Odell and H. Y. Fan, "Impedance characteristics of grain boundaries in high resistivity *n*-type germanium," and W. E. Taylor and H. Y. Fan, "D.C. Characteristics of high resistance barriers at crystal boundaries in germanium," *Phys. Rev.*, vol. 78, p. 354; 1950. See, also, footnote reference 1 for a review of the theory of grain boundaries.
[4] G. K. Teal and J. B. Little, "Growth of germanium single crystals," *Phys. Rev.*, vol. 78, p. 647; 1950.
G. K. Teal, M. Sparks, and E. Buehler, Proc. I.R.E., vol. 40, pp. 906–909; August, 1952.
[5] H. A. Gebbie, P. C. Banbury, and C. A. Hogarth, "Crystal diode and triode action in lead sulphide," *Proc. Phys. Soc.* (London), vol. B63, p. 371; 1950.
[6] A. F. Gibson, "Absorption spectra of single crystals of lead sulphide, selenide and telluride," *Proc. Phys. Soc.* (London), vol. 65 3, pp. 378; 1952.
P. C. Banbury, "Double surface lead sulphide transistor," p. 236.

[7] For a more complete treatment and references, see W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y.; 1950.

The result can be interpreted by saying that in the crystal the mass of the electron is in effect different from what it would be in free space. Experiments indicate that the effective mass of an electron in a crystal is about as large as the mass in free space and may be less by a factor of four in the case of semiconductors of interest to transistor electronics.[8]

The reader unfamiliar with the background of the modern energy-band theory of solids probably finds it surprising that the electron can move through the strong fluctuating electric fields in a crystal as though it were in free space. The reason that it can do this is that the motion of the electron is governed by wave mechanical laws. It is a property of wave propagation that waves can proceed unattenuated through lossless structures which have perfect periodicity. Examples of iterated electrical filters and loaded lines will remind the reader of this fact. A more striking example from electromagnetic theory is that of the artificially loaded dielectric which is used in lens systems for microwave antennas.[9] For the case of electrons in crystals, similar mathematical developments apply. It is found that the electron wave can cause the electron to propagate indefinitely through the crystal provided that the electrical potential in which it moves is perfectly periodic, as it will be if the crystal is in the perfect reference condition.

If energetic imperfections in the form of phonons are present, the perfect periodicity of a lattice will be destroyed. As a result, the electron once set in motion will not move indefinitely in one direction but will instead be scattered so that after a time its motion is changed to some other direction. The laws of scattering of electron waves by interference with phonons can be derived on the basis of quantum-mechanical theory. When the derivation is carried out, it is found that the effect of the phonons can be very accurately simulated by introducing a set of *phantom particles*.[10] These phantom particles are to be thought of as uncharged hard, spherical, ideally elastic masses. Phonons in a germanium crystal at room temperature correspond to phantom masses about 170 times that of the electron mass. Consequently, in any collision the electron does not lose much energy but bounces nearly elastically from the phantom particle. The choice of a spherical shape is dictated by the fact that scattering by phonons is isotropic, that is, after being scattered an electron is equally likely to proceed in any direction. The scattering of a light spherical mass by a much heavier spherical mass also has this isotropic property. The phantom particles should be thought of as moving with random thermal energies. The number of them and their cross sections are such that as an electron moves through the crystal it will, on the average, move for a distance of

about 10^{-5} cms between collisions with the phantom particles. We shall refer to this *mean free path* as l; we thus have

$$l \doteq 10^{-5} \text{ cm.} \tag{2}$$

This distance is approximately 1,000 times the interatomic distance between germanium atoms in the crystal.

If the electron is placed at rest in a germanium crystal at room temperature, then it will be jostled by the phantom particles which represent the phonons. As a result of this jostling, the electron will acquire thermal energy. A general theorem from statistical mechanics applies to this situation. This theorem, known as the principle of equipartition of energy, states that the average kinetic energy of a free particle at temperature T is given by the formula

$$\text{thermal kinetic energy} = (3/2)kT, \tag{3}$$

where T is the absolute temperature in degrees K and k is Boltzmann's constant. Since the electron has only about 1/170 the effective mass of a phantom particle, its speed will be 13 times greater. If we take the electron's mass as equal to the mass of a free electron and define a thermal velocity by the equation

$$mv_l^2/2 = kT, \tag{4a}$$

then we find that at room temperature the thermal velocity of an electron is

$$v_r \doteq 10^7 \text{ cm/sec.} \tag{4b}$$

In terms of the mean free path l and the thermal velocity v_T, we can define a *mean free time* τ;

$$\tau = l/v_r \doteq 10^{-12} \text{ sec.}$$

This is the average time during which an electron moves between collisions.

One of the convenient units of energy to use for describing electrons is the electron volt. This is the energy which an electron would acquire in falling through a potential difference of one volt. At room temperature T is about 300° on the Kelvin scale and

$$kT = 0.026 \text{ electron volts.} \tag{5}$$

We shall see below that the voltage which gives an electron thermal energy plays a critical role in the nonlinear behavior of p-n junctions and p-n junction transistors.

The model of electron collisions with phantom particles represents quite accurately the behavior of an electron interacting with the phonons. In particular it correctly represents the ability of electrons to interchange energy with the phonons. This interchange of energy is a very important process which occurs whenever an electrical field is applied to a semiconductor. Under these conditions electrons acquire energy from the electric field which does work upon them. In order for this energy to be converted to heat in the crystal it is necessary for it to be transferred to the thermal vibrations or phonons. From our model of the phantom particles it is evident that if the electrons acquire a greater velocity

[8] E. M. Conwell, "Properties of silicon and germanium," PROC. I.R.E., vol. 40, pp. 1327–1337; this issue.
[9] W. E. Kock, "Metallic delay lenses," *Bell Sys. Tech. Jour.*, vol. 27, pp. 58–82; 1948.
W. E. Kock and F. K. Harvey, "Refracting sound waves," *Jour. Acous. Soc. Amer.*, vol. 21, pp. 471–481; 1949.
[10] W. Shockley, "Hot electrons in Germanium and Ohm's law," *Bell Sys. Tech. Jour.*, vol. 30, p. 990; 1951.

than their normal thermal velocity, then upon collisions with the phantom particles they will lose their excess energy. A mathematical analysis of this process can be carried out in order to find out at what critical field Ohm's Law will fail. We shall return to this point briefly after discussing the mobility of the electrons below.

Diffusion: The behavior of an electron in the absence of an electric field is represented crudely in Fig. 5. Its path consists of a number of short straight line segments. On the average, each one of the segments has a length of l. At the end of each segment the electron has a collision
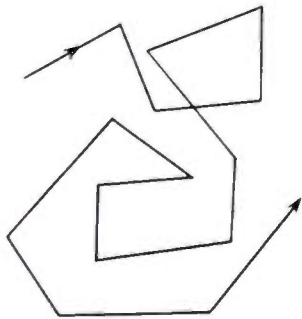
Fig. 5—The random diffusive motion of an excess electron.

with a phantom particle and starts off with approximately the same thermal velocity but in a completely random direction. As a result, it *diffuses* about in the crystal. The process of diffusion is a basic one in transistor electronics and deserves special discussion at this point. It is difficult, however, to explain the process of diffusion in terms of the motion of an individual electron. For this reason we shall consider a large number of electrons so that we can deal with their average statistical behavior. If a large number of electrons were actually present in a crystal which contained no other electronic or chemical imperfections, then they would strongly influence each other by their electrostatic repulsion. There are certain experimental conditions, which we shall discuss in Section 4, in which the effect of this electrostatic repulsion is suppressed by other factors. It would greatly prolong the argument to consider these other factors here. In order to proceed with the discussion of diffusion, therefore, we shall simply assume for the moment that the electrostatic interaction between the electrons is so small that we can neglect it in comparison with the diffusive motion.
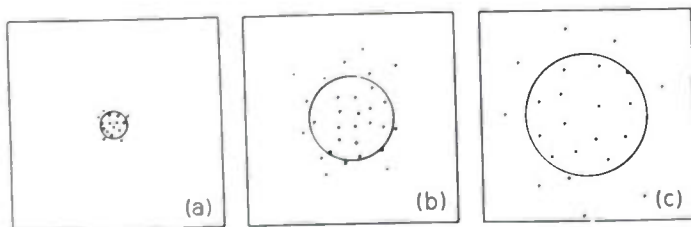
Fig. 6—The spreading of a group of electrons by diffusion.

The quantity which characterizes diffusion is the *diffusion constant D*. It has the dimensions of cm^2/sec. We shall illustrate with two examples how diffusion processes are governed by the diffusion coefficient and shall also show how the diffusion coefficient is related to the

microscopic quantities l and v_T. The first example is represented in Fig. 6. We imagine that a group of N electrons are all contained inside a very small sphere at $t = 0$. Then at later times, the random diffusive motion causes them to spread out. Their distribution in space at time t is then given by the formula, which we introduce just for this example,

$$n(r, t) = N(\pi Dt)^{-3/2} \exp(-r^2/4Dt), \qquad (6)$$

where r is the distance from the initial position of the electrons and $n(r, t)$ is the number of electrons per cm^3 at r at time t. The most important term is the exponential. It shows that at any instant of time the density of the electrons is a *three-dimensional error function*. The density is a maximum at the origin, and decreases by a factor of $\exp(-1) = 0.37$ at $r = 2(Dt)^{1/2}$ and by a factor of $\exp(-4) = 0.019$ at twice that distance. In effect, the electrons are contained in a sphere of radius slightly larger than $2(Dt)^{1/2}$. One way of describing the size of the distribution is to consider the average value of r^2 for the electrons. This is found to be

$$\text{average of } r^2 = 6Dt. \qquad (7)$$

The second example is shown in Fig. 7. In this case the electron density is independent of the y and z coordinates and increases with increasing x. The rate of increase is called the gradient of the electron density
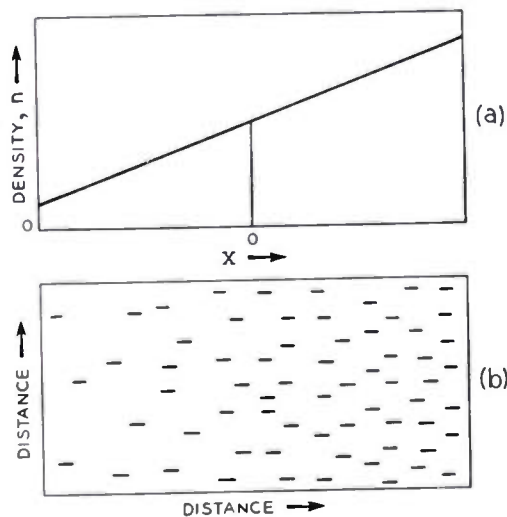
Fig. 7—A distribution of electrons with a density gradient.

and is the slope of the line in Fig. 7(a). Part (b) suggests the nature of the corresponding distribution of electrons in space. As these electrons diffuse, they tend to smooth out their distribution, with the result that a diffusion current of electrons will flow to the left in the figure. The strength of this current is

electrons per unit area per unit time

$$= D \text{ times density gradient.} \qquad (8)$$

We shall discuss below the way in which D may be derived from l and v_T on the basis of the situation represented in Fig. 7. There is, of course, only one diffusion constant, and it can be shown that either of the two equations (7) and (8) may be derived from the other.

In order to deal with the situation of Fig. 6 from a microscopic viewpoint we shall imagine that a group of N electrons are contained within a small sphere at time $t=0$. As a result of thermal velocities and collisions they spread out in a random fashion. In order to illustrate how this spreading results in (7) we shall consider what happens in one mean-free time. During this time each electron travels one mean-free path or, more accurately, parts of several mean-free paths. For purposes of this exposition, however, no important error will be made by assuming that all the electrons start and conclude a mean-free path simultaneously. Now consider the group of electrons whose radial distance from the origin is r when they start a mean-free path. After a time τ, about half of these will have moved outward and will be approximately at radius $r+l$, and half will have moved inward to $r-l$. The new average of r^2 for these electrons will thus be

$$\text{new av of } r^2 = (1/2)[(r + l)^2 + (r - l)^2]$$
$$= r^2 + l^2. \tag{9}$$

This result is somewhat in error because the outward motion is not necessarily radial; when this effect is taken into account, it is found that the increase in average of r^2 for this group of electrons in time τ is

increase in average of r^2 is $l^2/3$.

Since this result is independent of r, it will be obtained for all groups of electrons. In a time t, such increases will occur t/τ times so that

increase in average r^2 in time $t = l^2 t/3\tau = (v_l l/3)t$. (10)

Thus we see, on the basis of this very crude analysis, that r^2 will increase linearly in time and that the coefficient is given in terms of the microscopic quantities l and v_T.

The spreading of a group of electrons is one manifestation of diffusion of interest in transistor electronics. Another is the diffusion current produced by a concentration gradient. For this purpose we suppose that the electrons are distributed nonuniformly, as in Fig. 7, so that the number per unit volume is n and

$$n = n_0 + ax. \tag{11}$$

The concentration gradient is then

$$dn/dx = a. \tag{12}$$

Since there are more electrons to the right than to the left of the plane at $x=0$, we would expect a diffusion current to flow to the left. The magnitude of this diffusion current may be estimated roughly as follows: In one mean-free time, half the electrons in the region from $x=0$ to $x=+l$ will cross the $x=0$ plane towards the left and half those in the region from $x=0$ to $x=-l$ will cross it to the right. Now half the number of electrons per unit area of the $x=0$ plane within a distance l to the right of $x=0$ is

$$(1/2)l \text{ times } (n_0 + al/2) \tag{13}$$

since the average density in this region is the density at $x=l/2$. A similar treatment for the electrons to the left

leads to an expression with $-al/2$ in place of $al/2$. From this it follows that the net unbalance of flow causes

$$(l/2)(n_0 + al/2) - (l/2)(n_0 - al/2) = al^2/2 \tag{14}$$

electrons to cross the plane at $x=0$ moving towards the left in time τ. Hence the electron flow is

electrons per unit area per unit time

$$= al^2/2\tau = (lv_l/2)a. \tag{15}$$

If correct averaging procedures are employed, it is found that the coefficient of a becomes identical with D found in connection with the increase in r^2. Hence the general result is

particles per unit area per unit time

$$= D \text{ times density gradient.} \tag{16}$$

The direction of flow is in the direction of lower concentration.

Mobility: The effect of an electric field upon a diffusing electron is indicated in Fig. 8. The electric field

(INTERVALS BETWEEN COLLISIONS SHOWN EQUAL FOR SIMPLICITY)
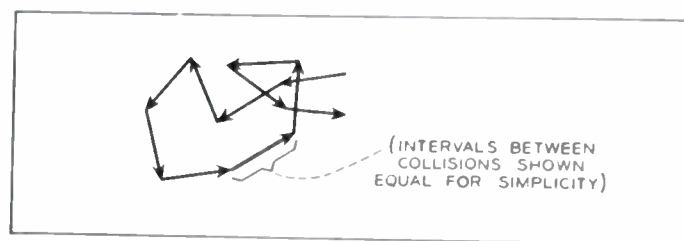
(a) RANDOM MOTION OF AN ELECTRON IN A CRYSTAL
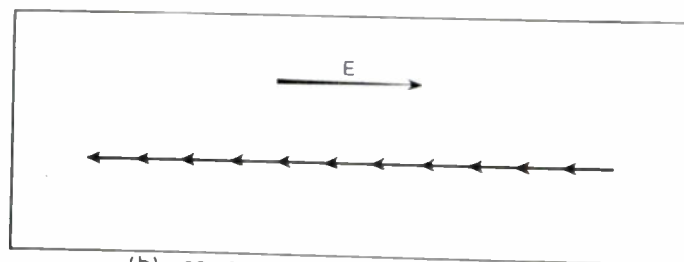
(b) COMPONENT OF ELECTRON MOTION ADDED BY ELECTRIC FIELD
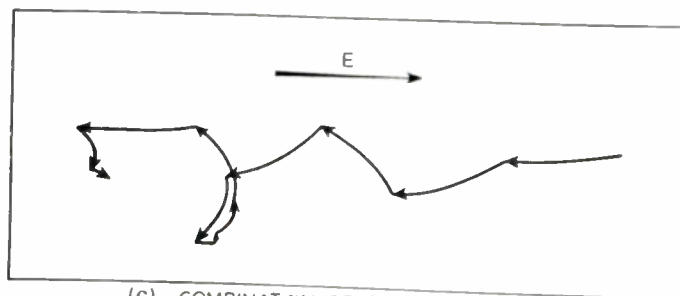
(c) COMBINATION OF (a) AND (b) ELECTRON MOTION IN AN ELECTRIC FIELD

Fig. 8—The effect of an electric field in superimposing drift on random motion.

which is shown as being directed towards the right in the figure exerts a force to the left upon the electron because of its negative charge. This force accelerates the electron with the result that the segments of the electron's path between collisions become curved rather than straight lines. The end points of the paths are consequently offset, as is indicated in part (c) of the figure. The offset produced by the electric field in each mean-free path can

be calculated from the falling-body formula. This formula is usually written in the form

$$S = (1/2)at^2. \tag{17}$$

The use of this equation to calculate the effect of the electric field makes use of the fact that the effect of an acceleration can be superimposed upon an already established velocity. (This is the theorem used in answering the question, "Where should you aim if the monkey drops when he sees the flash?") After a collision, the electron starts with a perfectly random velocity which will carry it one mean-free path in a time τ. During its motion it acquires additional velocity because of the force, and the additional motion is given by (17).

In order to apply (17) to our case we must take the acceleration as being equal to the force divided by the mass and must use the expression qE for the force where q represents the magnitude of the charge on an electron. This leads to the substitutions

$$a = F/m = qE/m \tag{18}$$

$$t = \tau. \tag{19}$$

The cumulative effect of the displacements given by equation (17) is represented in part (b) of Fig. 8. This results in a steady drift being superimposed on the random motion. Since one displacement S occurs in each interval τ of time, the average *drift velocity* is evidently given by the formula

$$v_d = S/\tau = (qE/m)\tau/2 = (q\tau/2m)E. \tag{20}$$

If a more sophisticated method of averaging is used which takes into account the fact that not all mean-free paths have exactly the same length, then it is found that a better value for the drift velocity is given by

$$v_d = (q\tau/m)E = \mu E. \tag{21}$$

In the last term of this equation we have introduced the symbol μ which is known as the *mobility*. It is the proportionality factor between electric field and drift velocity:

$$\mu = v_d/E = q\tau/m. \tag{22}$$

The linear relationship between electric field and drift velocity is valid so long as the electrons are able to dissipate the energy they acquire from the field sufficiently rapidly to the phonons.

If the electric field is too high, the electrons become "hot" in the sense that the energy of their random motion is increased. As a result they have collisions more frequently and their mobility is decreased. This effect gives rise to a fundamental deviation from Ohm's law in certain experiments and the resistance of a specimen increases with increasing electric field.[8,11] The critical field at which the mobility deviates appreciably from its low electric field value is simply related to the behavior of the phantom particles that simulate the phonons. When the drift velocity of the electrons equals the thermal

velocity of the phantom particles, then the linear relationship between drift velocity and electric field fails.[10]

Since both the diffusion constants and the mobility are determined by the microscopic quantities l, v_T and m, there should be a relationship between the two. This relationship can be exhibited in terms of the ratio

$$D/\mu = (1/3)lv_T/(q\tau/m) = mv_T{}^2/3q. \tag{23}$$

This equation is inaccurate because of the approximations made in evaluating D and μ. The correct equation is

$$D/\mu = mv_T{}^2/2q = kT/q = 0.026 \text{ volt at } 300° K. \tag{24}$$

This relationship between diffusion constant and mobility is known as the Einstein relationship. It was derived by Einstein on the basis of much more general considerations which show that such a relationship must hold regardless of the nature of the mechanism involved in producing diffusion and mobility. In semiconductor work it has usually proved more effective and accurate to measure the mobility and then to use the Einstein relationship to evaluate the diffusion constant rather than to measure the diffusion constant directly. Recently, however, the Einstein relationship has been directly verified by measuring simultaneously mobility and diffusion of both electrons and holes in germanium. We shall discuss the experiments involved in Section 4.

The orders of magnitude of the diffusion constant and the mobility can be estimated from the values given earlier in this section. Evidently, the diffusion constant will be approximately

$$D = (1/3)lv_T = 33 \text{ cm}^2/\text{sec}. \tag{25}$$

The mobility may then be estimated from the diffusion constant with the aid of the Einstein relationship, and the result is

$$\mu = (q/kT)D = 1,300 \text{ cm}^2/\text{volt sec}$$

$$= 1,300 \text{ (cm/sec) per (volt/cm)}. \tag{26}$$

These values are typical of mobilities for electrons and holes in silicon and germanium. The actual values may differ by 3 or 4 from these and are of course functions of the temperature. The best estimates for values for germanium and silicon are given in Table II. The effects

TABLE II

MOBILITIES IN CM²/VOLT SEC AND DIFFUSION CONSTANTS IN CM²/SEC

	Electrons		Holes	
	μ	D	μ	D
Si	$1,200 \pm 120$	30 ± 3	250 ± 50	65 ± 12
Ge	$3,600 \pm 180$	93 ± 5	$1,700 \pm 90$	43 ± 2

of impurities and temperature upon μ and D are reviewed in a recent article by Conwell.[8]

The combined effects of drift and diffusion are shown in Fig. 9. Here it is supposed that the group of electrons start at $t=0$ at one point indicated by the dot. Their subsequent distributions are represented by the circles,

11 E. J. Ryder and W. Shockley, "Mobilities of electrons in high electric fields," *Phys. Rev.*, vol. 81, p. 139; 1951.

The times since starting for the five circles are in the ratio 1:9:25:49:81. It is seen that the effect of diffusion is more powerful than the effect of drift at short times, and during the first interval of time, some electrons diffuse against the field. As the circle surrounding the electrons becomes larger, the rate at which it grows radially decreases and the dominant effect is that of drift in the electric field. If the circle is taken as the radius at which the electron density is 0.1 of its value at the center, then (7) leads to

$$R = (2.3 \times 4 \times Dt)^{1/2}, \qquad (27a)$$

where t is the time of flight. The distance drifted in the same time is

$$L = \mu Et = \mu Vt/L, \qquad (27b)$$

where V is the voltage drop in distance L. These equations lead to

$$\text{spread/drift} = R/L = (9.2kT/qV)^{1/2}. \qquad (27c)$$

From this it is seen that the spreading is small compared to the drift when V is large compared to $9.2\,kT/q$ or about $\frac{1}{4}$ volt. The largest circle in Fig. 9 corresponds to about 4 volts.
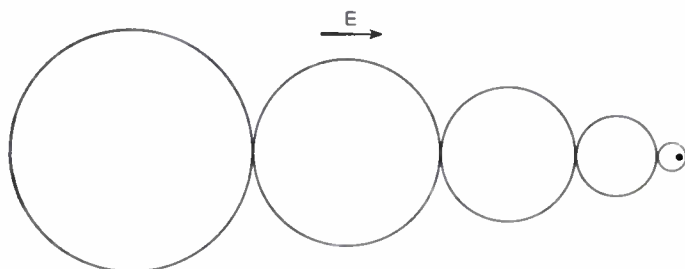
Fig. 9—The relative influence of diffusion and drift.

A magnetic field applied to the specimen will lead to an additional deflection of the electron paths and will produce a phenomenon known as the Hall effect. Although the Hall effect has proved to be of considerable interest in connection with semiconductors and transistor electronics, in general it does not appear to be of sufficient immediate concern with the functioning of transistors to warrant a complete discussion of it at this place. Consequently, we refer the reader to the literature for a discussion of the Hall and Suhl effects.[7]

C. The Behavior of a Hole

The word *hole* is used to describe the electronic imperfection produced by removing one electron from a valence bond. Such a disturbance obviously represents one positive charge equal to the electronic charge. Like the excess electron this charge will be shielded by the dielectric constant of the material. When the perfection of the crystal is disturbed in this way, electronic conduction can be thought of as taking place by a replacement process. An electron in an adjacent bond can jump into the hole in the incomplete bond, thus producing an electronic motion and a reciprocal motion of the hole. If an

electric field is applied such as to move electrons to the left, the hole will move towards the right as a result.

On the basis of this picture, what would we expect the attributes of a hole to be? In the first place, its charge as stated above would be plus-one electronic unit. We would expect the mean-free path to be about one interatomic distance. It is hard to tell what effective mass or mean-free time should be assigned to the hole on the basis of the description given above. The results drawn from this simple reasoning are in disagreement, however, both with theory and with direct observation of the behavior of holes. The conclusions drawn from more complete reasoning and from experiment are at first surprising. When the theory is worked out in detail, it is found that the application of the wave equation to the behavior of the electrons when a hole is present leads to the conclusion that the effective mean-free path for hole motions is of the same order of magnitude as for electrons. The effective mass for the hole is also found to be of the same order of magnitude.[12] Qualitatively then, the hole behaves just as does an electron except for the fact that it acts as though it were a positive charge. Quantitatively, D and μ are somewhat less for a hole than for an electron.

There is no simple way of seeing how the electron replacement process can lead to these long mean-free paths for holes. The analytical reasoning required to reach this conclusion inevitably seems to be complicated.

From the experimental point of view, on the other hand, the behavior of the hole may be regarded as an established fact. The mobility and diffusion constants for holes in germanium have been directly measured, as have those for electrons. It is found that the hole is approximately $\frac{1}{2}$ as mobile as an electron in this case. If the mean-free path were really as short as one interatomic distance, the ratio would have to be 1:1,000 instead of 1:2. Thus the important attributes of a hole may be regarded as determined by direct experimental observations. So we are justified in using these attributes of the behavior of holes in design theory and in the explanation of the way in which transistor devices function.

Although the hole has acquired a very substantial reality as a result of new experiments in transistor physics, its true nature should not be forgotten. The hole is after all simply a convenient way of describing the behavior of an incomplete assemblage of electrons. Attributing to the hole a positive mass, a positive charge, and a mean free path of about 10^{-5} cm leads to a correct description of the way in which this imperfection in the incomplete assemblage diffuses and drifts under the influence of electric and magnetic fields. These are the processes of chief importance in transistor electronics, and for them no error will be made in considering the hole to be a real particle. There are pitfalls, however, in the blind acceptance of this model, and there are circumstances in which the true

[12] An extensive treatment of this topic is given in footnote reference 1.

electronic nature of hole currents may become apparent. For example, adding a hole to a specimen will not increase its mass and its weight by the effective mass of the hole. Adding a hole is really removing an electron, and the mass of the specimen will be decreased by the true mass (not the effective mass) of an electron. The linear momentum of a current of holes in a specimen will be in the opposite direction from the motion of the holes since the momentum really arises from the motion of the assemblage electrons and this motion is in the opposite direction from the hole motion. These considerations are presented chiefly to prevent possible confusion which may arise if the model of the hole is taken too literally.[12a]

D. Photoconductivity, Lifetime, and Deathnium

In Fig. 10 we represent an experiment which may never be carried out in any laboratory. We see a transformer involving a primary winding and a secondary winding consisting of a doughnut of pure germanium. Also shown is a light source. By measuring the input resistance at the primary of the transformer, it is possible
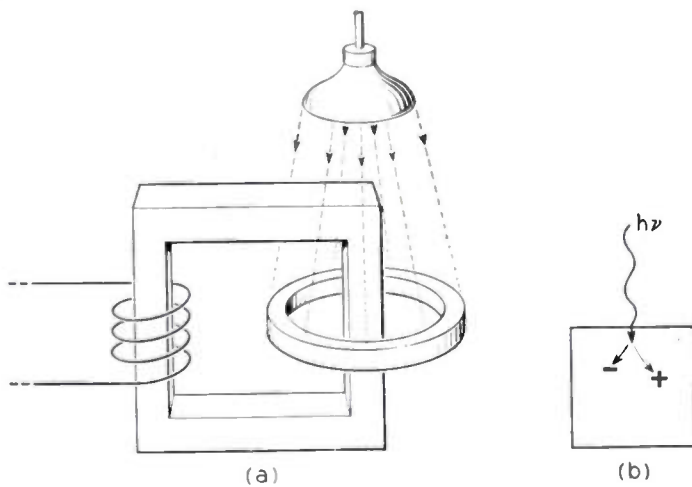
Fig. 10—Production of photoconductivity in a germanium ring by absorption of light; each absorbed photon generates one hole-electron pair.

in principle to measure the resistance of the secondary winding. We have chosen this structure so as to avoid in this discussion the problem of end contacts on the specimen of germanium. When the light is turned on, photons or particles of radiation fall upon the specimen. Some of these photons penetrate the germanium and disturb its electronic structure. This disturbance has been shown by experiments carried out by Goucher[13] to have unit efficiency in the sense that one hole and one

electron are produced for each photon absorbed. What happens is this: If a photon has sufficient energy, it will be absorbed in the crystal and deliver its energy to an electron in one of the electron-pair bonds. This electron will be ejected from the bond, thus becoming an excess electron in the crystal and leaving a hole behind it. Both the hole and the electron will move in the presence of electric fields, and consequently the doughnut will acquire conductivity as a result of the presence of the light. If the number of electrons present per cubic centimeter of the crystal is represented by n and the corresponding density of holes by p, then it can be readily shown that the conductivity of the torus will be given by the formula

$$\sigma = q(n\mu_n + p\mu_p), \tag{28}$$

where q is the charge on the electron and μ_n and μ_p are the mobilities of electrons and holes, respectively.[14] Under the influence of the ac electric field the holes will flow about the torus in one direction and the electrons in the opposite.

If the light is removed, the conductivity will die away as a result of the recombination of holes and electrons. In principle, this recombination can take place by the reverse of the mechanism of generation, the electron can drop into the hole, and radiation can be emitted. This recombination process appears to be of negligible importance in the literature of transistor electronics.

It is found that after the exciting cause is removed, the effect of recombination can well be described by an exponential decay towards a certain equilibrium value.[15] In other words, if the conductivity is measured as a function of time, it is found to obey a law of the sort

$$\sigma = \sigma_0 + \Delta\sigma \exp(-t/\tau). \tag{29}$$

The excess conductivity $\sigma - \sigma_0$ decays exponentially towards zero so that the conductivity finally becomes σ_0.

In the case of a sample of germanium at room temperature, σ_0 may be caused by the spontaneous breaking of hole electron bonds as a result of the presence of phonons in the crystal. This process produces a thermal equilibrium concentration of holes and electrons. In a pure sample of germanium these must evidently be present in equal numbers. Otherwise, the specimen would have a net electric charge. In the last part of this section, we shall discuss this *intrinsic conductivity* again.

The constant τ in (29) is called the *lifetime*. It is analogous to RC for a condenser-resistor combination. In each interval τ, the disturbance decays by a factor $\exp(-1)$. The lifetime τ is not related in any direct way to the mean-free time between collisions, which was given the same symbol. An electron will ordinarily have 10^6 to 10^8 collisions, and thus spend a like number of mean-free times before recombining with a hole.

[12a] Such confusion has recently occurred in connection with metals which conduct by the hole process. For such conduction it was found that the ratio of momentum to current was that expected for electrons and not for positive particles. (S. Brown and J. S. Barnett, *Phys. Rev.*, vol. 87, p. 601, 1952.) This result is not surprising, as Brown and Barnett imply; it is just what should be expected on reasoning presented above and is entirely consistent with the theory of the hole presented in this article and elsewhere in connection with semiconductors and the anomalous Hall effect.

[13] F. S. Goucher *et al.*, "Photon yield of electron-hole pairs in germanium," *Phys. Rev.*, vol. 78, p. 816; 1950, and "Theory and experiment for a germanium *p-n* junction," vol. 81, p. 637; 1951.

[14] This formula may be readily derived by considering the number of carriers crossing the unit area per unit time. See footnote reference 7, p. 202.

[15] J. R. Haynes and W. Shockley, "The mobility and life of injected holes and electrons in germanium," *Phys. Rev.*, vol. 81, pp. 835–843; 1951.

One fact which emerges with great prominence when the lifetimes of added holes and electrons are studied is that there are great variations between specimens which appear to be alike in all other properties. From a study of such observations, one concludes that the recombination process must be controlled by some imperfection present in the crystal. At least one such imperfection is known, and there are probably others that can be active as well. For purposes of convenience the word *deathnium* has been introduced to identify the imperfections that catalyse the recombination and thus limit the lifetime.
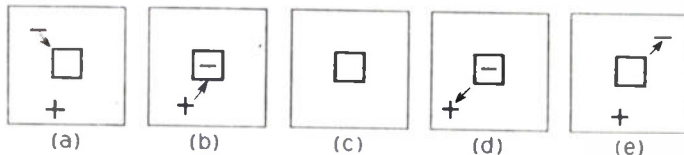
Fig. 11—Possible behavior of deathnium center.

In Fig. 11 we illustrate a way in which the deathnium recombination process can occur. We imagine that in the crystal there is an imperfection of some sort which is capable of trapping an electron. Part (a) of the figure shows the electron approaching this center and part (b) represents the situation after the electron is trapped. If a hole reaches the center while the electron is trapped in it, the electron may easily jump into the hole. The center is then left in its original condition and is ready to recombine another electron-hole pair. Evidently, the reverse process can also go on and the center can generate hole-electron pairs by proceeding through the steps represented by (c), (d), and (e) in the figure. In (d) the deathnium center is supposed to have captured an electron, not an excess electron in this case but one from an electron-pair bond. As a result, a hole is created which diffuses away from the center. Subsequently, in (e) the electron jumps out of the deathnium center which is then returned to its original state and may create other electron-hole pairs or recombine them as the case may be. Conservation of energy is achieved in these processes by the simultaneous absorption or generation of photons and phonons.

One form of deathnium is the imperfection produced by disordering the germanium lattice itself so as to knock some atoms out of their normal positions in the lattice, thus leaving vacancies behind. Such disturbances can be produced by bombarding the crystal with high-energy electron beams or nuclear particles.[16] Such a treatment of the germanium crystal can substantially reduce the lifetime.[17]

The surface of a germanium crystal often acts like a concentrated layer of deathnium. The nature of the recombination centers on the surface is not known, but may be similar to the incomplete germanium bonds that are associated with bombardment-produced disorder in the interior. The effect of surface recombination is so great in specimens of small cross section, for example 0.05 cm on an edge, that it sometimes dominates the recombination process and controls the lifetime. The surface recombination is described by a *surface-recombination velocity*. This velocity may be controlled by surface treatment, some of the results being quoted in the review article by Conwell.[8] In subsequent parts of this article we shall refer to the lifetime in rods of germanium. This lifetime will, in each case, be the effective lifetime resulting from the combined action of recombination in the volume and on the surface. The mathematical procedures for combining these are somewhat involved and will be found in the references cited in connection with *p-n* junctions and *p-n* junction transistors.

E. Intrinsic Conductivity

In the previous portion of this section we emphasized chiefly conductivity induced by the absorption of photons. As was pointed out there, this photoconductivity tends to die away because holes and electrons recombine through the deathnium process. We also pointed out that deathnium can work backwards and generate hole-electron pairs. As a result, a thermal equilibrium density of holes and electrons will be present in an unilluminated specimen. The equilibrium density depends upon the temperature, and we shall analzye below the factors which determine the equilibrium condition.

The statistics of the recombination and generation of holes and electrons by the deathnium process are somewhat involved.[18] Fortunately, it is not necessary to consider this particular process in order to reach conclusions about the thermal equilibrium density of holes and electrons. There is a general theorem of statistical mechanics which states that the equilibrium state of a system is independent of the mechanisms which permit it to be achieved. We have discussed two processes for the generation of hole-electron pairs: One was the absorption of photons and the other the reverse of the deathnium process. In order to determine the equilibrium number we can consider a case in which no deathnium is present and only the photon process is active. This process can also work backwards in the sense that a hole and an electron can recombine and emit a photon. Fortunately, the statistics for the photon process are relatively simple.

Under conditions of thermal equilibrium there will be a certain thermal equilibrium density of black-body radiation present within the specimen. Some of the photons in this distribution have sufficient energy to break the electron-pair bonds and create electron-hole pairs. At temperatures of interest here the number of photons of this class is proportional to

$$\exp(-h\nu/kT) = \exp(-E_G/kT), \qquad (30)$$

[16] J. W. Cleland, J. H. Crawford, Jr., K. Lark-Horovitz, J. C. Pigg, and F. W. Young, Jr., "The effect of fast neutron bombardment on the electrical properties of germanium," *Phys. Rev.*, vol. 83, pp. 312–319; 1951.

[17] See footnote reference 7, p. 347. The cross section for recombination by a bombardment center is estimated from preliminary experiments to be about 10^{-15} cm².

[18] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, p. 387; 1952.
W. Shockley and W. T. Read, "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, p. 835; 1952.

where $h\nu$ is the energy of a photon. We have introduced the symbol E_G to represent the energy required to break a bond. This symbol stands for the *energy gap* in the *energy-band theory of solids*.[8,19] According to this theory each possible state of motion of an electron in the crystal corresponds to an energy level. The electrons in the valence bonds are in one group of energy levels, known as the *valence-bond band* or *full band*, and electrons which are excess electrons and can move about the crystal are in a higher group or band of energy levels, known as the *conduction band* or *empty band*. Between these two bands of energy levels there is a gap for which there are no energy levels. In order to create an electron-hole pair it is necessary to give an electron enough energy to carry it from one band to another across the energy gap.

The rate of generation of electron-hole pairs by blackbody radiation will be proportional to the exponential of (30). The thermal equilibrium state is then found by equating the rate of recombination to the rate of generation. The rate of recombination may be estimated by considering one electron. The chance that this electron will recombine with a hole is evidently proportional to the number of opportunities which it has; and this is proportional to the density p of holes. Hence the rate of recombination of holes with electrons will be directly proportional to the product of their densities. (This simple relationship does not apply when the recombination must take place through a deathnium center.) The proportionality constant which relates the rate of recombination to the product np will be dependent upon temperature because the speed with which the holes and electrons diffuse through the crystal depends on temperature and the distance of approach at which recombination occurs also will depend on temperature. The temperature dependence of this proportionality factor is unimportant, however, compared to the temperature dependence of the photon density given by (30). Hence, so far as temperature dependence is concerned, the condition for thermal equilibrium will be given approximately by[20]

$$np = \text{const} \exp(-E_G/kT). \qquad (31)$$

This equation is reminiscent of a mass-action law in chemistry in which the tendency of the reaction to go in one direction is proportional to the product of the concentrations of the reacting constituents.

In a pure sample of germanium in which there is no charge density due to chemical impurities (a subject which we shall discuss in the next section), the hole and electron densities must be equal in order that the specimen be electrically neutral. This leads to the equation

$$n = p = n_i, \qquad (32)$$

where n_i stands for the concentration in a pure or *intrinsic* sample of material. Evidently, n_i is given by the relationship

$$n_i^2 = \text{const} \exp(-E_G/kT). \qquad (33)$$

The conductivity of such a sample will then be

$$\sigma_i = q(\mu_p + \mu_n)n_i. \qquad (34)$$

We shall have occasion to refer to this intrinsic conductivity in subsequent sections and shall also consider cases in which n and p are not equal, but are related by the constancy of their product given by (31).

3. Chemical Imperfections and Conductivity

The chemical impurities of chief importance in silicon and germanium are known as donors and acceptors. The reason for the importance of these classes of impurities is that their presence in the otherwise perfect crystal induces the presence of electrons or holes, and thus makes the crystal a conductor, or rather a semiconductor. As for the case of the excess electron and the hole, there is a high degree of symmetry between the behavior of a donor and of an acceptor, and we shall discuss the electron-producing donors before the hole-producing acceptors.

A. Donors and n-Type Germanium

Typical donor elements for silicon and germanium are antimony and arsenic. These elements have an inner core and valence electrons as does germanium; but instead of having four valence electrons, they have five. There is a body of experimental evidence that indicates that arsenic or antimony atoms when present in a crystal of germanium or silicon substitute themselves in place of normal atoms;[21] rather than squeezing into spaces where there are normally no atoms. Surrounding a normal lattice position there are four neighbors and places for a total of eight electrons in the electron-pair bonds. An arsenic atom in these surroundings has one too many electrons for its share of the bonds. Consequently, one electron becomes an excess electron and wanders throughout the crystal.

The situation surrounding an arsenic atom is represented crudely in Fig. 12. It is evident that the arsenic atom represents a localized positive charge of $+1$ electronic unit. This conclusion follows from the fact that the electron-bond structure surrounding it would be adequate to neutralize a charge of $+4$ on the core but not $+5$. The resulting unbalanced electric field distorts the bond configuration and produces a dielectric polarization. On Fig. 12 we represent a section through a cube surrounding the arsenic atom. This cube is drawn so that its sides cut midway between nearest neighboring atoms. In the case of a perfect crystal, there would be no electrical charge within the cube. In the case of the arsenic atom, there is one net unit of positive charge within this cube due to the extra charge on the arsenic core. Because of the high dielectric constant of germanium, which is 16, there is a nearly compensating inward shift of the electrons across the surface of the cube so that the total net charge within the cube is only $\frac{1}{16}$ of an electronic charge. This shielding of the charge by the

---

[19] See footnote reference 7 for a discussion.
[20] See footnote reference 8 for values of np.

[21] See footnote 7, p. 25 of reference.

dielectric effect minimizes the attraction of the positive arsenic atom for a wandeirng excess electron. The attraction is much smaller than it would be between an ionized arsenic atom and an electron in free space. At room-temperature thermal energy of motion of the excess electron has a far greater effect than the feeble attraction, and the electron will hardly notice the presence of the arsenic atom.

NET CHARGE INSIDE = q/16

Fig. 12—Dielectric polarization produced by a donor.

At low temperatures, however, thermal agitation is less. If the temperature is low enough, the feeble attraction of the arsenic atom will completely overcome the thermal motion and the excess electron will become trapped in the neighborhood of the positive charge and will be unable to conduct. Information about the strength of the attraction can thus be found by studying conductivity as a function of temperature. Some of the findings based on this method of attack are discussed in a review paper by Conwell.[8]

The principal effect of the extra arsenic atom in the lattice at room temperature is thus to produce a bound charge of +1 unit which is shielded from electrons or other charges by the dielectric constant of the material. This charge cannot move in electric fields since the force holding the atoms themselves in position are very great. As a result, the arsenic atoms do not enter appreciably into the electrical conduction process at any operating temperatures for transistors. It is important that they do not because the permanence of many transistor structures depends upon the chemical imperfections remaining fixed in place.

A crystal containing a relatively large number of donors will have a large net positive charge unless some compensating imperfections are present. In this case the compensating imperfections will evidently be electrons.

Since these electrons are not bound to the donor atoms at room temperature but are free to conduct, the crystal will be conducting. Such a specimen is referred to as *n*-type since the current is carried by localized negative charges, namely, the electrons.

The donor density in atoms per cm³ is usually denoted by N_d. The conductivity will thus be

$$\sigma = q\mu_n = q\mu_n N_d \qquad (35)$$

since n and N_d are equal.

If the number of electrons is not exactly equal to the number of donors, then the specimen will have a net electrical charge. This electrical charge will set up an electric field which will tend to restore electrical neutrality. It is important to note that in a specimen of macroscopic size the degree of neutralization must be extremely good or else enormous electric fields will be set up. For example, if the concentration of donors has a typical value of 10^{15} per cc and the neutralization is 10 per cent imperfect throughout a sphere whose radius is 10^{-2} centimeters, of 4/1,000 of an inch, then an electric field of about 40,000 volts per centimeter will be set up. In some cases, considered below, we shall deal with situations in which neutralization is quite imperfect; but the volumes or charge densities involved will be considerably smaller.

In an *n*-type specimen the product of hole and electron densities must have the value discussed in Section 2E. Evidently, if n is much larger than the value n_i in a pure crystal, p will be much less than n_i and hence much less than n. Consequently, the charge density of holes will be nearly negligible and the requirement of electrical neutrality will lead to equal numbers of electrons and donors.

B. Acceptors and p-Type Germanium

Aluminum and gallium are typical acceptors for silicon and germanium. These elements have three valence electrons rather than four. As a result, they cannot complete the paired electron structure of the four bonds that surround them when they replace an atom in the normal crystal lattice. If an electron from a bond somewhere else is used to fill in this hole, the acceptor atom will acquire a localized negative charge. This negative charge will produce a polarization like that shown in Fig. 12 for a donor except that the effect will be to repel electrons from the bonds in its neighborhood rather than to attract them.

The hole, which is set free when an electron completes the bonds, will be attracted by the negative charge of the acceptor and all of the remarks made about the behavior of an electron will apply in this case with proper change in sign.

The behavior of the hole is, of course, actually the result of the co-operative behavior of the electrons in the valence bonds. It is difficult to explain how the electrons simulate the attributes of a positive particle, but as was pointed out in Section 2, they do. Their behavior in the presence of an acceptor is also described with considerable accuracy in terms of the hole as a positive particle

moving near a fixed negative center. We shall not try to justify these conclusions by attempting an exposition of the theory, but will again remind the reader that both theory and experiment show that a hole behaves much as does an excess electron except for its positive charge.

In a specimen containing a uniform density of acceptors, and no donors, the charge density of the acceptors must be balanced by that of holes. Such a specimen will have p-type conductivity since the charge carriers are positive and the conductivity will be

$$\sigma = q\mu_p N_a, \tag{36}$$

where N_a is the acceptor density.

C. Compensation and Intrinsic Germanium

Since the chief effect of donor and acceptor densities is to produce a distribution of charge density, it is evident that their effects may be combined algebraically into a single *chemical charge density* ρ_c,

$$\rho_c = q(N_d - N_a). \tag{37}$$

An impure or doped crystal having equal values of N_d and N_a is said to be *compensated;* it will behave electrically much like a pure crystal. The phenomenon of *compensation* is of great technological importance; it permits control of conductivity and conductivity type (i.e., n-or-p-type) to be achieved by adding donors or acceptors without first removing those that are present.[22]

There are disadvantages to attempting to control properties entirely by compensation, however. The weak fields of the donors and acceptors do have an influence on mobility if the impurity density is high enough.[8] Small fractional errors in compensated densities will also have a proportionately larger effect on the chemical charge density, and this effect makes control more difficult.

For most specimens of germanium, a condition of *intrinsic conductivity* can be reached by raising the temperature. As the temperature is increased, there is an increase of n_i the density holes and electrons would have in a pure crystal. If n_i is large compared to the difference between N_d and N_a, then both p and n will be approximately equal to n_i with a relatively small diffreence to neutralize the chemical charge density. Under these conditions the conductivity is nearly the same as for a pure crystal, and represents an *intrinsic* property, that is a property inherent in the germanium itself. If a crystal has a difference in N_d and N_a much less than n_i at room temperature, it is said to be intrinsic at room temperature. At a lower temperature it may become definitely n-type or p-type.

4. Injection and Detection of Minority Carriers

In Fig. 13 we represent one of the earliest and most basic experiments of transistor physics. This experiment was designed to test the theory of operation of the point-contact transistor. The theory of the point-contact transistor is that the emitter point-contact introduces holes into the n-type base material. The collector point-contact, which is biased negative, withdraws these holes. Current in the emitter point-contact thus controls the current in the collector contact.

Haynes tested this theory by making an elongated transistor consisting of a rod of germanium somewhat less than a millimeter square and about 1 cm in length.[23] Emitter and collector point-contacts separated by several millimeters were made to this specimen. In a recent form of the experiment,[24] which we shall use for illustration, the experimental procedure was as follows: A pulse of positive charge was applied to the emitter point, the duration of the pulse being short compared to other times involved in the experiment. Current drawn by the collector point was then observed on an oscilloscope. If the emitter point injects holes into the rod, then these holes will flow towards the right as a result of the sweeping field set up by the battery applied to the two ends of the rod. If the holes move in the fashion described in Section 2, then the hole pulse at successive intervals of time will be as represented in part (b) of the figure. When the leading edge of the pulse of holes arrives at the collector point, the collector current will start to increase; and when the middle of the pulse is at the collector, the current will reach its maximum value; after this the current will decay.
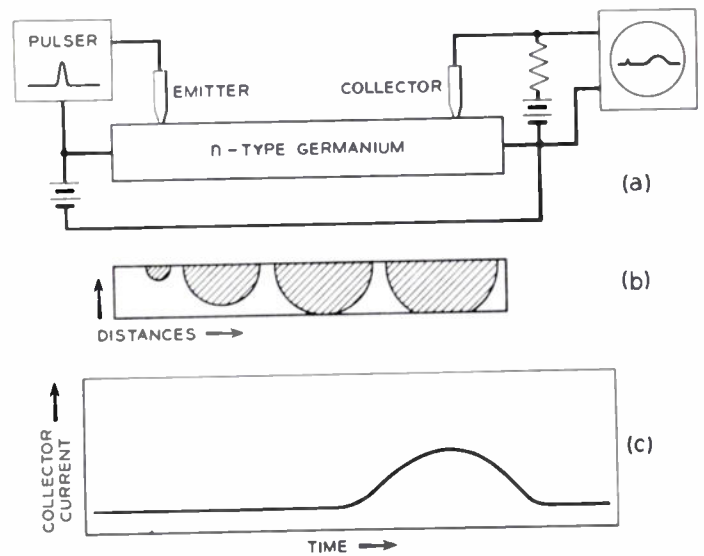
Fig. 13.—Observation of drift and diffusion of an injected group of minority carriers.

The results observed by Haynes and his collaborators are in agreement with prediction, the appearance of the oscilloscope traces being similar to those shown in part (c) of Fig. 13. From measurement of the electric field, the distance, and the time, it is possible to determine both the mobility and the diffusion constants from this experiment and to see whether or not the two are related by the Einstein relationship. The relationship is

[22] The development of this viewpoint and of the related techniques is reported by J. H. Scaff, H. Theurer, and E. E. Schumacher, "P-type and N-type silicon and the formation of the photovoltaic barrier in silicon ingots," *Jour. Met.*, vol. 185, pp. 383–388; 1949.

[23] J. R. Haynes and W. Shockley, "Investigation of hole injection in transistor action," *Phys. Rev.*, vol. 75, p. 691; 1949.

[24] "Experimental verification of the relationship between diffusion constant and mobility of electrons and holes," submitted to *Phys. Rev.*

indeed found to be satisfied, and the orders of magnitudes of the diffusion constant and the mobility are in accord with the orders of magnitude that one would estimate on the basis of theory.[24]

However, it is not at all evident at first that the motion of the holes in Haynes' experiment should be that described in Section 2. After all, a relatively large number of holes are introduced, and the electric field that the charge of these holes could produce is far from negligible. Furthermore, a large number of electrons are present, and their motions will be interfered with by the electric fields set up by the holes. Thus more careful consideration is certainly necessary to show that the pulse of holes introduced in the rod will actually move as though the disturbance in electric field set up their presence were negligible. The critical condition necessary for this approximation to be true is that the concentration of the holes should be small compared with the concentration of electrons already present. We shall consider this point below.

Let us then consider first a case in which a small group of added holes are present in an n-type semiconductor in the absence of any sweeping field. If these holes were suddenly introduced by some magical means before the electrons could move, they would set up a large positive electrostatic potential. This electrostatic potential would tend to make the group of holes blow apart, and at the same time would tend to attract electrons inwards. Since there are vastly more electrons in any element of volume than there are holes, according to our assumption that the holes represent a small disturbance, the motion of electrons would predominate and the potential would be reduced to a small value before any large motion of holes took place. The situation would then be as represented in parts (a) and (b) of Fig. 14. The concentration of electrons would be larger than its normal value in the presence of the holes, but a slight excess of hole-charge density would remain. What we wish to show next is that the motion of the holes for this simple case is governed largely by diffusion, and the effect of drift in the electric field due to the holes is negligible relatively, providing the concentration of the minority carriers, holes in this case, is very small compared to that of the majority carriers.

The reasoning on which this conclusion is based may be understood in terms of the arrows in Fig. 14(a). These arrows are supposed to be proportional to currents expressed in terms of particles per unit area per unit time. We shall start by assuming that the drift current of holes is negligible. This assumption can then be justified by the following considerations: (1) Since electron and hole concentrations are nearly equal, so are their concentration gradients; since the diffusion constant for electrons is twice that for holes, the electron diffusion current is twice that for holes. This is represented in part (a) of the figure. (2) Electrical neutrality requires that equal numbers of holes and electrons flow out of the region. Hence, the electron drift current must cancel approximately half of the electron diffusion current. We now come to the essential point. (3) Since the

hole density is much less than the electron density, and since the same electric fields acts on both, the hole drift current must be much less than the electron drift current. This leads to the conclusion that the hole drift current must be shown by the small arrow in Fig. 14(a).
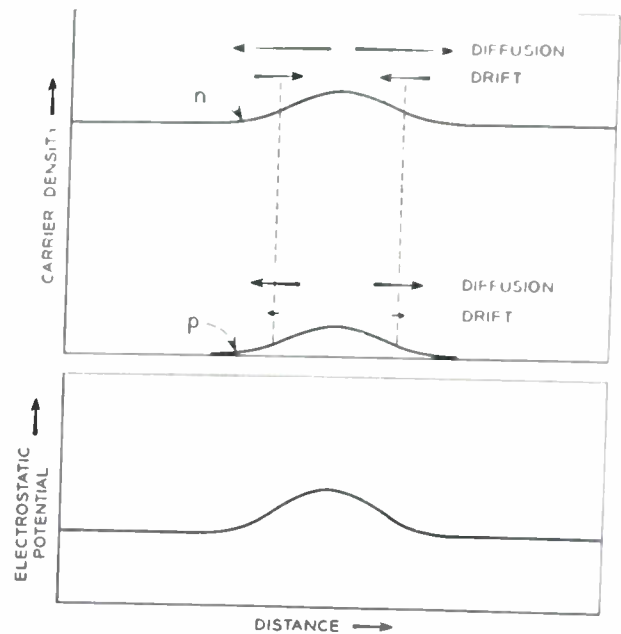
Fig. 14—Diffusion and drift for a small density of added minority carriers.

This important result is applicable in the theory p-n junctions, junction transistors, and the Haynes' experiment discussed in Fig. 13. The essential conclusion may be stated as follows: The influence of small concentrations of minority carriers in the presence of larger concentrations of majority carriers produces such a small effect upon the electric field that the motions of the minority carriers can be calculated as though they took place in the electric fields which would be present if no minority carriers were introduced.

The application of this reasoning to Fig. 13 shows that the hole distribution should drift in the sweeping field in the manner proposed at the start of this section. The effect of somewhat larger minority densities has also been investigated. For information on this important subject the reader is referred to the literature.[23,25,26]

5. *P-N Junctions as Rectifiers and Photocells*

From the discussion of Section 3 it is evident that in a single crystal there may be both p-type and n-type regions. The boundary between such regions is called a p-n junction. Such p-n junctions have interesting electrical and optical properties. An understanding of their nature is basic to many phases of transistor electronics.[27]

[25] C. Herring, "Theory of transient phenomena in the transport of holes in an excess semiconductor," *Bell Sys. Tech. Jour.*, vol. 28, pp. 401–427; 1949.

[26] W. van Roosbroeck, "Theory of the flow of electrons and holes in germanium and other semiconductors," *Bell Sys. Tech. Jour.*, vol. 29, pp. 560–607; 1950.

[27] The theory of p-n junctions presented here is discussed more fully in W. Shockley, "The theory of p n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, pp. 435–489; 1949. See, also, footnote reference 7.

Fig. 15 represents a *p-n* junction. In discussing its electrical properties, we will be concerned with the five kinds of imperfections shown in the lower part of the figure. From a mechanical point of view the crystal is practically homogeneous and perfect. A typical concentration for impurities in the crystal might be 10^{15} cm^{-3}.
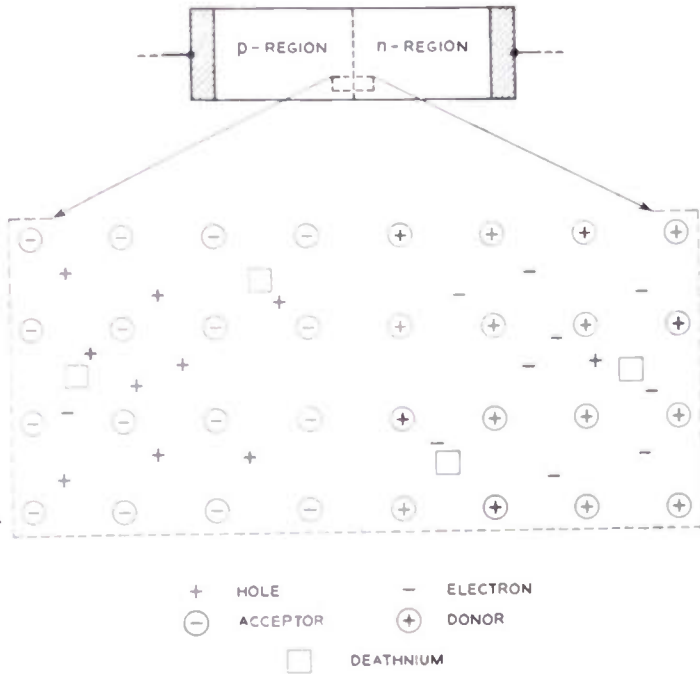
Fig. 15—A *p-n* junction and the distribution of imperfections in it.

This density of imperfections is so low that if one were to traverse a row of atoms from end to end in the crystal one would, on the average, strike only about ten imperfections. Thus the crystal structure itself is only slightly altered by the presence of the imperfections. From the electrical point of view, on the other hand, the imperfections have profound effects.

As is shown in Fig. 15, the electrons are found chiefly in the *n*-region where they neutralize the chemical space charge of the donors and the holes are similarly found in the *p*-region. In order for this situation to persist, as it does under conditions of thermal equilibrium, there must be an electric field present in the crystal. The idea that an electrical field is present in a conductor under conditions of thermal equilibrium is at first surprising. However, the necessity for such an electric field can readily be seen in terms of Fig. 15. Let us first of all suppose that no electric field is present across the junction; then as a result of diffusion, holes will cross the junction from left to right and electrons will similarly diffuse across the junction from right to left. As a result, there will be a net transfer to the right of positive charge across the junction. This will result in an electric field which will build up to just such a point that further current flow is prevented.

The electric field in the *p-n* junction is represented in Fig. 16(a) in terms of an electrostatic potential ψ. The remaining parts of the figure show how this electric field and electrostatic potential arise from the charge densities involved. The chemical charge densities are shown in parts (b) and (c) of the figure. In this example it has

been assumed, as represented by the N_d and N_a curves, that the transition from *n*-type to *p*-type occurs abruptly at the junction and that compensation of one impurity type by another is not involved. In the presence of the electrostatic potential shown in part (a) of the figure the holes tend to be pushed to the left. As a result, the hole density drops to a small value before the junction itself is reached. Electrons having a negative charge tend to move to the points of highest electrostatic potential, and thus they also are not found near the center of the junction. As a consequence, the chemical charge density is not compensated by holes or electrons in the immediate neighborhood of the junction. Thus an electrostatic dipole layer is formed at the junction, the charge density being as shown in part (f) of the figure. This dipole layer is seen to be of just the nature necessary to produce electrostatic potential shown in part (a).
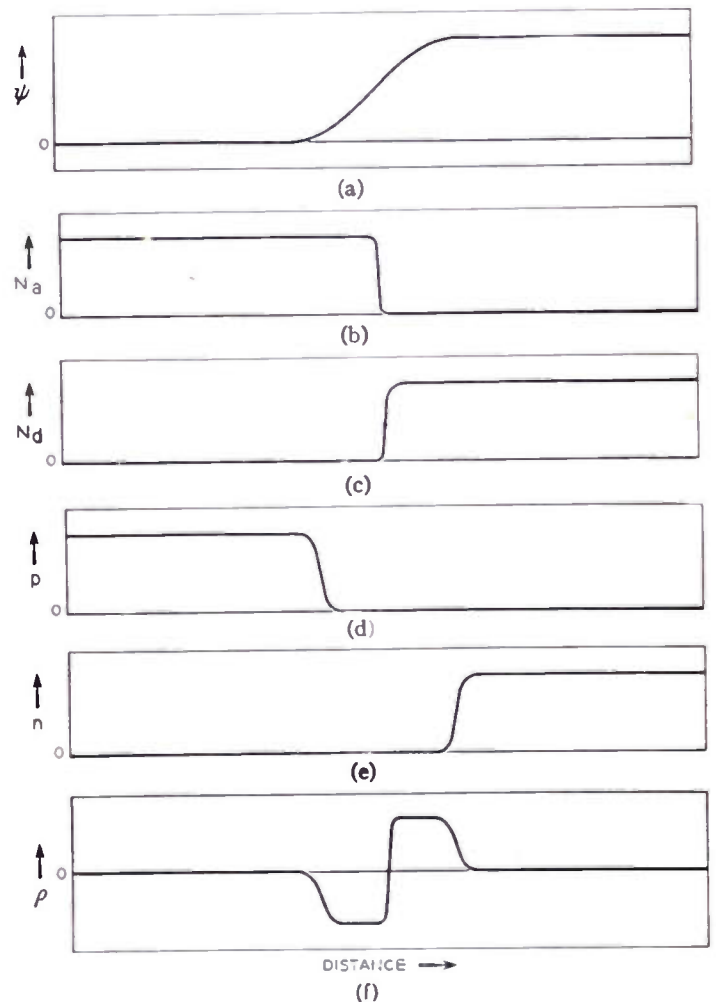
Fig. 16—Potential and charge distribution in a *p-n* junction.

Mathematically, what is done in order to determine the shape of the electrostatic potential in Fig. 16(a) is to solve a differential equation. If the dependence of electrostatic potential upon distance is regarded as the unknown, then from it and certain principles of statistical mechanics it is possible to write an expression for the charge density due to the holes and the electrons. These charge densities can be combined with those due to chemical imperfections in order to obtain a differential equation for the electrostatic potential. This differential

equation is Poisson's equation, which relates derivatives of the electrostatic potential to the charge density. When this equation is solved, it is found that the situation in the *p-n* junction under thermal equilibrium conditions is as represented in Fig. 16.

Under conditions of thermal equilibrium no net current of either holes or electrons will flow across the junction. It is advantageous, however, to consider this equilibrium situation as arising from compensating currents. We shall illustrate this by considering the flow of holes back and forth across the junction. Although the density of holes is small in the *n*-region, it is still appreciable and plays a vital role in the behavior of the *p-n* junction. Let us consider the course of a hole that arrives in the *n*-region by climbing the potential hill as illustrated in Fig. 17. Having climbed the hill and arrived at the plateau of constant electrostatic potential in the *n*-type region, it will then proceed to move by a random diffusive motion of the type discussed in Section 2. The most probable outcome of this motion will be that it will diffuse to the steep part of the hill and slide back down into the *p*-type region. We shall not be concerned with holes which follow this particular course. On the other hand, it may, by chance, diffuse more deeply into the *n*-type region. In this event, it will on the average diffuse for lifetime τ, and subsequently it will be captured by a deathnium center in which it recombines with an electron.
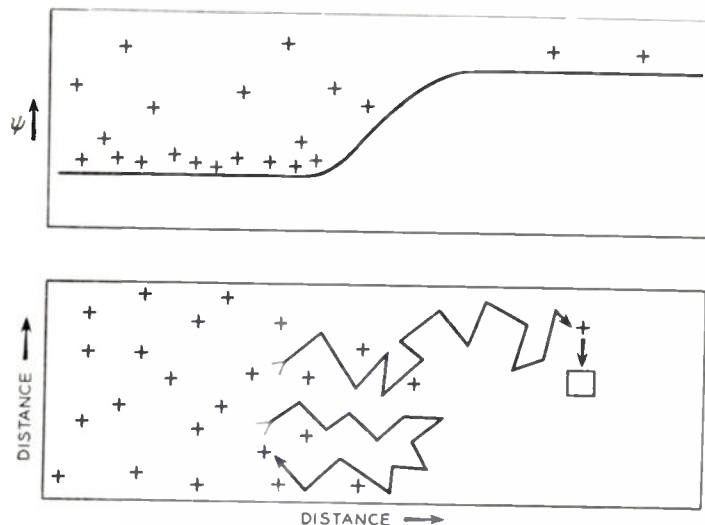
Fig. 17—Hole current from *p*-region to *n*-region in a *p-n* junction.

The average depth to which holes diffuse in the *n*-type region depends upon the lifetime. The holes spread out in the region by diffusion in the manner described in Section 4. When the suitable differential equation describing this process is solved, it is found that the average depth to which they penetrate is given by the equation

$$L = \sqrt{D\tau},$$

where L is known as the diffusion length, D is the diffusion constant for holes, and τ is the lifetime for holes in the *n*-region. Thus under equilibrium conditions a current of holes flows from the *p*-region into the *n*-region

and penetrates on the average one diffusion length L before recombining with electrons.

Under equilibrium conditions a *principle of detailed balance* holds. This principle of statistical mechanics says that each process and its opposite occur equally frequently. Hence we must conclude that the flow of holes from the *p*-region into the *n*-region, followed by recombination, must be exactly balanced by a reverse process. The reverse process is thermal generation of holes through deathnium centers, followed by diffusion to barrier where they slide down into *p*-type region.
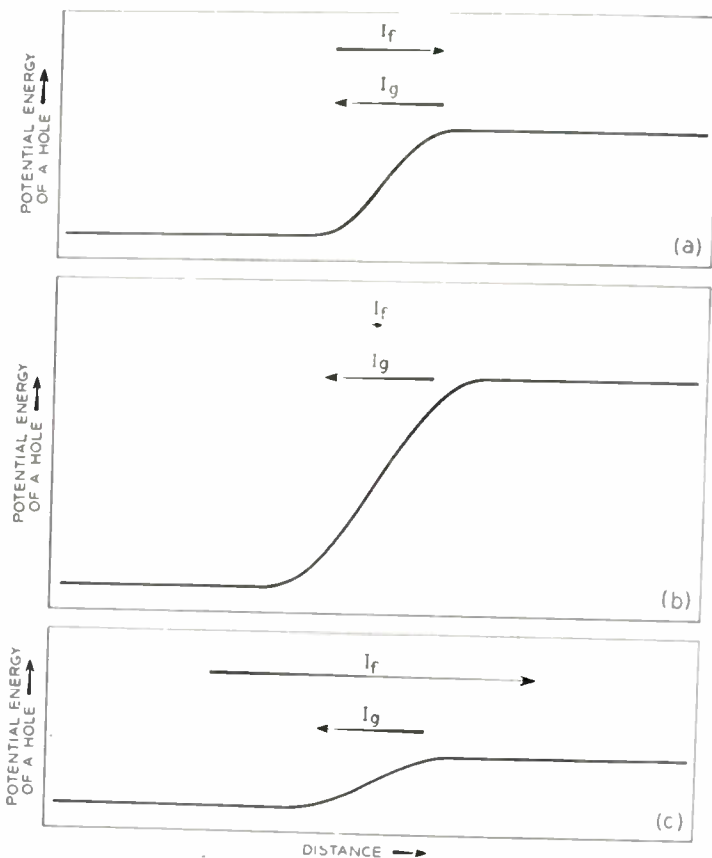
Fig. 18—Dependence of recombination and generation currents upon bias. (a) Thermal equilibrium. (b) Reverse bias. (c) Forward bias.

The application of voltage to the terminals of the device shown in Fig. 15 destroys the exact balance of the two currents just discussed. In considering the application of voltage we shall neglect any voltage drops at the contacts between the metal electrodes of Fig. 15 and the semiconductors. At the end of this section we will return briefly to the reasons why such voltage drops may be neglected. The effect of the application of voltages upon the currents is represented in Fig. 18. In part (a) of this figure we show the thermal equilibrium condition. The two currents previously discussed are represented by I_f and I_g, these currents standing, respectively, for the current of holes entering the *n*-region and recombining and the current generated in the *n*-region and diffusing to the barrier.[27a] For the condition of thermal equilibrium

[27a] The subscript f and g may be thought of as "forward" and "generation." This mixed choice avoids subscript v, which might be either "reverse" or "recombination." In this decision, "forward" is equivalent to "recombination" and "reverse" to "generation."

these two currents are equal and opposite. In part (b) of the figure the situation for a large "reverse" bias is shown. For reverse bias, negative voltage is applied to the p-region and positive to the n-region so that the electrostatic potential difference between the two regions is increased. If the electrostatic potential is sufficiently high, corresponding to the situation shown in part (b), then practically no holes can climb the potential hill and I_f drops nearly to zero. This situation is represented by showing I_f as a vector of negligible length whereas I_g has practically the same value as it has for the case of thermal equilibrium. In general, the diffusion length L is large compared to the width of the dipole or space-charge region. Hence the region where I_g arises is practically unaffected by reverse bias and I_g is thus independent of reverse bias. This independence of current upon bias is referred to as *saturation*.

When forward bias is applied, the situation shown in Fig. 18(c) occurs and I_f increases. This increase is described in terms of the energy difference for a hole in the n-region and p-region. This energy difference is equal to the charge of the electron times the electrostatic potential differences between the two sides. We can apply a general theorem from statistical mechanics to a consideration of the number of holes which, by chance, acquire sufficient energy to climb the potential hill. This theorem states that each time the potential hill is increased by one thermal unit of energy, kT, then the number of holes capable of climbing the higher hill is reduced by a factor of $1/e$. Since the potential barrier is already present under conditions of thermal equilibrium, it follows also that each lowering of the barrier by an amount kT will increase the current by a factor of e. The change in height of the barrier caused by the applied voltage V is $-qV$, where the polarity is so chosen that positive values correspond to plus potentials applied to the p-region. For $V=0$ is the equilibrium case, and for this case I_f is equal to I_g. Hence, in general, the recombination current is

$$I_f = I_g \exp qV/kT. \tag{38}$$

This gives rise to a total current of holes from p-region to n-region, given by the difference

$$I_f - I_g = I_g(\exp (qV/kT) - 1). \tag{39}$$

This current is zero when $V=0$, increases exponentially to large values for positive V, and decreases to a negative saturation value of I_g when V is negative and substantially larger than kT/q.

Similar reasoning can be applied to the electron current flowing across the junction. The applied potential which lowers the potential barrier for holes evidently lowers it also for electrons; consequently, large electron currents flow under the same voltage conditions that produce large hole currents. In both cases these large currents correspond to flows of minority carriers into each region. In both cases the current is in the sense of transferring positive charge from the p-region to the n-region. In one case this is carried in the form of positive imperfections, namely holes, moving from p to n,

and in the other case it is due to negative imperfections, electrons, moving in the opposite direction. For the case of reverse biases the potential rise is larger and the holes tend to be retained in the p-region and the electrons in the n-region. A saturation current due to generation in both regions flows. If the total saturation current is called I_s, then the total current for any applied voltage V is given by the formula

$$I = [\exp (qV/kT) - 1]I_s. \tag{40}$$

Evidently, I_s is the sum of the two generation currents. This equation is found to be well satisfied for p-n junctions in germanium, and a comparison of the rectification curve[28] as measured with the theoretical formula is given in Fig. 19. It should be noted that the separation between the forward and reverse branches of the curves corresponds to a factor of e when the voltage is $kT/q = 25$ mv. This is exactly the factor predicted by (40). This agreement between theory and experiment is evidence that the imperfections which carry the current in a p-n junction have the charge of the electron. If they had twice this charge, a value of 12.5 mv should be obtained; if half the charge, the value should be 50 mv.
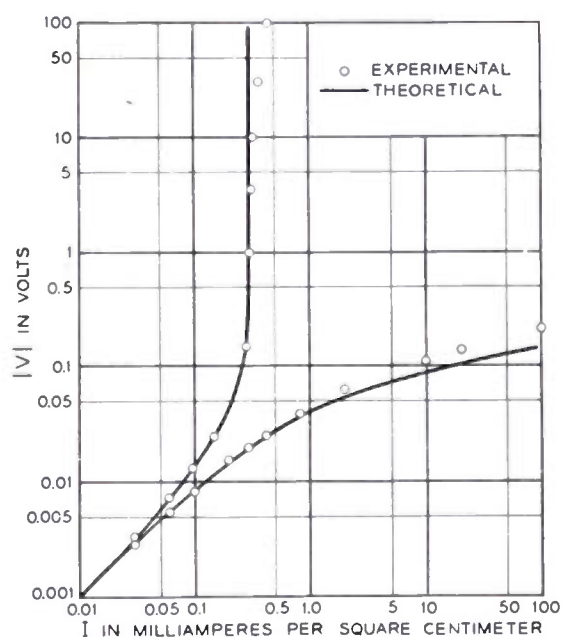
Fig. 19—Rectification characteristic for a p-n junction.

For large forward biases, the potential barrier between the n and p-type is nearly eliminated. Under these conditions large concentrations of minority carriers flow across the barrier and density of the majority carriers may be substantially disturbed. Under these conditions it is no longer valid to consider that the minority carriers diffuse in a field-free region on the basis of reasoning like that presented in Section 4. Although these large signal conditions are of general interest, we shall not consider them further in this section.

There are a number of ways in which the diffusion theory for rectification in the p-n junction can be tested experimentally. We shall consider a test based upon a

[28] See the second part of footnote reference 13.

photoelectric effect. As was discussed in Section 2, photons of sufficient energy can generate hole-electron pairs when they are absorbed in germanium. This generation adds to the thermal generation produced in the deathnium centers. If the light is focused on the specimen in the form of a small spot, then it is possible to generate minority carriers in either region at a precise distance from the junction. The current flowing across the junction as a result of generation by the light should then decrease with increasing distance of the light from the junction. It can be shown that the probability that a minority carrier generated at a distance x from the junction diffuses to junction before recombining is simply

$$\exp(-x/L), \tag{41}$$

where L is the diffusion length of the minority carrier in the region where it is generated. This exponential dependence of response upon distance from the junction has been verified directly by Goucher and his colleagues.[28] They have found also that value of L determined by studies of this sort is consistent with that necessary to explain value of I_s in rectification formula.

For purposes of illustration we shall consider the value of the diffusion length for a typical example. A typical lifetime for a minority carrier is 10^{-4} sec and the diffusion constant for electrons in germanium is 93 cm²/sec. These lead to a diffusion length of about 0.1 cm or 1 mm.

Since the light acts as a current generator, its effect may be readily included in an equivalent circuit for a p-n junction. This is illustrated in Fig. 20. Here I_1 is the
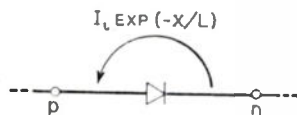
Fig. 20—Equivalent circuit for a p-n junction photocell.

current of minority carriers generated by the light. The equivalent circuit shown corresponds to a case in which the light is focused at a distance x from the junction. If the light is distributed, the appropriate average of the exponential probability factor should be used. If the light falls on both the n-type and the p-type region, this average should take into account the fact that the diffusion length for holes in the n-region is probably different from that for electrons in the p-region. Fig. 20 emphasizes the importance of considering the light as acting as a current generator. If the equivalent circuit of Fig. 20 is operated in an open circuit condition, then a photovoltage will be developed. This photovoltage will in general be nonlinear in the light intensity because of the rectification characteristics of the rectifier. In the equivalent circuit of Fig. 20, however, the current generator is linear in the light intensity. This independence of the current generator of biases applied to the p-n junction has been verified over a wide range of experimental conditions by Goucher and his colleagues.

In closing this subject we shall consider how the magnitude of the current in a p-n junction depends upon the properties of the n- and p-type regions which compose it. For this purpose it suffices to consider I_g, the

current carried by holes generated in the n-region. The consideration for electrons is entirely analogous. We shall accordingly suppose that the lifetime of a hole in the n-type region is τ and the normal concentration of holes in that region is p_n. Each of these holes normally lives for an average time τ and then recombines with an electron through a deathnium center. This implies that the rate at which holes are recombining with electrons per unit volume is

$$p_n/\tau. \tag{42}$$

Under equilibrium conditions this rate will be exactly equal to the rate of thermal generation of holes through the reverse process. The current I_g, which flows to the junction as a result of generation, is due chiefly to holes generated within one diffusion length of the junction. In fact for a junction of unit area, the current flowing to the junction will be equal to the generation within a slab of unit area and depth L. Hence the current density across the junction will be equal to

$$q(p_n/\tau)L = qp_n\sqrt{D/\tau}. \tag{43}$$

From this equation we see that the current flowing to the junction increases with the rate of generation, which is proportional to $1/\tau$, but only as the square root rather than as the first power. The reason for this square-root law is that although the rate at which holes are generated is proportional to $1/\tau$, the effective region from which holes may reach the junction decreases as τ decreases. The effective generation is the more important so that the reverse current increases with decreasing lifetime but only to the square-root power as shown above. The value of p_n in the n-region depends upon the electron concentration in this region since, as discussed in Section 3D, the product of hole times electron concentration is a constant at a given temperature. Furthermore, to a good approximation, the electron concentration is proportional to the conductivity σ_n of the n-region. When this factor is taken into account and the diffusion equations for hole flow to the junction are solved, it is found that I_g for holes is given by

$$I_g = b(1 + b)^{-2}(\sigma_i^2/\sigma_n L)(kT/q), \tag{44}$$

where

I_g = saturation current per unit area due to holes   (45)

$b = \mu_n/\mu_p$ (46)

μ_n = mobility of an electron   (47)

μ_p = mobility of a hole   (48)

$\sigma_i = q(\mu_n n_i + \mu_p n_i)$

    = conductivity of intrinsic germanium   (49)

$\sigma_n = q\mu_n n$ = conductivity of n-region   (50)

L = diffusion length of a hole in n-region.   (51)

A similar expression holds for electron current. The dependence on constants discussed above has been verified for a p-n junction studied by Goucher.[28]

From the equation for I_0 due to holes it is evident that large I_s and small junction resistances can occur for small lifetimes τ and short diffusion lengths L. On this basis we can give a reason for not being concerned about the end electrodes shown in Fig. 15. Metal electrodes, in effect, may be regarded as regions of very short lifetimes for either holes or electrons. Hence the resistance of a metal-to-semiconductor contact may be quite low compared to that of the p-n junction, and can be neglected. This point of view is an over simplification, however, as we shall see in connection with the discussion of point contacts. However, it will serve as a temporary justification for neglecting voltage drops across the end electrodes in Fig. 15.

6. The p-n Junction Transistor

In terms of the concepts of currents carried by diffusion in p-n junctions we can readily understand the mode of operation of a junction transistor.[29] Among its numerous features of interest, the junction transistor is possibly the most calculable of all electronic amplifying devices. The essential feature that gives the junction transistor its unusual simplicity is that in it the pattern of flow is essentially one dimensional.

Fig. 21 gives a diagrammatic representation of an n-p-n transistor. In a preferred form it consists of a single crystal of germanium divided into three regions. These regions are alternately n-type, p-type, and n-type. Metallic contacts are made to each of the three regions; these contacts have low resistance and are of such a nature that they carry current to and from the semiconductor chiefly as a flow of majority carriers, and thus do not inject minority carriers into the body of the germanium. As is indicated on the figure, the contacts are described by the symbols ϵ, b, and c which stand, respectively, for emitter, base, and collector.

Under conditions of thermal equilibrium, the distribution of imperfections in the junction transistor is as shown in part (b) of Fig. 21. In this case of this transistor, the role of deathnium is not so vital as in a simple p-n junction, and symbols representing the deathnium imperfections have been omitted. The potential energy for a hole in the structure is represented in part (c) and the potential energy for an electron in part (d). The electrons and holes both tend to concentrate in regions where their potential energies are low. There are, however, a few minority carriers of each type in each region.

When the junction transistor is used as an amplifier, biases such as those in Fig. 22(a) are applied. These biases produce the potential energy distribution for electrons shown in Fig. 22(b). The positive bias applied to the collector region biases the p-n junction between it and the base in the reverse direction. The current across this collector junction, however, is much greater than that which would be predicted on the basis of the

(a)
(b)
(c)
(d)
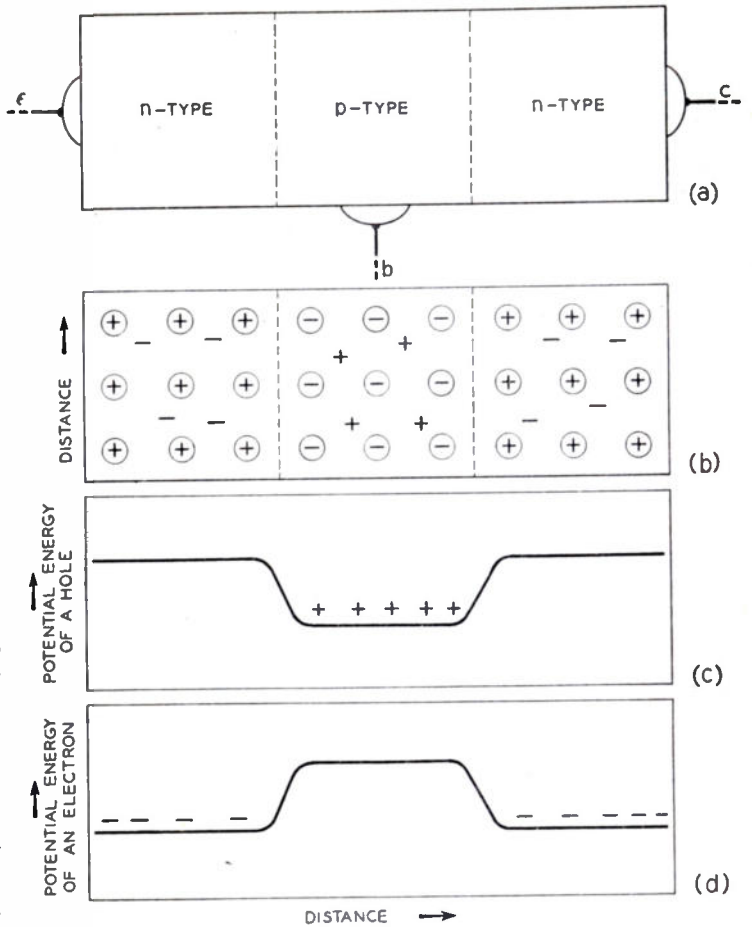
Fig. 21—An n-p-n structure under conditions of thermal equilibrium.

(a)
(b)
(c)
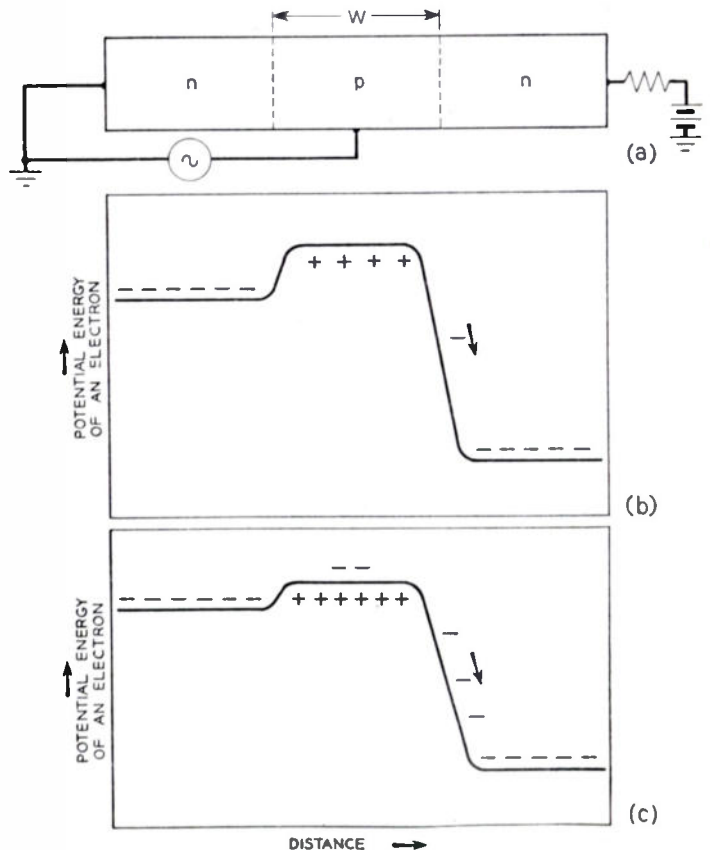
Fig. 22—A junction transistor biased to amplify or oscillate.

[29] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, pp. 435–489; 1949.
W. Shockley, M. Sparks, and G. K. Teal, "p-n junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; 1951.
M. Sparks, "The junction translator," *Sci. Am.*, pp. 28–32; July, 1952.

p-n junction theory of the previous section. The reason for the difference is that in the case of the transistor, the minority carriers flowing out of the base layer do not have to be generated in the base layer, but can instead enter it from the emitter region. In effect, it is as if an infinite supply of deathnium were located at the junction between the emitter and the base. Consequently, the length quantity which should be used for electron current in the p-n junction formula (44) should be the width W of the base layer rather than the diffusion length for electrons in the base layer. The hole current flowing from the collector region across the junction is, of course, perfectly normal providing only that the length of the collector body is large compared to the diffusion length for holes in it. This current can then be calculated from the formula (44).

In Fig. 22(b) the potential distribution corresponds to no signal applied between emitter and base. If a positive signal is applied to the base as represented in part (c), then the emitter-base p-n junction will be biased forward. As a result, a larger number of electrons will be able to diffuse into the base. If the diffusion length in the base is large compared to its thickness W, then an electron is unlikely to recombine in the base as compared to the probability that it will diffuse entirely through the base, reach the collector junction, and slide down the potential hill into the collector. Hence most of the current of electrons which flows out of the emitter body will diffuse through the base layer and reach the collector. The fraction of this current which recombines with holes in the base can be shown to depend upon the ratio of the natural diffusion length in the base material to the width W of the base region. This fraction is approximately equal to

$$(W/L)^2/2. \tag{52}$$

For small values of W/L this fraction is nearly negligible. Consequently, for many purposes one can consider that all of the electron current entering the base region from the emitter arrives at the collector.

If it were true that all changes in the emitter current appeared in the collector current, then there would be no change in base current and, when operated with the emitter grounded, the junction transistor would behave much like an ideal negative grid triode which draws negligible low-frequency ac control current. However, the junction transistor does require finite base currents in order to modulate the emitter and collector currents. One of these sources of base current is the recombination of electrons with holes in the base layer. This is unimportant in many cases, and we shall neglect it as mentioned above. The other source of base current consists of a flow of holes from the base into the emitter. This current is an ordinary junction current and can be calculated by the methods described in Section 5. The ratio of this hole current to the electron current can be shown to be

$$W\sigma_p/L\sigma_n. \tag{53}$$

This result may be seen from (44) of Section 5 by using L for the diffusion length of holes in the n-region

and W for the effective diffusion length of electrons in the p-region. Physically, the reason that the ratio of hole current to electron current takes this form can be seen as follows: The current flowing across the junction may be hampered by two factors. The more heavily doped the region into which the minority carriers must penetrate, the more difficult it is for them to arrive there. Once they have arrived as minority carriers across the junction, the probability of their staying and recombining is less the larger the diffusion length. Hence the factors σ_n and L are both unfavorable to hole flow into the emitter region. The corresponding unfavorable factors for electron flow into the base region are σ_p and W. Hence the hole flow is smaller than the electron flow by the ratio given in the previous equation.

Although a finite current must be furnished to the base in order to change the collector current in a junction transistor, the current gain factor may be quite large. We can estimate the current gain from the above reasoning. When the base potential is altered, there is a change in electron current flowing through emitter and through base to collector and a change in hole current from base to emitter. If the collector junction is saturated, as is usually a good approximation, the hole flow from the collector region to the base will be unaltered. To the degree of approximation discussed above, we neglect electron recombination in the base layer, and we then conclude that the added base current consists entirely of hole current into the emitter region. From this it is evident that the current gain factor is simply the ratio of change in electron current to change in hole current across the emitter junction, and is therefore

$$L\sigma_n/W\sigma_p. \tag{54}$$

This current gain may be 100 fold or more in a junction transistor.

In addition to having high current gain when operated with a grounded emitter, the junction transistor has impedances that are favorable to high gain. In particular, the output impedance or collector impedance is high due to the fact that the collector in general will be saturated; as soon as the collector bias is greater than a few times kT/q, say perhaps 0.1 volt, the collector current will be substantially independent of the collector bias just as is the case for the saturation current of a simple p-n junction. Under these conditions the output impedance would be nearly infinite. Actually, output impedances of the order of megohms are obtained for good junction transistors. The dependence of electron current upon base potential, on the other hand, is very sensitive to base potential so that 39,000 micromhos of transconductance will arise for each milliampere of emitter current. (This gives a figure of merit for transconductance per unit current of q/kT or 39 volts^{-1}. As a result of the high current gain, high transconductance, and high collector impedance, very high gains per stage can be achieved through junction transistors, and typical values may be as high as 50 db.[30]

[30] R. L. Wallace, L. G. Schimpf, and E. Dickten, "A junction transistor tetrode for high frequency use," PROC. I.R.E., vol. 40, pp. 1395–1401; this issue.

If the lifetime in the emitter and collector regions are high and the conductivities also high, then the saturation current across the collector junction may be of the order of a few microamperes. Also, the collector current saturates at a voltage a few times larger than kT/q so that amplifiers and oscillators can be made with power supplies furnishing less than a microwatt of power.[31]

In the above discussion of the operation of the junction transistor, we have tacitly made several approximations. In particular, it has been supposed that the voltage drops in the transistor occurred entirely across the p-n junctions. Actually, there may be potential drops within the uniform body of the semiconductors due simply to their ohmic resistances. In particular, there may be a drop along the p-type base layer due to the current leaving it and flowing into the emitter. This voltage drop will cause uneven bias between emitter and base to be present. In addition, capacitances have been neglected. The capacitance of a p-n junction may be quite large, and this neglect can become as important at high frequencies. As has been discussed by Wallace and his collaborators, these capacitances and the series resistance in the base layer may play a vital role in the high-frequency behavior of junction transistors.[32]

An interesting extension of the theory of the junction transistor has recently been made by Early.[33] His investigation is concerned with the effect of reverse bias at the collector junction upon the effective thickness of the base layer. The essential mechanism is that discussed in the subsequent section on depletion regions.
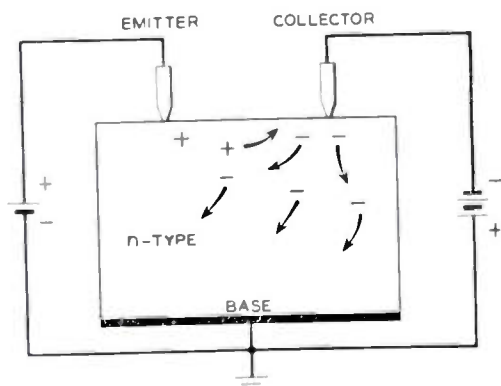
Fig. 23—Hole and electron currents in a point-contact transistor.

7. COMMENTS ON POINT-CONTACT TRANSISTORS

The general functioning of the point-contact transistor has been described by Bardeen and Brattain.[34] Here we shall restrict our comments chiefly to aspects that are concerned with imperfections in germanium. The role of imperfections is of particular importance near the collector point. In Fig. 23, the current flow in a point-contact transistor is represented. Holes injected at the emitter point flow in the electric field set up by the collector current. They are thus drawn to the collector and add their current to the normal reverse current of the collector contact. In addition, the hole current causes an enhanced flow of electrons from the collector point into the base body.

The behavior of the collector contact depends upon electrical forming treatments. Forming involves the passage of reverse currents much larger than the working currents and dissipation of relatively high powers at the collector point. The nature of the mechanism of the forming process is one of the central problems of the point-contact transistor. Experiments have shown that at least two effects of forming can be observed.

Using micromanipulator techniques, Valdes[35] has investigated the area around a formed collector point. His studies show that a portion of the n-type block is converted to p-type by forming. The size of the p-type region depends upon the intensity of the forming process, and may be as large as 2×10^{-2} cm in diameter over the surface and 1.7×10^{-3} cm deep. Normal forming probably produces volumes whose linear dimensions are about three times smaller. The presence of a p-type region can lead to high values of the current multiplication factor at the collector by the "hook" mechanism. In order to invoke this mechanism, Valdes also supposes that there is an n-type region adjacent to the collector point. Under these conditions, holes flowing into the p-layer tend to bias it forward in respect to this n-region at the collector. The theory of current multiplication on this basis can be understood in terms of the mechanisms discussed in the last two sections.[36] In essence, the p-layer acts like the base of an n-p-n transistor and a hole current into this layer produces a much larger electron current from one n-region to the other.

In passing, we may remark that photocells involving this same multiplying action have been studied by Shive.[37] Such cells are highly sensitive and the absorption of one photon may lead to a flow of many hundreds of electronic charges through the device.

It is not entirely certain that in actual transistors the current multiplying mechanism is that of the hook. Another mechanism that is apparently in better accord in some ways with experiment is that of hole trapping.[36] This involves the concept of an imperfection, the trap, which is capable of capturing a hole and restraining it for an appreciable time before either releasing it or permitting it to recombine with an electron. Fig. 24 illustrates the action of hole traps near a collector point. In this region it is supposed that the reverse bias has removed most of the electrons and produced a positive space charge. The positive space charge, in turn, is effective in withdrawing electrons from the metal point contact. Holes flowing into this region will enhance the space charge and provoke an additional electron flow. The triangular symbols represent imperfections that

[31] R. L. Wallace and W. J. Pietenpol, "Some circuits properties and applications of n-p-n transistors," PROC. I.R.E., vol. 40, pp. 753–767; July, 1951. See, also, *Bell Sys. Tech. Jour.*, vol. 30, pp. 530–563; 1951.

[32] R. L. Wallace, L. G. Schimpf, and E. Dickten, *op. cit.*

[33] J. M. Early, "Effects of space charge layer widening in junction transistors," PROC. I.R.E., vol. 40, pp. 1401–1407; this issue.

[34] J. Bardeen and W. H. Brattain, "Physical principles involved in transistor action," *Phys. Rev.*, vol. 75, p. 1208; 1949.

[35] L. B. Valdes, "Transistor forming effects in n-type germanium," PROC. I.R.E., vol. 40, pp. 445–448; April, 1952.

[36] W. Shockley, "Theories of high values of alpha for collector contacts on germanium," *Phys. Rev.*, vol. 78, pp. 294–295; 1950.

[37] J. N. Shive, personal communication.

may be either neutral or positive in charge. If neutral, they may become positive by trapping a hole. If a hole in flowing through the space-charge region spends much more time in traps than it does in motion, then the amount of space charge it produces should be greatly enhanced.
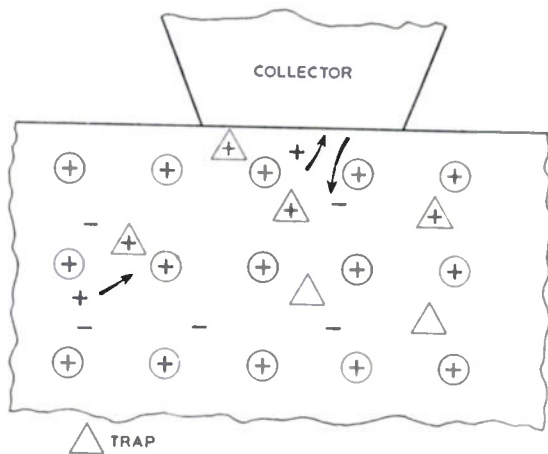
Fig. 24—The enhancement of positive space charge at the collector due to mobile and trapped holes.

The trap mechanism predicts a collector current that is nonlinear in the emitter current. The reason is that when the hole current is large enough, the traps become full and there is no further enhancement of the space charge due to added increments of current. In addition, the ability of the traps to capture and hold the holes will be dependent upon temperature. The dependence of the current multiplication upon emitter current and temperature has been investigated by Sittner.[38] His findings are in agreement with the theory of the trap mechanism insofar as the general trends are concerned. From his data he is able to estimate the density of the traps and their binding energy for holes. The former is about 10^{13} cm^{-3} and the latter is 0.3 electron volts. The nature of the traps, however, is a subject for speculation.

From the point of view of imperfections in nearly perfect crystals, the situation at point contacts is seen to be in a highly unsatisfactory state so far as fundamental knowledge is concerned. From the point of view of manufacturing control, on the other hand, the situation is quite different, and point-contact transistors can be formed to highly precise tolerances.[39]

8. Depletion Regions in *p-n* Junctions

In this section we shall consider some important features of *p-n* junctions that are frequently disregarded in considering the current voltage characteristics. In Section 5 the potential distribution in a *p-n* junction was described. It was shown in Fig. 16 that a dipole layer forms which produces the potential distribution. Within this dipole layer the density of holes and electrons is small compared to the density of chemical imperfections. As a result, the space charge of the donors and

acceptors is not neutralized. If a large reverse bias is applied to a *p-n* junction, then the size of the dipole layer becomes larger as is illustrated in Fig. 25. Within this dipole layer the density of carriers is small and arises from the reverse currents of thermally generated minority carriers in the small *p*-region and the *n*-region. The fact that the dipole layer is depleted of its normal concentration of carriers is of importance in certain applications of transistor electronics. We shall emphasize the importance of the lack of carriers by referring to the layer as a *depletion* region. Such layers are sometimes referred to as *space-charge* layers, and the words emphasize that there is a net charge density within them.
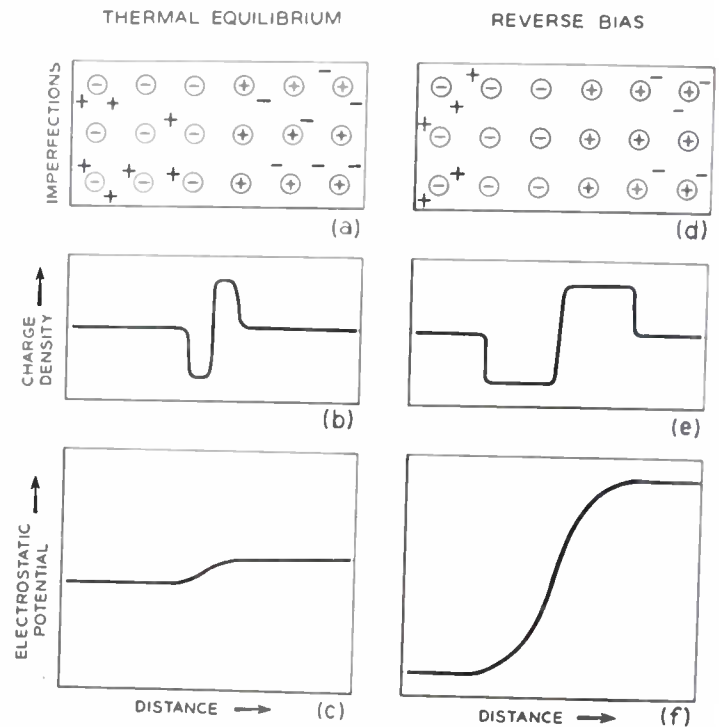
Fig. 25—Depletion region produced by reverse bias in a *p-n* junction.

The observation of depletion layers in *p-n* junctions has been made by both electrical and electromechanical means. The electrical method consists of measuring the capacitance of a *p-n* junction at reverse bias.[40] Under these conditions the depletion layer acts substantially as a region of dielectric. There is a small saturation current flowing across it, but this current does not contribute appreciably to the ac impedance. The ac voltage changes across the layer are produced by flows of carriers in the *p*- and *n*-regions and add their charges at the edges of the depletion region. Thus the *p-n* junction acts like a condenser having the dielectric constant equal to that of germanium and a plate to plate to spacing equal to the thickness of the depletion layer. On the basis of electrostatic theory one can predict how the capacitance should depend on reverse bias for any assumed concentration of donors or acceptors. The theory appears to apply to *p-n* junctions that have been studied within the accuracy of experimental measurement.

[38] W. R. Sittner, "Current multiplication in the type A transistor," Proc. I.R.E., vol. 40, pp. 448–454; 1952.
[39] J. A. Morton, "Present status of transistor development," *Bell Sys. Tech. Jour.*, vol. 31, pp. 441–442; 1952.
[40] G. L. Pearson, W. T. Read, Jr. and W. Shockley, "Probing the space-charge layer in a *p-n* junction," *Phys. Rev.*, vol. 85, p. 1055; 1952.

A much more graphic way of demonstrating the presence of the depletion layer involves the use of a probe point. In this demonstration p-n junctions were biased at voltages as high as 85 volts, and the potential on the surface of the germanium was determined by making contact to it with a very sharp tungsten point. It was found under these conditions that the potential was substantially constant within the n-region and the p-region themselves, and then varied rapidly throughout a narrow region. In Fig. 26 some potential plots obtained by this means are shown.[40] It is to be observed that the width of the depletion region is in some cases greater than 1/1000 of an inch.
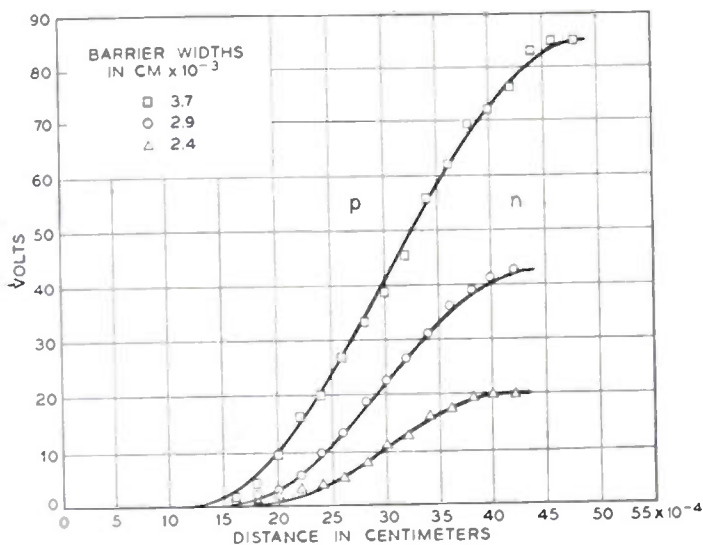
Fig. 26—Probe voltage versus distance through carrier—depletion region.

If the electric field in the depletion region is sufficiently large, then a new effect occurs. As was discussed in Section 1, an electric field tends to displace the valence electrons in respect to the atomic cores. In other words, it tends to break the bonds. This breaking of bonds is aided by a quantum-mechanical effect first discussed by Zener.[41] According to Zener's theory, a sort of spontaneous breaking of the valence bonds occurs before the electric field is large enough to rupture them directly. As a result of this spontaneous breaking induced by strong electric fields, generation of hole electron pairs will occur in the depletion region. Such field-generated currents, or Zener currents, have been observed in p-n junctions.[42] Their effect upon the rectification characteristic is illustrated in Fig. 27. Zener diodes is the name applied to p-n junctions which exhibit such currents. By adjusting the densities of donors and acceptors when the junction is fabricated, it is possible to cause the voltage at which the Zener current becomes important to come at almost any assigned value. Voltages less than 1 volt and as high as 2,000

volts have been observed. A stable Zener current can be drawn from such a junction for an indefinite period, thus making such junctions potentially useful as voltage-regulating or voltage-limiting devices. The temperature coefficient of the critical voltage is, fortunately, quite small.
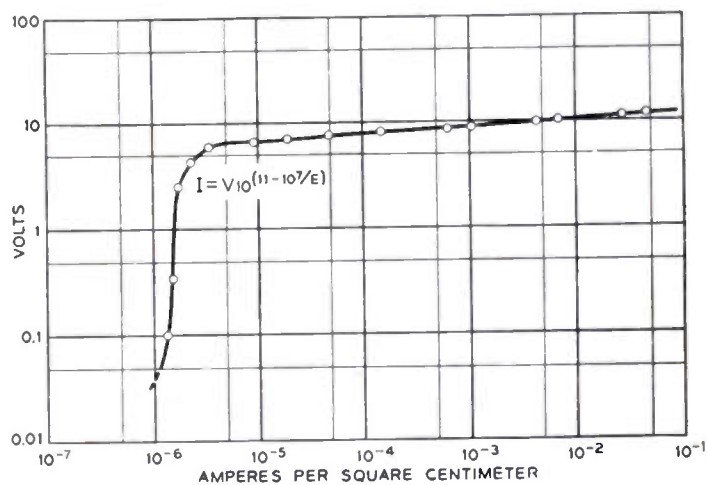
Fig. 27—Zener current in a p-n junction.

9. Unipolar and Analog Transistors

Although transistors and vacuum tubes are different in structure and mode of operation, vacuum-tube concepts have often been used to describe the functioning of transistors. Usually the analogy between the two is rather strained. In this section, however, we shall describe a new class of *unipolar transistors*, some of which are so conveniently described by analogy with vacuum tubes that they are referred to as *analog transistors*. We shall present the material of this section in terms of these analogies.

As a starting point for describing these new transistors by analogy with vacuum tubes, we shall discuss an analogy between a p-n junction biased in the reverse direction and a vacuum condenser. In a vacuum condenser there is a potential difference between the two metal plates. Between the plates there is an electric field, and under ordinary conditions of operation the electron emission from the plates can be neglected. Thus in the vacuum condenser the space between the two plates is free of charge. As we discussed in the last section, when a p-n junction is biased in the reverse direction a region of carrier depletion is formed. If the effect of the small number of carriers in the region of depletion is negligible, then the structure will behave as a condenser. However, in general there will be an appreciable space charge in the depletion region and the electric field will not be the same as in a true vacuum condenser.

A closer approximation to a true vacuum condenser can be produced by suitably adjusting the donor and acceptor concentrations. In Fig. 28 we represent a situation in which the structure consists of a p-type region, a region which is substantially pure or intrinsic, and an n-type region. If an electric field which produces a potential difference is applied across the structure, then

[41] C. Zener, "A theory of the electrical breakdown of solid dielectrics," *Proc. Roy. Soc.* (London), vol. 145, p. 523; 1934.

[42] K. B. McAfee, E. J. Ryder, W. Shockley, and M. Sparks, "Observations of Zener current in germanium p-n junctions," *Phys. Rev.*, vol. 83, p. 650; 1951.

the distribution of charge will be as shown in Fig. 28. In this case the dipole layer consists of two separated distributions of charge much as it does in the case of a metal parallel plate condenser. Thus we see in this example that the *p*- and *n*-regions act as the metal condenser plates and the substantially pure region as the vacuum. This analogy will hold so long as reverse potentials are applied. For forward potentials, however, there will be injection of majority carriers from both sides into the middle region.
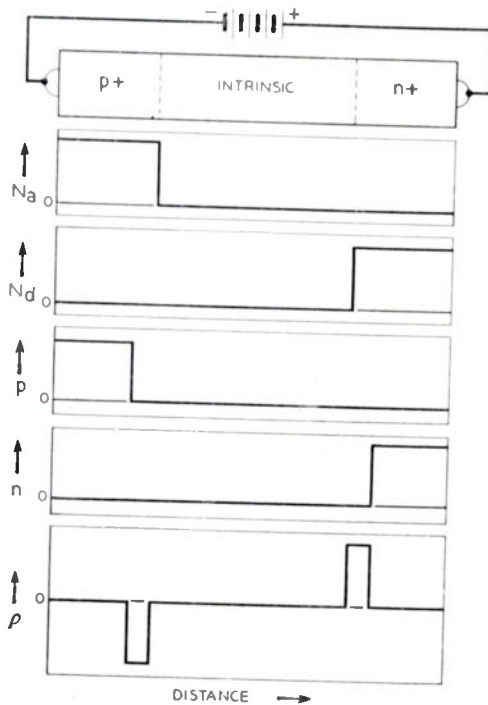
Fig. 28—Semiconductor analog to a vacuum condenser.

From the reasoning presented above, it is evident that heavily doped *p*- and *n*-regions are like cold non-emitting electrodes in a vacuum-tube structure, provided that the electrical fields at their surfaces are properly chosen. The choice must be such that the electric fields at the boundaries of the regions are *majority carrier retaining* so that the *p*- and *n*-type regions will furnish only small saturation currents of minority carriers. For suitably designed structures, these small currents can be neglected. On the basis of these ideas we can understand how a transistor may be made whose mode of operation will be closely analogous to that of a vacuum-tube triode. We shall refer to such a transistor as an analog transistor.

The mode of operation of the analog transistor may be described by reference to the state of affairs in the vacuum-tube triode. If normal operating voltages are applied to a vacuum-tube triode whose cathode is cold and nonemitting, then at the grid and cathode the electric field is electron extracting and at the plate it is electron retaining. If the cathode is heated so that electron emission occurs, then the field at the cathode is

reduced by the space charge of the emitted electrons. However, the fields at the electrodes still remain electron extracting at grid and cathode and electron retaining at the plate.

In the analog transistor the geometrical structure of the vacuum tube is converted to semiconductor, as represented in Fig. 29, by replacing the cathode and anode by *n*+ regions, the grid by a *p*+ region, and the space by pure germanium. Potentials of the same polarities are applied. Under these conditions the analog-grid will draw only a small saturation current since the field at its surface is hole retaining so that the majority carriers do not flow out of it. Likewise, the majority carriers are retained in the analog-plate. At the analog-cathode, however, the field is such as to extract majority carriers. Electrons will thus flow out of the analog-cathode, setting up a negative space charge. Since the analog-grid is negative in respect to the analog-cathode, none of these electrons will reach it and the grid will be surrounded by a depletion region. The electron flow between the grid "wires" can evidently be controlled by a signal voltage applied to the analog grid. Thus a mode of operation closely analogous to that of a vacuum tube is achieved.

A space-charge law strictly similar to Child's Law can be derived for the space-charge limited emission of the *n*+ region of Fig. 29. For this purpose, as in the case of Child's Law, the complex geometry of Fig. 29 is replaced by a plain parallel model. Under these conditions one finds that the current emitted per unit area is given by the formula

$$I = 9\kappa\epsilon_0\mu V^2/8x^3, \tag{55}$$

where κ and ϵ_0 are the dielectric constant of germanium and the permitivity of free space, respectively, μ is the mobility of an electron, V is the voltage between the parallel planes, and x is their spacing. In similar notation the conventional expression for Child's Law is

$$I = (4/9)\epsilon_0(2q/m)^{1/2}V^{3/2}/x^2. \tag{56}$$

The analog for Child's Law governs the current flow in the analog transistor in much the same way that Child's Law does in the case of a vacuum tube.

It should be pointed out, however, that there are important differences between the two laws. In the case of the analog of Child's Law it should be noted that there is no conservation of momentum of the electrons. After travelling one mean-free path they are scattered so that no coherent motion is preserved in the direction of the electric field from one collision to the next. One interesting consequence of this is that electron flow to the analog-grid need not occur when the analog-grid is positive in respect to the analog-cathode, provided that the field is electron extracting at the analog-grid. It should also be pointed out that the analog of Child's Law presented above is valid only for relatively low electric fields. For high electric fields the relationship between drift velocity and electric field ceases to be linear, with

he result that, in effect, the mobility decreases.[43] This act must also be taken into account in designing transistors making use of space-charge limited emission.
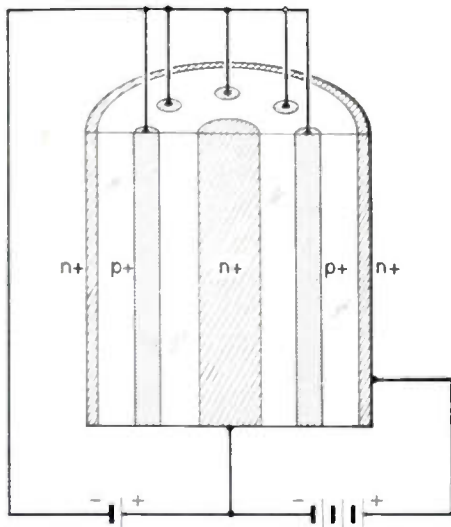
Fig. 29—Compositional structure of an analog transistor.

The transistor of Fig. 29 is a strict analog of a vacuum-tube triode so far as the polarities and nature of the electron currents are concerned. It is possible of course to have an anti-analog transistor in which the roles of donors and acceptors are interchanged. In such an anti-analog transistor the source of holes would be the anode and the electrode to which they flow would be the cathode. This situation poses certain difficulties in nomenclature for analog transistors. It is evident that the words cathode and anode are unsuitable since the role played by an electrode will be opposite in a hole-flow analog or *p*-type analog transistor from what it is in an electron-flow or *n*-type analog transistor.

It is evident that the analog transistor works according to principles somewhat different from those of the ordinary type of transistor. As was mentioned above, the phenomenon of minority carrier injection into a region where majority carriers are present does not occur. Thus the conventional role assigned to an emitter electrode in an ordinary transistor is not played by any of the electrodes of the analog transistor. Furthermore, except for ac charging currents and small reverse saturation currents, practically all of the current flowing in the transistor is carried by carriers of one sign. This is different from the point-contact transistor and the filamentary transistor in which both signs are involved to a significant degree. In the junction transistor, also, both controlled and controlling currents flow in the base layer at the same time. Thus a difference between the analog transistor and the more conventional type is a separation in space of the two kinds of current flow.

In order to distinguish between the more conventional transistors and the analog types, we propose to use the words *bipolar* and *unipolar*. In terms of this

[43] See footnotes 10, 11.

terminology the *analog* transistor structures are the *unipolar* type. This is true for both the type in which the current is carried by electrons and the type in which the current is carried by holes. Elsewhere in this issue we discuss another type of unipolar transistor, referred to as the *field-effect* transistor.[44] This transistor differs from the analog type in having unbalanced chemical impurity densities in the region where the carriers flow. These chemical impurities neutralize some of the space charge of the current carriers in a way somewhat reminiscent of the behavior of positive gas ions in a gas-discharge tube.

Since the role of emitter is not played in the normal way in the unipolar transistor and since the collector also functions somewhat differently, it appears advantageous to introduce new terminology for the electrodes in the unipolar types. The choice proposed for the electrodes is as follows: *Source* for the electrode from which the carriers enter the region of relatively high electric fields; *drain* for the electrode at which they arrive and out of which they flow; in the analog transistor the control electrode will be called the *grid* because of its close analogy with vacuum-tube structures. In the field-effect transistor it is proposed to call the control electrode the *gate*. The fact that gate and grid have the same initial letter leads to the use of a common subscript for these two similarly functioning electrodes. The choice of these names has been based partly on an attempt to find names which describe functions and partly on the value of the names from a phonetic and abbreviational point of view. It should be noted that none of the new subscripts are the same as those encountered in bipolar transistors. Furthermore, it may be noted that the names selected are all monosyllabic.

CONCLUSION

This article has attempted to show how the basic mechanisms of current flow in semiconductors arise from imperfections. Another purpose has been to show that the theoretical concepts of the hole and the electron, known for over two decades since the pioneering theories of Wilson were published, have reached the stage of acquiring quantitative attributes usable for engineering design. This quantitative status has become possible largely as a result of new experiments in transistor physics made possible by the growth of transistor technology. Although not all of the physical principles that may have engineering utility have been discussed at length, enough examples have been presented to show that transistor electronics is a large and diversified field. New combinations of principles may be expected to lead to new and useful devices. One new class of devices is constituted by the unipolar transistors, first published in this article. The conclusion is that transistor electronics is a large and diverse field. It may be expected to show rapid growth for many years to come.

[44] W. Shockley, "A unipolar 'field-effect' transistor," PROC. I.R.E. vol. 40, pp. 1365–1377; this issue.

Present Status of Transistor Development*

J. A. MORTON†, SENIOR MEMBER, IRE

Summary—The invention of the transistor provided a simple, apparently rugged device that could amplify—an ability which the vacuum tube had long monopolized. As with most new electron devices, however, a number of extremely practical limitations had to be overcome before the transistor could be regarded as a practical circuit element. In particular, the reproducibility of units was poor—units intended to be alike were not interchangeable in circuits; the reliability was poor—in an uncomfortably large fraction of units made, the characteristics changed suddenly and inexplicably; and the "designability" was poor—it was difficult to make devices to the wide range of desirable characteristics needed in modern communications functions. This paper describes the progress that has been made in reducing these limitations and extending the range of performance and usefulness of transistors in communications systems. The conclusion is drawn that for some system functions, particularly those requiring extreme miniaturization in space and power as well as reliability with respect to life and ruggedness, transistors promise important advantages.

INTRODUCTION

WHEN THE TRANSISTOR was announced not quite four years ago, it was felt that a new departure in communication techniques had come into view. Here was a mechanically simple device which could perform many of the amplification functions over which the electron tube had long held a near monopoly. The device was small, required no heater power, and was potentially very rugged; moreover, it consisted of materials which might be expected to last indefinitely long, and it did not appear to be too complicated to make.

However, as might be expected for a newly invented electron device, the practical realization of these promises still required the overcoming of a number of obstacles. While the operation of the first devices was well understood in a general way, several items were limiting and puzzling. For example,

(a) Units intended to be alike varied considerably from each other—*the reproducibility was bad*.

(b) In an uncomfortably large fraction of the exploratory devices, the properties changed suddenly and inexplicably with time and temperature, whereas other units exhibited extremely stable characteristics with regard to time—*the reliability was poor*.

(c) It was difficult to use the theory and then existing undeveloped technology to develop and design devices to a varied range of electrical characteristics needed for different circuit functions. Performance characteristics were limited with respect to gain, noise figure, frequency range, and power—*the designability was poor*.

Before the transistor could be regarded as a practical circuit element, it was necessary to find out the causes

of these limitations, to understand the theory, and to develop the technology further in order to produce and control more desirable characteristics.

Over the past two years measurable progress has been made in reducing, *but not eliminating*, the three listed limitations.

These advances have been obtained through an improved understanding, improved processes, and, very importantly, through improved germanium materials. As a result,

(a) The beginnings of method have evolved in the use of the theory to explain and predict the electrical network characteristics of transistors in terms of physical structure and material properties.

(b) It is now possible to evaluate some of the effects and physical meaning of empirically derived processes, and thereby to devise better methods subject to control. Previously, inhomogeneities in the material properties masked the dependence of the transistor electrical properties even on bulk properties (such as resistivity) as well as on processing effects.

(c) On an exploratory development level, it is now possible to make transistors in the laboratory to several sets of prescribed characteristics with usable tolerances and satisfactory yields.

(d) Such transistors are greatly improved over the old ones in so far as life and ruggedness are concerned, and some reduction in temperature dependence has been achieved. However, it is not to be inferred that all reliability problems are solved.

(e) It has become possible in the laboratory to explore experimentally some of the consequences of the theory, with the result that point-contact devices with new ranges of performance are indicated. Even more importantly, new *p-n* junction devices have been built in the laboratory, and these junction devices have indicated an extension in several performance characteristics.

(f) By having interchangeable and reliable devices with a wider range of characteristics, it has become possible to carry on exploratory circuit and system applications on a more realistic basis. Such applications effort is, in turn, stimulating the development of new devices towards new characteristics needed by these circuit and system studies.

It is the purpose of the remainder of this paper to give an over-all but brief summary of recent progress made at Bell Telephone Laboratories in reducing the above-mentioned limitations on reproducibility, reliability, and performance. Since a fair number of device types are currently under development, each with different characteristics to be optimized, the data will be presented as a sort of montage of characteristics of several different device types. It is not to be inferred that any

one type of transistor combines all of the virtues any more than such a situation exists in electron-tube art. Moreover, it will be impossible in a paper of practical length to present complete detailed characteristics on all or even several of these devices under development; nor would it be appropriate since most of these data are on devices currently under development. Rather, what is desired is a summary of progress across the board to give the reader an integrated and up-to-date picture of the current state of transistor electronics.

Reproducibility Status

A. Description of Transistors

Before quantitative data comparing the characteristics of past with present transistors are presented, it will be useful to briefly review physical descriptions of the various types of transistors to be discussed. Fig. 1 shows a cutaway view of the now familiar point-contact cartridge-type transistor. All of the early transistors
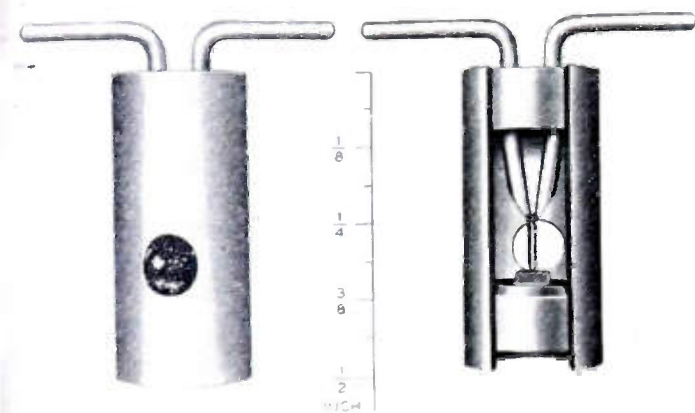
Fig. 1—The Type A transistor structure.

were of this general construction, and the characteristics of a particular one, called the Type A,[1] will be used as a reference against which to measure results now obtainable with new types under current development. Fig. 2 is a semischematic picture of the physical operation of such a device. Pressing down upon the surface of a small die of *n*-type germanium are two rectifying metal electrodes, one labelled *E* for emitter, the other *C* for collector. A third electrode, the base, is a large-area ohmic contact to the underside of the die of germanium. The emitter and collector electrodes obtain their rectifying properties as a result of the *p-n* barrier (indicated by the dotted lines) existing at the interface between the *n*-type bulk material and small *p*-type inserts under each point. When the collector is biased with a moderately large negative voltage (in the reverse direction) so that the collector barrier has relatively high impedance, a small amount of reverse current flows from the collector to the base in the form of electrons as indicated by the small black circles. Now, if the emitter is biased a few tenths of a volt positively in the forward direction, a current of

[1] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, p. 367; 1949.

holes (indicated by the small open circles) is injected from the emitter region into the *n*-type material. These holes are swept along to the collector under the influence of the field initially set up by the original collector electron current, thus adding a controlled increment of collector current. Because of their positive charge these holes can lower the potential barrier to electron flow
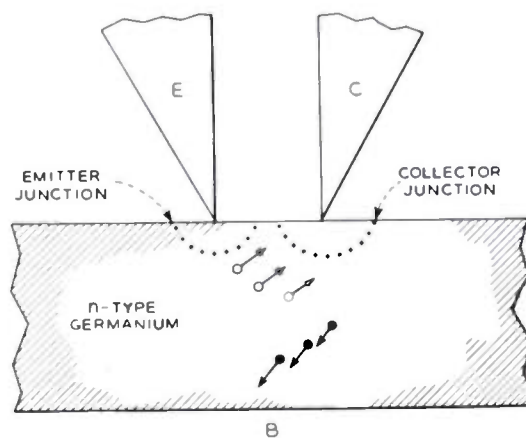
Fig. 2—Schematic diagram of a point-contact transistor.

from collector to base, and thus allow several electrons to flow in the collector circuit for every hole entering the collector barrier region. This ratio of collector-current change to emitter-current change for fixed collector voltage is called "alpha," the current gain. In point-contact transistors alpha may be larger than unity. Since the collector current flows through a high impedance when the emitter current is injected through a low impedance, voltage amplification is obtained as well.

Some of the new transistors are point-contact transistors similar in physical appearance to the Type A.
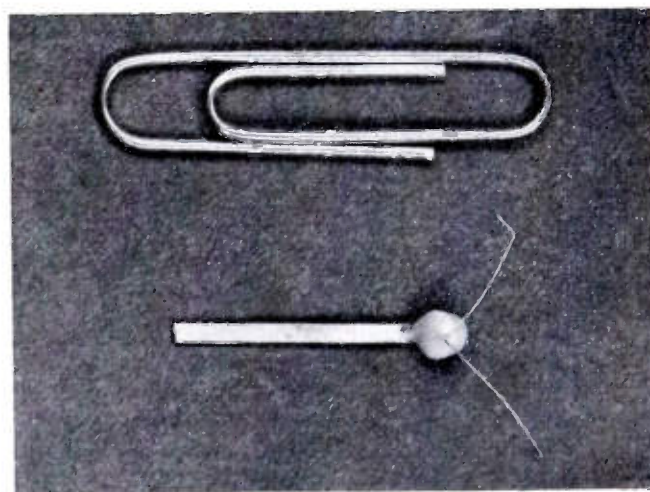
Fig. 3—M1689 point-contact transistor.

However, their electrical characteristics will be shown to be significantly improved over the old Type A, not only in-so-far as reproducibility and reliability are concerned but also as to range of performance.

For use in miniature packaged circuit functions, the point-contact transistor has been miniaturized to contain only its bare essentials. Fig. 3 is a photograph of a
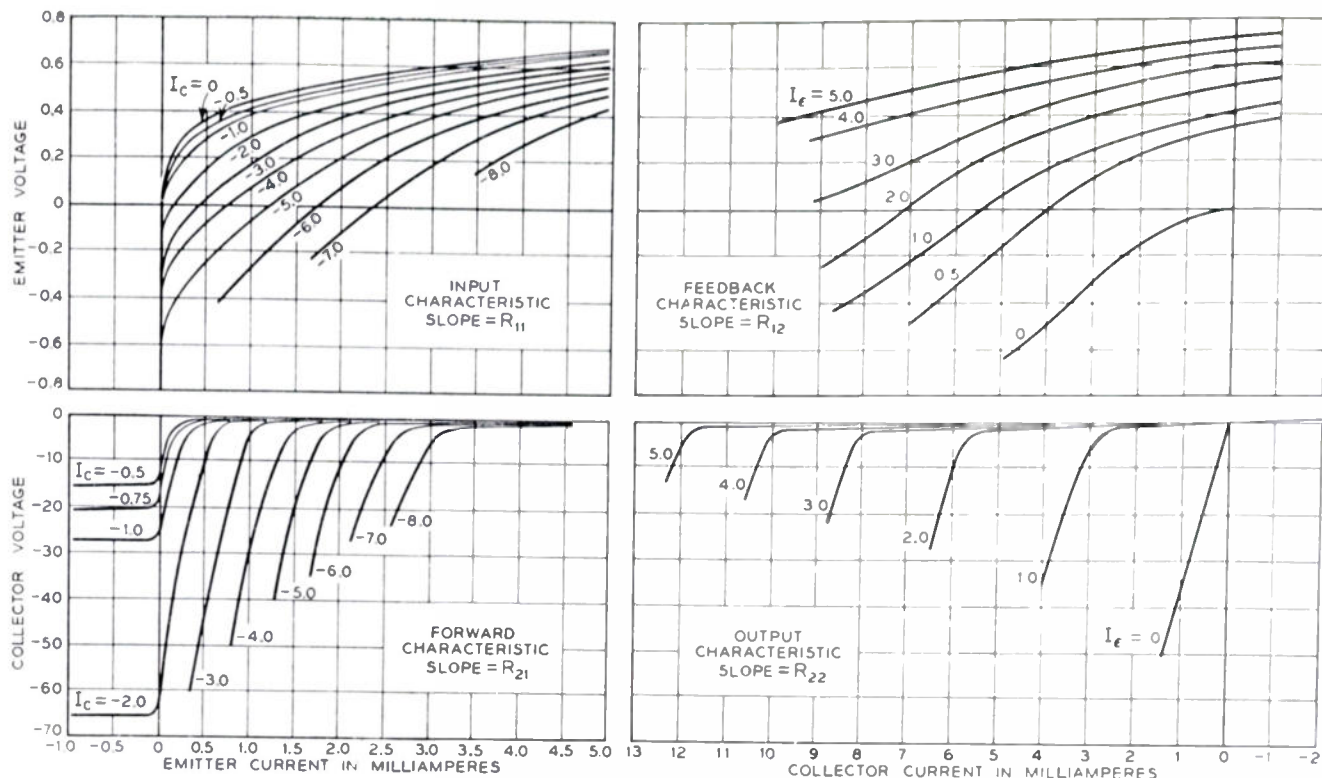
Fig. 4—Static characteristics of the M1689 transistor.

so-called "bead" transistor (compared to a paper clip for size) and several of the current development types are being made in this form.

In Fig. 4 is shown the family of static characteristics representative of the M1689 bead-type transistor. Note in particular the collector family which gives the dependence of collector voltage upon collector current with emitter current as parameter. These characteristics may be thought of as the dual to the plate family of a triode.[2] The slope of these curves is very nearly the small-signal ac collector impedance of the transistor.[3]
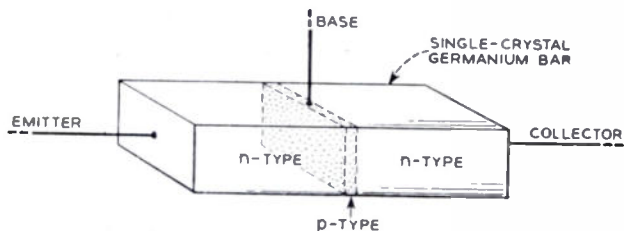
Fig. 5—The n-p-n junction transistor.

For a fixed collector voltage of −20 volts, when the emitter current is changed from zero to 1 ma, note that the collector current correspondingly changes slightly more than 2 ma, indicating a current gain, alpha, of slightly more than two.

Newest member of the transistor family, recently described by Shockley, Sparks, Teal, Wallace, and Pieten-

pol, is the n-p-n junction transistor.[4,5] Fig. 5 is a schematic diagram of such a structure. In the center of a bar of single-crystal n-type germanium there is formed a thin layer of p-type germanium as part of the same single crystal. Ohmic nonrectifying contacts are se-
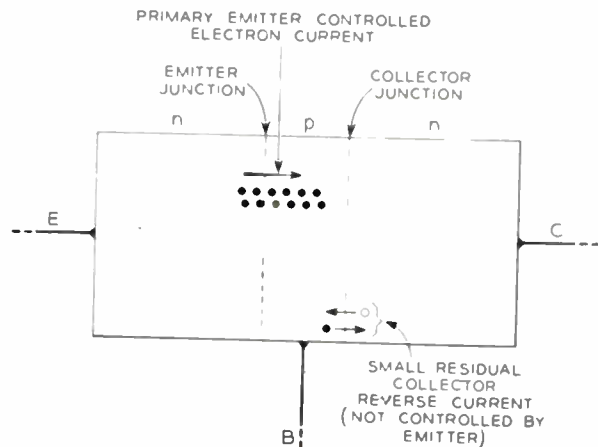
Fig. 6—Schematic diagram of a junction transistor.

curely fastened to three regions as shown, one being labelled "emitter," one "base," and one "collector." In many simple respects, except change in conductivity type from p-n-p in point-contact (see Fig. 2) to n-p-n in the junction type, the essential behavior is similar.

As shown in Fig. 6, if the collector junction is biased in the reverse direction, i.e., electrode C biased positively with respect to electrode B, only a small residual

[2] R. L. Wallace and G. Raisbeck, "Duality as a guide in transistor circuit design," *Bell Sys. Tech. Jour.*, vol. 30, p. 381; 1951.
[3] As shown by Ryder and Kircher,[1] the ac collector impedance, $r_c = R_{22} - R_{12}$, where R_{22} is the open-circuited output impedance and R_{12} is the open-circuit feedback impedance. Usually, $R_{22} \gg R_{12}$.

[4] W. Shockley, M. Sparks, and G. K. Teal, "P-N transistors," *Phys. Rev.*, vol. 83, p. 151; 1951.
[5] Wallace and Pietenpol, "Some circuit properties and applications of n-p-n transistors," *Bell Sys. Tech. Jour.*, vol. 30, p. 530; 1951.

back current of holes and electrons will diffuse across the collector barrier as indicated. However, unlike the point-contact device, this reverse current will be very much smaller and relatively independent of the collector voltage because the reverse impedance of such bulk barriers is so many times higher than that of the barriers produced near the surface in point-contact transistors. Now again, if the emitter barrier is biased in the forward direction (a few tenths of a volt negative with respect to the base is adequate), then a relatively large forward current of electrons will diffuse from the electron-rich *n*-type emitter body across the reduced emitter barrier into the base region. If the base region is adequately thin so that the injected electrons do not recombine in the *p*-type base region (either in bulk or on the surface), practically all of the injected emitter current can diffuse to the collector barrier; there they are swept through the collector barrier field and collected as an increment of controlled collector current. Hence, again, since the electrons were injected through the low forward impedance and collected through the very high reverse impedance of bulk type *p-n* barriers, very high voltage amplification will result. No current gain is possible in such a simple bulk structure and the maximum attainable value of alpha is unity. However, because the bulk barriers are so much better rectifiers than the point surface barriers, the ratio of collector reverse impedance to emitter forward impedance is many times greater, more than enough to offset the point-contact higher alpha; thus, the junction unit may have much larger gain per stage.[1,4,5] Fig. 7 is a photograph of a developmental model of such a junction transistor called the M1752.
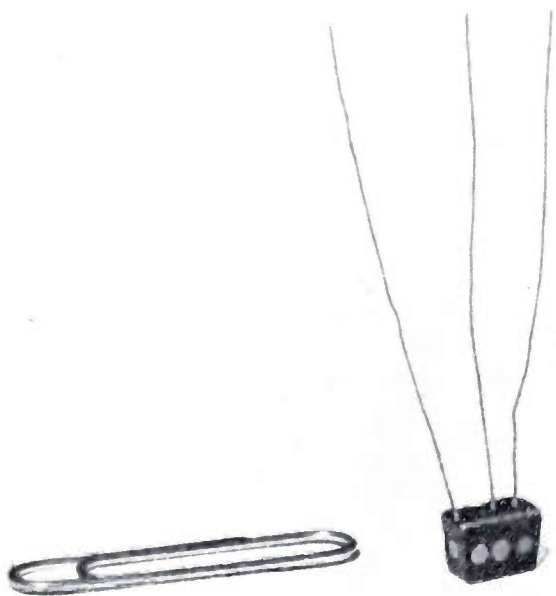
Fig. 7 —The M1752 junction transistor.

The upper part of Fig. 8 is a collector family of static characteristics for the M1752 *n-p-n* junction transistor. By way of comparison to those of the point-contact

family, note the much higher reverse impedance of the collector barrier (relatively independent of collector voltage) and the correspondingly smaller collector cur-
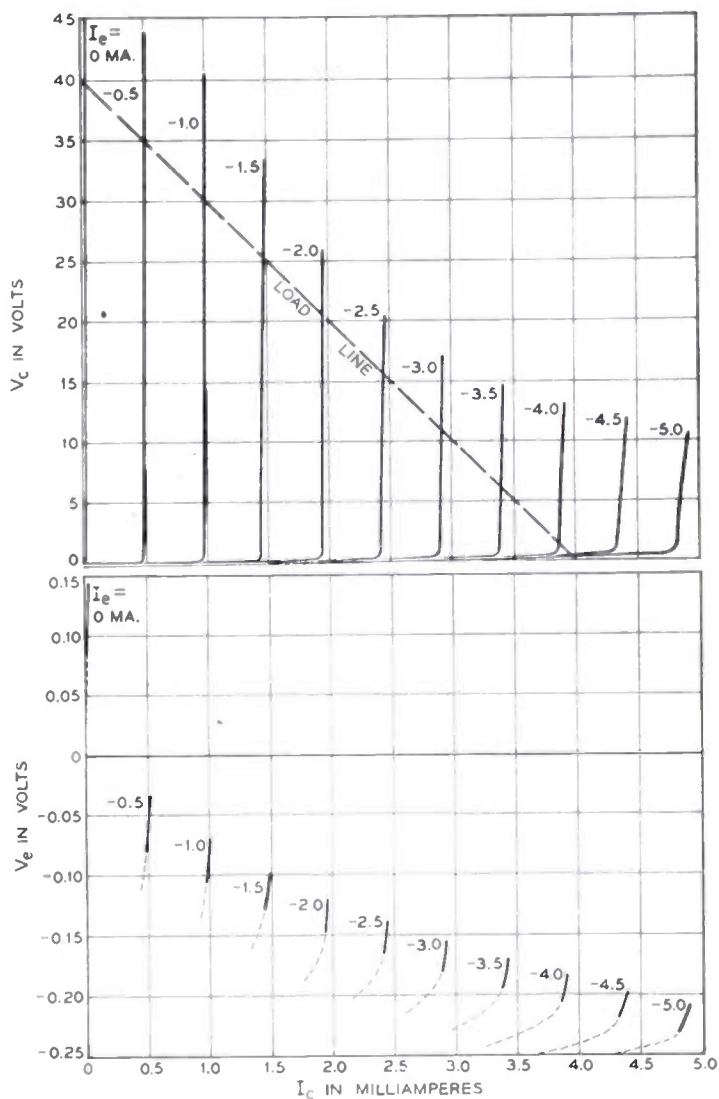
Fig. 8—Static characteristics of the M1752 junction transistor.

rents when the emitter current is zero. In fact, Fig. 9 is an expanded plot of the lower-left rectangle of the collector family of Fig. 8. The almost ideal straight-line character and regular spacing of these curves persists down to voltages as low as 0.1 volt and currents of a few microamperes. Thus, essentially linear Class A amplification is possible for as little collector power as a few microwatts. Constant collector power-dissipation curves of 10, 50, and 100 μw are shown dotted for reference.

B. Reproducibility of Linear Characteristics

In describing progress in the reproducibility of those transistor characteristics pertinent to small-signal linear applications, one possible method is to give the statistical averages and dispersions in the linear open-circuit impedances of the transistor as defined by Ryder and Kircher.[1] Such a procedure, of course, implies a state of

statistical control in the processes leading to a reasonably well-behaved normal distribution for which averages and control limits can be defined. This situation can be said to be in effect for most transistors under current development.
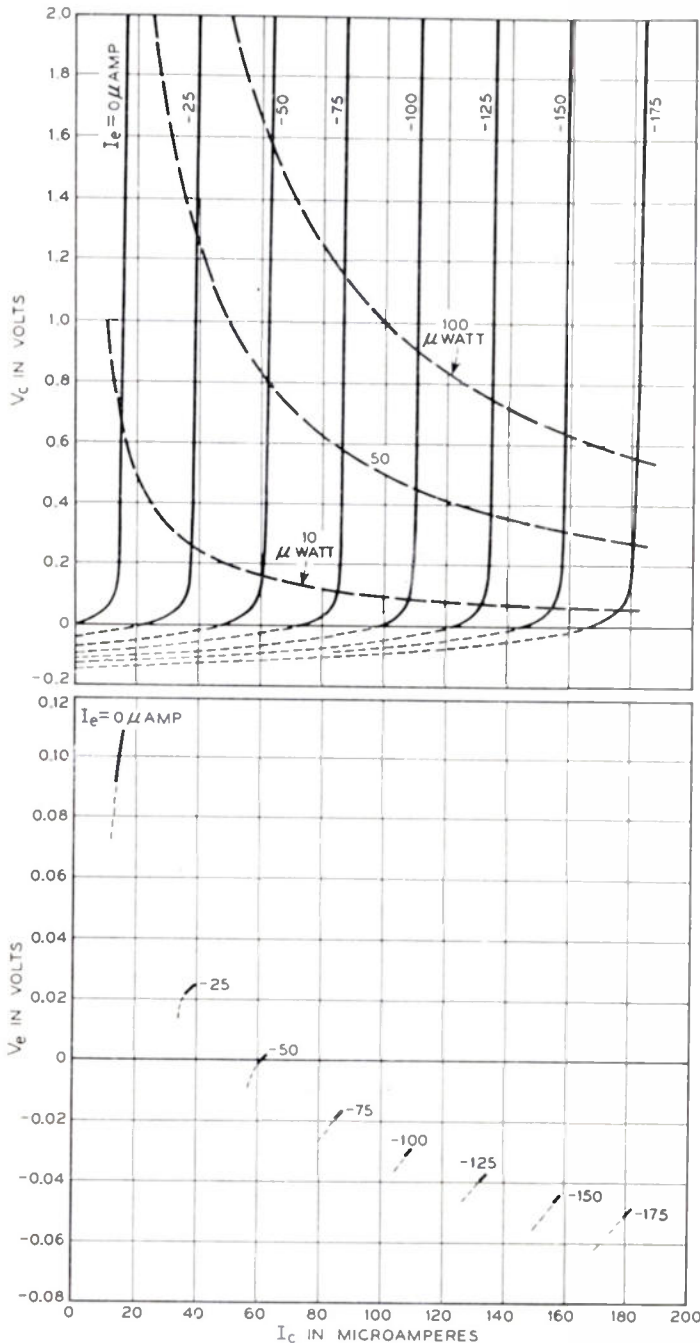
Fig. 9—Expanded plot of the microwatt region of the static characteristics of the M1752 transistor.

However, for the old Type A unit, control simply was not in evidence; so that in quoting figures on Type A's, ranges for commensurate fractions of the total family will be given. In order that symbols and terminology will be clear, it will be useful to review briefly the method of defining the linear characteristics of all transistors. In Fig. 10 is shown a generalized network representing the transistor in which the input terminals are emitter-base

and the output terminals are collector-base. Then, over a sufficiently small region of the static characteristics, the linear relations between the incremental emitter and collector voltages and currents may be represented by

$$v_e = i_e z_{11} + i_c z_{12}$$
$$v_c = i_e z_{21} + i_c z_{22}$$
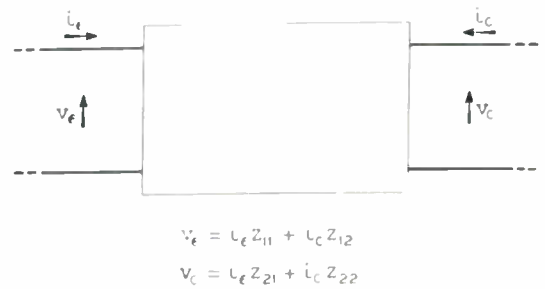
Fig. 10—The general linear transistor.

the pair of linear equations shown.[1] The coefficients are simply the open-circuit driving point and transfer impedances of the transistor, or the slopes of the appropriate static characteristics at fixed dc operating currents. These equations may be represented by any one of a large number of equivalent circuits of which one shown in Fig. 11 is perhaps currently most useful. In this circuit r_e is very nearly the ac forward impedance

$$\bar{r}_e = 250 \text{ OHMS} \qquad \bar{r}_c = 20,000 \text{ OHMS}$$
$$\bar{r}_b = 250 \text{ OHMS} \qquad \bar{a} = 2$$
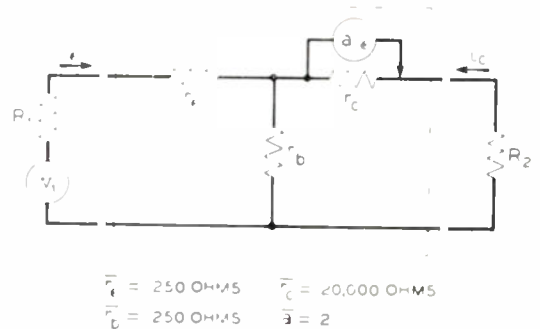
Fig. 11—Equivalent circuit and average element values of the Type A transistor.

of the emitter barrier, r_c is very nearly the ac reverse impedance of the collector barrier, r_b is the feedback impedance of the bulk germanium common to both, and a is the circuit current gain representing carrier collection and multiplication, if any. It turns out this is very nearly equal to the current multiplication factor a of the collector barrier mentioned before. Average values of these elements for the Type A transistor are given in Fig. 11. In Fig. 12 are given the ranges of these param-

ELEMENT	RANGE SEPTEMBER 1949	RANGE JANUARY 1952
a	4 : 1	±20%
r_c	7 : 1	±30%
r_e	3 : 1	±20%
r_b	7 : 1	±25%

Fig. 12—Reproducibility of point-contact linear characteristics.

eters for the Type A, as of September, 1949, and the control limits[6] for the same characteristics for new point-contact transistors now under development. For September, 1949 the ranges are taken about the average values shown in Fig. 11 for the Type A transistor. The control limits given for the present situation apply to a number of different types of point-contact transistors so that the present average values of these equivalent circuit elements depend upon the type of transistor considered. In Fig. 13 are given the average values of the

TYPE	M 1729	M 1752
r_e	120	25
r_b	75	250
r_c	15.000	5×10^6
a	2.5	0.95

Fig. 13—Average characteristics of the M1729 and typical characteristics of the M1752 transistors.

characteristics of the M1729 point-contact video amplifier transistor which bears the closest resemblance to the older Type A transistor. By way of contrast are given some typical values of the elements for the M1752 junction transistor which is not yet far enough along in its development to have design centers fixed nor reliable dispersion figures available.

As Ryder and Kircher have shown,[1] transistors in the grounded-base connection may be short-circuit unstable if $a > 1$ and r_b is too large, since r_b appears as a positive feedback element. The curve in Fig. 14 is a plot of the short-circuit stability contour when r_e and r_c have the
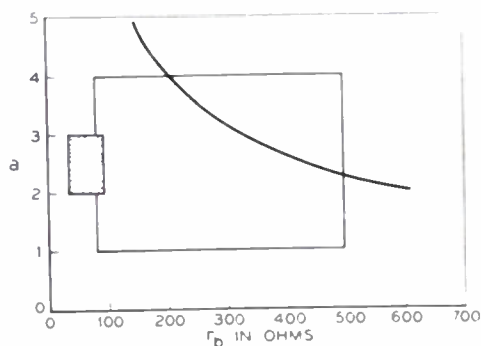
Fig. 14—Stability contour and ranges of a and r_b.

nominal values of 700 and 20,000 ohms. Transistors having a and r_b sufficiently large to place their representative points above this contour will be short-circuit unstable, i.e., they will oscillate when short-circuited, Those having an $a - r_b$ point below the stability contour will be unconditionally stable under any termination conditions. The large unshaded rectangle bounds those

[6] "Quality control of materials," ASTM Manual, pt. III, pp. 55–114; January, 1951.

values of a and r_b which were representative of the Type A transistor in September, 1949. It is apparent that the circuit user of Type A units had approximately a 50-per cent chance of obtaining a short-circuit unstable unit from a large family of Type A units. The smaller shaded rectangle bounds the values of a and r_b now realized in the M1729 transistor presently under development. Not only has the spread in characteristics been greatly reduced as shown, but also the design centers have been moved to a region for which all members of the M1729 family are unconditionally stable.

It is of interest to note that spreads of the order of ± 20 to ± 25 per cent are of the same magnitude as those dispersions now existing among the characteristics of presently available well-controlled electron tubes. These kinds of data on reproducibility of the linear equivalent circuit element values hold for practically all classes of point-contact devices now under development for cw transmission service. While it is too early to prove that such a situation pertains as well to junction transistors, there is every reason to expect similar results after a suitable development period.

C. Reproducibility of Large-Signal Characteristics for Pulse Application

When electron devices are employed for large-signal applications, particularly those of switching and computing, it is well known that the characteristics must be controlled over a very broad range of variables from cutoff to saturation. In September, 1949 very little attempt was made to control such pulse-use characteristics. In the intervening time, transistor circuit studies have proceeded to the point where it is possible to define certain necessary large-scale transistor characteristics which, if met, permit such transistors to be used interchangeably and reproducibly in a variety of pulse circuit functions, such as binary counters, bit registers, regenerative pulse amplifiers, pulse-delay amplifiers, gated amplifiers, and pulse generators. Moreover, it has been possible to meet these requirements on a developmental level with good yields in at least three types of point-contact switching transistors. The scope of this paper will not permit a detailed accounting of the technical features of this situation, and such an account will be forthcoming in future papers on these particular studies. However, a brief description of some of the more important pulse characteristics and their tolerances is certainly pertinent.

In practically all of the transistor pulse-handling circuits examined to date, one characteristic common to all is the ability of the transistor, by virtue of its current gain, to present various types of two-state negative-resistance characteristics at any one or all of its pairs of terminals. A typical simple circuit and corresponding characteristic is shown in Fig. 15 for the emitter-ground terminals when a sufficiently large value of resistance is inserted in the base to make the circuit unstable. In region I where the emitter is negative, the input resistance

is essentially the reverse characteristic of the emitter as a simple diode. In region II as the emitter goes positive, alpha, the current gain rises rapidly above unity. If R_b is sufficiently large and alpha, the current gain, is greater than unity, the emitter to ground voltage will begin to fall because of the larger collector current increments driving the voltage of the node N negative more rapidly than the emitter current drop through r_e would normally carry it. This transition point is called the "peak point." If then $a(r_b+R_b)$ is sufficiently large, in this sense, the input resistance may be negative in this region II. When the internal node voltage has fallen to a value near that of the collector terminal, the "valley point" has been reached. At this point, the emitted hole current has reduced the collector impedance to a minimum value beyond which a is essentially zero; the transistor is said to be saturated. From this point on the input impedance again becomes positive and is determined almost en-

in a given circuit (with different transistors of the same type) the following points of the characteristic:

(a) The off impedance of the emitter—he desires that this be greater than a certain minimum.

(b) The peak point V_{ep}—he desires that this be smaller than a certain maximum.

(c) The value of the negative resistance—he desires that this be greater than a certain minimum.

(d) The valley point V_{ea}, I_{ea}—he desires that these be greater than certain minima.

(e) The slope in region III—he desires that this be smaller than a certain maximum so that he may control it by external means.

Fig. 16—Bistable circuit and characteristics showing trigger voltage requirements.

It may be shown that these conditions can be satisfied for useful circuits by specifying certain maximum and minimum boundaries on the static characteristics. Fig. 17 is an idealized set of input or emitter characteristics. By specifying a minimum value for the reverse

Fig. 15—Emitter-ground negative-resistance circuit and characteristics.

tirely by the base and emitter impedances. By terminating the emitter-ground terminals in various ways with resistor-capacitor-bias combinations, such a network can be made to perform monostable, astable, or bistable functions. Under such conditions, the emitter current, and correspondingly the collector current, switch back and forth between cutoff and saturation values. For example, Fig. 16 shows a value of emitter bias and load resistance such that there are three possible equilibrium values of emitter current and voltage. It may be shown that the two intersections in regions I and III are stable whereas that in region II is unstable. Hence, if the stable equilibrium is originally in I, a small positive pulse Δ_p applied to the emitter will be enough to switch from stable point I to stable point II and, conversely $-\Delta_v$ will carry it from the high current point to the low current point. The circuit designer is interested in reproducing

Fig. 17—Idealized emitter characteristics—slope $= R_{11}$.

resistance in region I, condition (a) above is satisfied. By specifying a maximum slope in region II, III, condition (e) is satisfied. Now refer to the idealized collector

family in Fig. 18. By specifying a maximum value to V_{C3}, it is possible to insure condition (d); and by specifying a minimum value for r_{eo}, condition (b) can be satisfied.
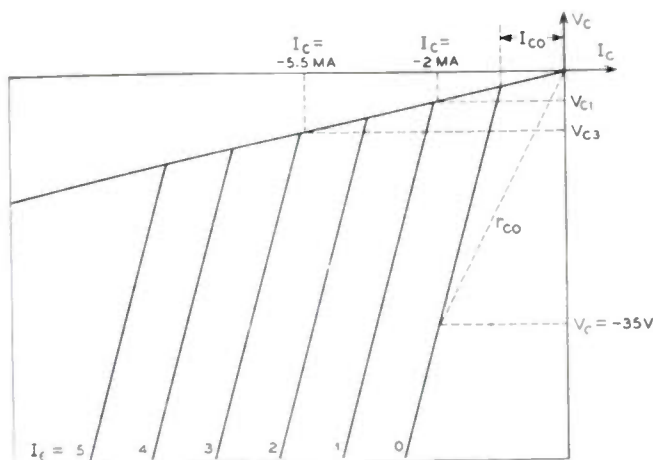
Fig. 18—Idealized collector characteristics.
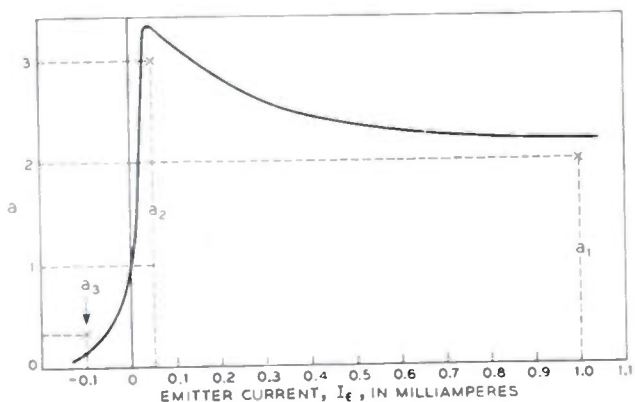
Fig. 19—Effective alpha characteristics.

fied. Finally, in Fig. 19, by demanding that alpha, as a function of I_e, go through a transition from a negligible value (at small negative I_e) to a value well in excess of unity (at a correspondingly small positive value of I_e) and maintain its value well in excess of unity at large values of I_e, conditions (b) and (c) can be met.

Fig. 20 gives the characteristic specifications which must be met by the M1689 bead-type switching transistor now under development. With these kinds of limits, circuit users find it possible to interchange such M1689 units in various pulse circuits and obtain over-all circuit behavior reproducible to the order of about ±2 db.

Reliability Status

A. Life

Reliability figures of merit are not too well defined for electron tubes, and the same situation certainly holds at present for transistors. However, insofar as these quantities can be presently defined, Fig. 21 shows a comparison between the present status and that in September, 1949. Estimates of the half-life of a statistical family of devices are, at best, arbitrary and necessarily amount to extrapolations of survival curves, assuming that a known

survival law will continue to hold.[7] In September, 1949 life tests on Type A units had been in effect some 4,000 hours. With the assumption of an exponential survival

TEST	CONDITIONS	MINIMUM	MAXIMUM
r_{co} – OFF COLLECTOR DC RESISTANCE	$V_C = -35$ V DC $I_\epsilon = 0$ MA DC	17,500 OHMS	—
V_{C1} – ON COLLECTOR VOLTAGE	$I_C = -2$ MA DC $I_\epsilon = 1$ MA DC	—	-3V DC
V_{C3} – ON COLLECTOR VOLTAGE	$I_C = -5.5$ MA DC $I_\epsilon = 3$ MA DC	—	-4V DC
OFF EMITTER RESISTANCE	$V_C = -10$ V DC	50,000 OHMS	—
ON EMITTER RESISTANCE R_{11}	$V_C = -10$ V DC $I_\epsilon = 1$ MA DC	—	800 OHMS
a_1	$V_C = -30$ V DC $I_\epsilon = 1.0$ MA DC	1.5	—
a_2	$V_C = -30$ V DC $I_\epsilon = +0.05$ MA DC	2.0	—
a_3	$V_C = -30$ V DC $I_\epsilon = -0.1$ MA DC	—	0.3
R_{12} – OPEN CIRCUIT FEEDBACK RESISTANCE	$V_C = -10$ V DC $I_\epsilon = +1$ MA DC	—	500 OHMS
R_{21} – OPEN CIRCUIT FORWARD RESISTANCE	$V_C = -10$ V DC $I_\epsilon = +1$ MA DC	15,000 OHMS	—
R_{22} – OPEN CIRCUIT OUTPUT RESISTANCE	$V_C = -10$ V DC $I_\epsilon = +1$ MA DC	10,000 OHMS	—

Fig. 20—Tentative characteristics for the M1689 switching transistor.

law, it was not possible, on the basis of a 4,000-hour test, to estimate the slope sufficiently accurately to warrant a half-life estimate in excess of 10,000 hours. These same Type A units have now run on life test for approximately 20,000 hours. With the more reliable estimate of survival slope now possible, the half-life is now estimated to be somewhat in excess of 70,000 hours. It should be emphasized, however, that these are Type A units of more than two years ago, made with inferior materials and processes. It is believed that those units under current development, being made with new materials and processes, are superior; but, of course, life

RELIABILITY FIGURE OF MERIT	SEPTEMBER 1949	JANUARY 1952
AVERAGE LIFE	$\doteq$ 10,000 HOURS	>70,000 HOURS
EQUIVALENT TEMPERATURE COEFFICIENT OF r_C	-1% PER DEG C	-1/4% PER DEG C
SHOCK	?	> 20,000 G
VIBRATION	?	20 - 5000 CPS NEGLIGIBLE TO 100 G

Fig. 21—Reliability status.

[7] Estimates of life, of course, depend upon definitions of "death." For these experiments, the transistors were operated as Class A amplifiers. A transistor is said to have failed when its Class A gain has fallen 3 db or more below its starting value.

tests are only a few thousand hours old. Although these new data are encouraging, it is still too early to extrapolate the data such a long way.

B. Temperature Effects

Transistors like other semiconductor devices are more sensitive to temperature variations than electron tubes. In terms of the linear equivalent circuit elements, the collector impedance, r_c, and the current gain, a, are the most sensitive. Over the range from $-40°$ to $80°$ C the other elements are relatively much less sensitive. For Type A transistors these temperature variations in r_c and a are shown in Fig. 22. While these curves are definitely not linear, an average temperature coefficient for r_c of about -1 per cent per degree was estimated for the purpose of easy tabulation and comparison in Fig. 21.
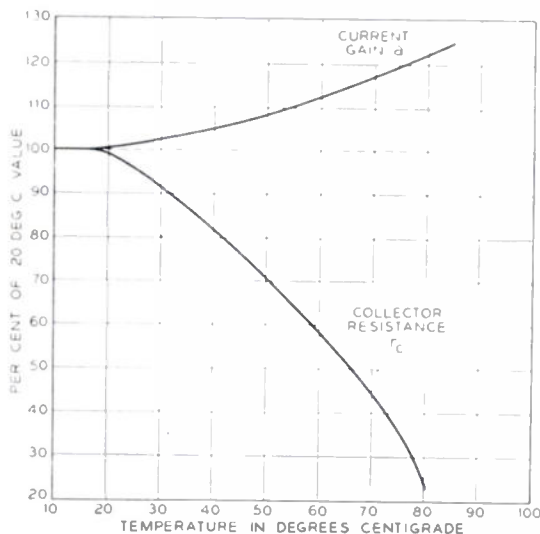
Fig. 22—Collector resistance and a versus temperature for Type A transistor.

Thus, for the early Type A, r_c fell off to about 20 to 30 per cent of its room-temperature value when the temperature was raised to $+80°$ C; at the same time a increased from 20 to 30 per cent over the same temperature range. Today, this variation has been reduced by a
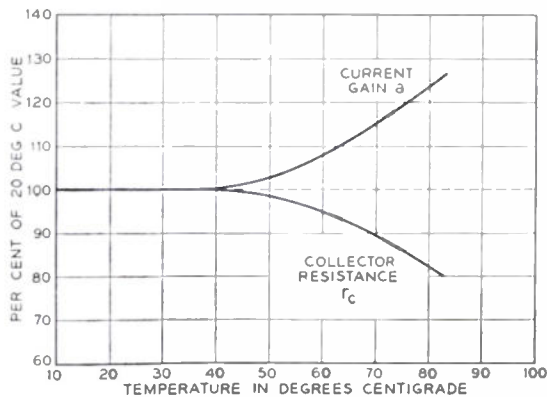
Fig. 23—Collector resistance and a versus temperature for Type M1729 transistor.

factor of about four for r_c in most point-contact types, the variations in the current gain being relatively unchanged. Fig. 23 illustrates the temperature dependence

of r_c and a for the M1729 transistor now under development. Again, for purposes of easy comparison in Fig. 21, the actual dependence of Fig. 23 was approximated by a linear variation and only the slope given in Fig. 21. For linear applications such as the grounded-base amplifier, the Class A power gain is approximately proportional to $a_2 r_c$; hence, the gain of such an amplifier will stay essentially constant within a db or two over the temperature range from $-40°$ to $+80°$ C. For pulse applications, and of importance to dc biasing with point-contact transistors, is the fact that the dc collector current (for fixed emitter current and collector voltage) will change at about the same rate as does r_c, the small signal collector impedance. Similar improvements have been made in these variations for switching transistors and Fig. 24 consists of a series of graphs showing how the
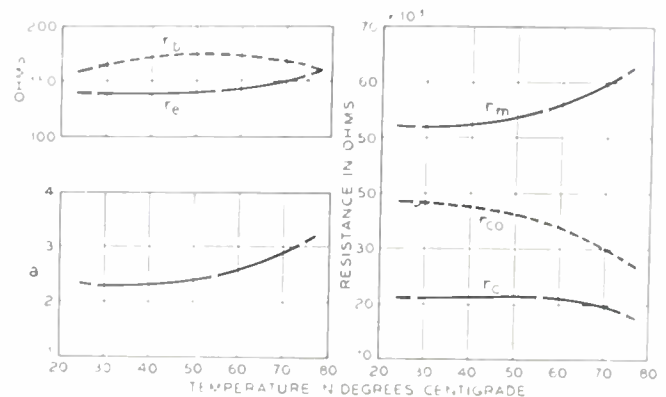
Fig. 24—Temperature behavior of the M1689 transistor.

M1689 bead-type switching transistor changes the pulse characteristics defined in Fig. 20 with respect to temperature. For those switching functions examined to date, it is believed that these data mean reliable operation to as high as $+70°$ C in most applications and perhaps as high as $+80°$ in others.

In junction transistors the laws of temperature variation are not so well established, the device being in a much earlier stage of development. Preliminary data indicate smaller variations in the small signal parameters, such as a and r_c. On the other hand, variations in the dc current, particularly I_{co}, are many times greater, of the order of 10 per cent per degree C.[8] The only saving grace here is the fact that I_{co} is normally very much less than the actual operating value of I_c.

In summary, it may be said that while significant improvements have been made in temperature dependence to the point where many applications appear feasible, it is not to be inferred that the temperature limitation is completely overcome. Much more development work of device, circuit, and system nature is required to bring this aspect of reliable operation to a completely satisfying solution.

C. Shock and Vibration

With regard to mechanical ruggedness, current point-contact transistors have been shock tested up to 20,000 g.

[8] I_{co} is the collector current at zero emitter current.

with no change in their electrical characteristics. Vibration of point-contact and junction transistors over the frequency range from 20 to 5,000 cps at accelerations of 100 g. produces no detectable modulation of any of the transistor electrical characteristics, i.e., such modulation, if it exists, is far below the inherent noise level. At a few spot frequencies in the audio range, vibration tests up to 1,000-g. accelerations similarly failed to produce discernible modulation of the transistor characteristics.

MINIATURIZATION STATUS

A. Space Requirements

In smallness of size, the transistor is entering new fields previously inaccessible to electron devices. The cartridge structure (see Fig. 25), such as the Type A, has a volume of 1/50 cubic inch, compared to about $\frac{1}{8}$

Fig. 25—Miniaturization in space and power drain.

cubic inch for a subminiature tube and about 1 cubic inch for a miniature tube. Under current development, the M1689 bead point-contact transistor has substantially similar electrical characteristics to the M1698[9] cartridge switching unit, but occupies only about 1/2000 cubic inch. The M1752 junction bead transistor has a volume of approximately 1/500 cubic inch, but this may be reduced to the same order as the point-contact bead, if necessary. For further substantial size reductions in equipment, the next move must comprise the passive components. It should be pointed out that the low voltages, low power drain, and correspondingly lower equipment temperatures should make possible further reductions in passive component size.

B. Power Requirements

The transistor, of course, has the inherent advantage of requiring no heater power; moreover, significant advances have been made in the past two years in reducing the collector voltage and power required for practical operation. Consider the minimum collector voltage for which the small-signal Class A gain is still within 3 to 6 db of its full value. In September, 1949 the Type A transistor could give useful gains at collector voltages

[9] The M1698 transistor is a cartridge-type point-contact transistor with electrical characteristics designed for switching and pulse applications. This unit is proving useful in the laboratory development of new circuits or in cases where miniature packages are unnecessary.

as low as 30 volts. Today, several point-contact devices (M1768 and M1734) perform well with collector voltages as low as 2 to 6 volts even for relatively high-frequency operation. One junction transistor, the M1752, can deliver useful gains at collector voltages as low as 0.2 to 1.0 volt. Under these same conditions, the minimum collector power for useful gains may be low as 2–10 mw for point-contact devices and as low as 10 to 100 μw in the case of the junction transistors.[10] Class A efficiencies have been raised for the point-contact devices to as high as 30–35 per cent and for junction transistors this may be as high as 49 per cent out of a maximum possible 50 per cent. Class B and C efficiencies are correspondingly close to their theoretical limiting values.

PERFORMANCE STATUS

Exact electrical performance specifications for the transistor depend, of course, upon the intended applications and the type of transistor being developed for such an application. These types are beginning to be specified; and in fact, they are already so numerous that mention of only a few salient features of some of them will be attempted. Bear in mind, as was pointed out before, that no one transistor combines all the virtues any more than does any one tube type. Fig. 26 attempts to

PERFORMANCE FIGURE OF MERIT	TYPE A SEPTEMBER 1949	JANUARY 1952	NEW DEVELOPMENT TYPE
a – CURRENT GAIN	5 X	50 X	JUNCTION
SINGLE STAGE CLASS A GAIN	18 DB	22 DB	POINT – M1729, M1768
		45 DB	JUNCTION – M1752
NOISE FIGURE AT 1000 CPS	60 DB	45 DB	POINT – M1768
		10 DB	JUNCTION – M1752
FREQUENCY RESPONSE f_c	5 MC	7–10 MC	POINT – M1729
		20–50 MC	POINT – M1734
CLASS A POWER OUTPUT	0.5 WATT	2 WATTS	JUNCTION
SWITCHING CHARACTERISTICS	NONE	GOOD	POINT – M1698, M1689 M1734
FEEDBACK RESISTANCE r_b	250 OHMS	70 OHMS	POINT – M1729
LIGHT/DARK PHOTOCURRENT RATIO	2:1	20:1	JUNCTION – M1740

Fig. 26—Performance progress.

compare the progress made in several important performance merit figures by development of several point-contact and junction types during the last two years. Again the reference performance is that of the Type A as of September, 1949.

Some switching and transmission applications need transistors having high current gain. By going to a point-

[10] In some special cases, depending upon the application, practical operation may be obtained for as little as 0.1 to 1.0 μw.

junction structure, useful values of alpha as high as 50 are now possible with laboratory models.

For straight transmission applications, the single-stage gain of point-contact types (M1768, M1729) has been increased to 20–24 db, whereas for the M1752 junction type the single-stage gain may be as high as 45–50 db.

For high-sensitivity low-noise applications, the point-contact devices have been improved to have noise figures of only about 40–45 db, whereas the M1752 *n-p-n* transistor has been shown to have noise figures in the 10- to 20-db range. All such noise figures are specified at 1,000 cps, and it should be remembered that they vary inversely with frequency at the rate of about 11 db per decade change in frequency.

For video, IF, and high-speed switching applications, measurable improvement has been attained in the frequency response. For video amplifiers up to about 7 mc, the M1729 point-contact transistor is capable of about 18–20 db gain per stage. For high-frequency oscillators and microsecond pulse switching, the M1734 point-contact transistor is under development. Preliminary models of 24 mc IF amplifiers using the M1734 have been constructed in the laboratory, these amplifiers having a gain of some 18–24 db per stage and a bandwidth of several megacycles. However, more work needs to be done on the M1734 to reduce its feedback resistance. For pulse-handling functions, such M1734 units work very nicely as pulse generators and amplifiers of $\frac{1}{2}$-μsec pulses, requiring only 6–8 volts of collector voltage and 12–20 mw of collector power per stage. The amplified pulses can have amplitudes as large as 4–5 volts out of a total collector voltage of 6 volts and rise times as little as 0.01–0.02 μsec.

By increasing the thermal dissipation limits of junction transistors, the Class A power output has been raised to 2 watts in laboratory models. This, however, does not represent an intrinsic upper limit but rather a design objective for a particular application.

Characteristics suitable for switching are now available in the M1698, M1689, and M1734 point-contact types, as previously described; but this is a continually evolving process and more work certainly remains to be done. At present, it is possible to operate telephone relays requiring as much as 50 to 100 ma with M1689 and M1698 point-contact transistors.

New junction-type phototransistors[11] represent a marked advance over the earlier point-contact type.[12] While their quantum efficiencies are not as high as those of the point-contact types, nevertheless the light/dark current ratios are greatly improved and the collector impedance has been raised 10–100 times, thus making possible much greater output voltages for the same light flux.

[11] W. J. Pietenpol, "P-N junction rectifier and photocell," *Phys. Rev.*, vol. 82, no. 1, pp. 121–122; April 1, 1951.
[12] J. N. Shive, "The phototransistor," *Bell Lab. Rec.*, vol. 28, no. 8, pp. 337–342; 1950.

Some Selected Applications

A. Data Transmission Packages

To determine the feasibility of applying transistors in the form of miniature packaged circuit functions, several of the major system functions of a pulse-code-data transmission system have been studied. This investigation has been undertaken under the auspices of a joint-services engineering contract administered by the Signal Corps.

It was desired that these studies should lead to the feasibility development of unitized functional packages combining features of miniaturization, reliability, and lower power drain. Accordingly, it was necessary to carry on in an integrated fashion activities in the fields of system, circuit, and device development to achieve these ends. In particular, circuit and system means have been developed to perform with transistors the functions of encoding, translation, counting, registering, and serial addition. The M1728 junction diode, M1740 junction photocell, and M1689 bead switching transistor are direct outgrowths of this program, and are the devices used in the circuit packages.

At this point, the major system functions listed below have been achieved with interchangeable transistors:

1. 4-digit reversible binary counter.
2. 6-digit angular position encoder.
3. 6-digit gray-binary translator.
4. 5-digit shift register.
5. 2-word serial adder.

DEVELOPMENT PACKAGE TYPE	PACKAGE FUNCTION	DEVELOPMENT TRANSISTOR, DIODE TYPES USED
M 1731-1	REGENERATIVE GATE	M 1689 M 1727
M 1732-1 M 1736 M 1790	BIT REGISTER	M 1689 M 1727 M 1734
M 1733-1 M 1792	PULSE AMPLIFIER	M 1689
M 1735-1 M 1747-1 M 1748-1 M 1751-1 M 1751-2 M 1751-3	DIODE GATE	M 1727 400 A
M 1745-1 M 1791	BINARY COUNTER	M 1689 400 A
M 1749-1	PHOTOCELL READOUT	M 1740
M 1746-1	DELAY AMPLIFIER	M 1689

Fig. 27—Development transistor—circuit packages.

These major system functions are, in turn, built of some seven types of smaller functional packages listed in Fig. 27. The end result of this exploratory development can be said to have demonstrated the feasibility of such a data transmission system in the sense that a workable (though not yet optimal) system can be synthesized from reproducible transistor-circuit packages which

Fig. 28—Bit register package.
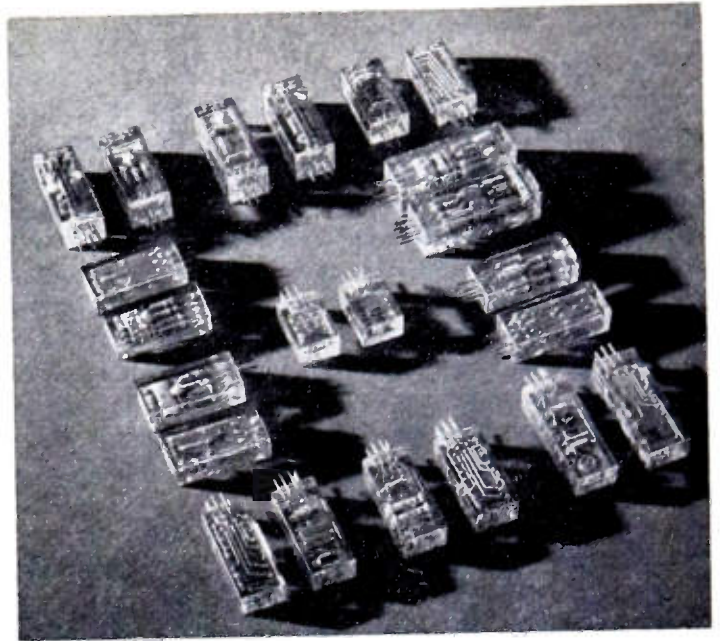
Fig. 29 —Package construction illustrated.

have been produced at reasonable yields and with reasonable (though not yet complete) service reliability. Further development work would be needed in all phases to make such a system of packages suitable for field use. It is estimated that the present laboratory model requires about one-tenth the space and power required to do the same job with present tube art. Fig. 28 is a photograph of a transistor bit-register package and

Fig. 29 is another photograph of such packages showing both sides of the various types employed.[13] Actual final packages would probably not use such clear plastics, and Fig. 30 shows some packages in which the plastic has been loaded with silica to increase its strength and thermal conductivity. The assembly in Fig. 30 consists of a six-digit position encoder at the left, followed by six regenerative pulse amplifiers which in turn feed a six-digit combined translator-shift register.

[13] The Auto-Assembly Process used in the construction of these packages is a Signal Corps Development.
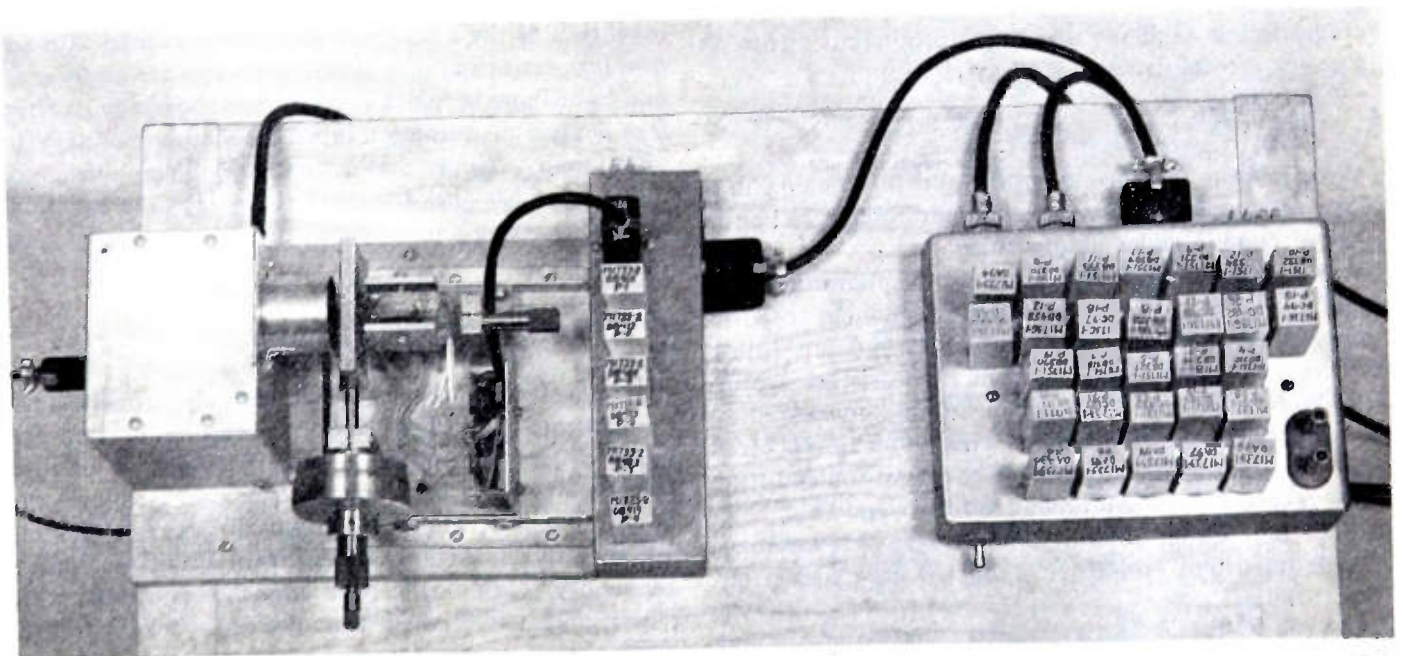
Fig. 3 —Laboratory model of encoder-transistor-register using transistor packages.

B. N-P-N Transistor Audio Amplifier and Oscillator[14]

To the right in Fig. 31 is shown a transformer-coupled audio amplifier employing two M1752 junction transistors. This amplifier has a pass band from 100–20,000 cps and a power gain of approximately 90 db. Its gain is relatively independent of collector voltage from 1–20 volts, only the available undistorted power output increasing as the voltage is increased. At a collector voltage of 1.5 volts it draws a collector current of approximately 0.5 ma per unit for a total power drain of 1.5 mw. Under these conditions it will deliver Class A power output of about 0.7 mw. The noise figure of such an amplifier has been measured to be in the range from 10–15 db at 1,000 cps, depending upon the operating biases.
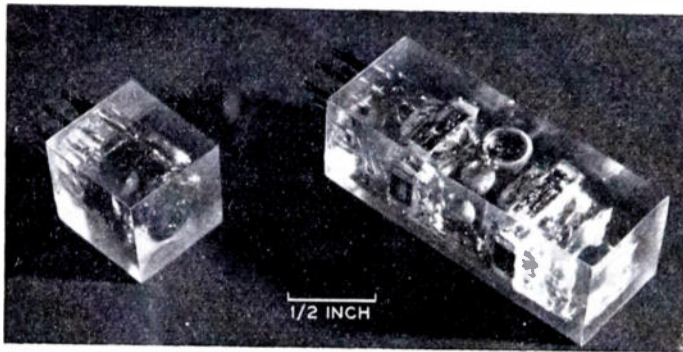
Fig. 31—Packaged oscillator and amplifier using junction transistors.

To the left of Fig. 31 is shown a small transistor audio oscillator having a single M1752 transistor, a transformer, and one condenser. To see just how little power was the minimum necessary to produce stable oscillations such an oscillator was tried at increasingly lower collector supply voltages. It was found that stable oscillations could be maintained down to collector supply voltages as low as 55 mv and collector current as low as 1.5 μa for a total drain of 0.09 μw.

CONCLUSION

With respect to reproducibility and interchangeability, transistors now under development appear to be the equal of commercial vacuum tubes.

With regard to reliability, transistors apparently have longer life and greater mechanical ruggedness to withstand shock and vibration than most vacuum tubes. With regard to temperature effects, transistors are inferior to tubes and present upper limits of operation are 70–80° C for most applications. This restriction is often reduced in importance by the lower power consumption which results in low equipment self-heating. This, however, is the outstanding reliability defect of transistors.

With regard to miniaturization, the comparison fig-

ures are so great as to speak for themselves. Operation with a few milliwatts is always feasible and, in some cases, operation at a few microwatts is also possible.

With regard to performance range, it is believed that the above results imply the following tentative conclusions:

In pulse systems (up to 1–2-mc repetition rates) transistors should be considered seriously in comparison to tubes, since they provide essentially equal functional performance and have marked superiority in miniature space and power. Bear in mind that in some reliability figures they are superior whereas in the matter of temperature dependence they are inferior to tubes.

In cw transmission at low frequencies (<1 mc) essentially the same conclusions are indicated, primarily because of junction transistors. In the range from 1–100 mc, tubes are currently superior in every functional performance figure (except perhaps noise and bandwidth) so that for transistors to be considered for such applications, much greater premium must be placed on miniaturization and reliability than for the first two applications areas.

Thus, it might be assumed that, even though there are many outstanding development problems of a circuit and device nature to be solved, it is appropriate for circuit engineers to explore seriously the application possibilities of transistors—not only in the hope of building better systems but also to influence transistor development towards those most important systems for which their intrinsic potentialities best fit them. It should not be inferred that all important limitations have been eliminated—nor, on the other hand, that the full range of performance possibilities have been explored.

If one remembers the history of engineering research and development in older related fields, it seems apparent that a relatively short time has elapsed since the invention of the first point-contact transistor. Already, new properties and new types of devices are under study and some have been achieved in the laboratory. It therefore is possible, and certainly stimulating, to infer that more than a single new component is involved; that much more lies ahead than in the past; that, indeed we may be entering a new field of technology, i.e., "transistor electronics."

[14] The material of this section represents a summary of some work by Wallace and Pietenpol, described more completely in footnote reference 4.

Properties of Silicon and Germanium*

ESTHER M. CONWELL†

Summary—This article provides the latest experimental information on those fundamental properties of germanium and silicon which are of device interest, currently or potentially. Electrical properties, especially carrier density and mobility, have been treated in greatest detail. Descriptive material has been provided to the extent necessary to give physical background.

INTRODUCTION

IN THIS ARTICLE it is proposed to present the latest available information on the properties of germanium and silicon and, in particular, the best experimental values for important physical quantities. It need scarcely be mentioned that in a field which is developing as rapidly as this some of the information will be obsolete shortly. The general state of development is such, however, that major changes are not anticipated.

It is assumed that the reader is familiar with the elements of the band description of semiconductors to the extent that these are covered in the article by Shockley in this issue[1] or in chapter I of "Electrons and Holes in Semiconductors."[2] For most purposes in this article electrons in the conduction band can be thought of as negatively charged particles moving with random thermal energy through the crystal, differing only from a free electron gas in that the free electron mass is replaced by the appropriate effective mass. Similarly, holes in the valence band can be thought of as free positively charged particles with a different effective mass. In each case the effective mass is the same order of magnitude as the free electron mass.

Electrical properties have been taken up first. Since conductivity depends on the density of current carriers and their mobility, both of which are more fundamental, sections on these precede the section on conductivity. Magnetic, optical, thermal, and mechanical properties have been presented in much less detail than the others because they are currently of less device interest. As is necessary to supplement this type of survey, extensive references have been given. No attempt has been made, however, to provide a complete bibliography.

DENSITY OF CURRENT CARRIERS

A pure, perfect crystal of germanium or silicon at the absolute zero of temperature would have no electrons in the conduction band and no holes in the valence band. The current carriers present at higher temperatures originate in the valence band, chemical impurities in the lattice, and lattice defects of some types. Of primary importance in determining the number of carriers that each source supplies under a given set of external conditions is its ionization energy, i.e., the energy required to liberate a bound electron or hole. This will now be considered for each of the three in the order stated.

To raise an electron from the valence band to the conduction band, creating a hole-electron pair, requires energy equal to the energy gap between these bands. Since the electron energy levels depend on the spacing between atoms, the energy gap would be expected to depend somewhat on temperature and pressure. The energy gap has been found to depend linearly on temperature according to

$$\epsilon_G(T) = \epsilon_G(0) - \beta T, \tag{1}$$

where $\epsilon_G(T)$ represents the value of the energy gap at the absolute temperature T and β is a constant. The values of these quantities found empirically are listed in Table I.[3]

TABLE I

SOURCE OF CARRIERS	IONIZATION ENERGY IN ELECTRON VOLTS	
	GERMANIUM	SILICON
Valence band	0.75-.0001T[4]	1.12-.0003T[5]
B, Al, Ga, In (Low concentration)	0.01-[4]	0.08[5]
P, As, Sb (Low concentration)	0 01+[6]	0.05[5]
Zn (Low concentration)	0.03[7]	
Thermal acceptors (Cu)	0.03[8]	

[4] J. Bardeen and W. Shockley, *op. cit.* Note, however, that optical experiments given $\epsilon_G(0) = 0.72$ ev. See footnotes for discussion of the precision of the temperature coefficients for germanium and silicon.
[5] G. L. Pearson and J. Bardeen. *loc. cit.*
[6] P. P. Debye. private communication. Bell Telephone Laboratories, Inc., the slope of $\log_e nT^{-3/2}$ versus $1/T$, and are therefore only approximately equal to the ionization energy. For a discussion of this point, see W. Shockley, *op. cit.*, p. 471. The quoted values are, however, Boltzmann's constant times.
[7] W. C. Dunlap, Jr., "Zinc as an acceptor in germanium," *Phys. Rev.*, vol. 85. p. 1945; 195a. Also. P. P. Deby, private communication, Bell Telephone Laboratories Inc., Murray Hill, N. J. Again this value is approximate for the reasons discussed in footnote reference 14.
[8] W. DeSorbo and W. C. Dunlap, Jr., "Resistivity of heat-treated germanium between 11°K and 298°K," *Phys. Rev.*, vol. 83, pp. 869, 879; 1951.

In both silicon and germanium the energy gap increases when pressure is applied. The temperature dependence of the energy gap can be used to compute its pressure dependence if the assumption is made that the change of energy gap with temperature is due entirely to thermal expansion.[9] The changes in energy gap per unit dilation computed from the values of β given in Table I are 3×10^1 and 5 ev per unit dilation for silicon

[1] W. Shockley, "Transistor electronics: imperfections, unipolar and analog transistors," PROC. I.R.E., vol. 40, pp. 1289–1314; this issue.
[2] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y.; 1950.

[3] The stated values of β, however, are in considerable doubt because they involve the assumption that the effective mass of an electron or hole is equal to the free electron mass. This can be only approximately true. See formula (2) and associated text and footnote.
[9] See, however, footnote reference 20(a) in J. Bardeen and W. Shockley, "Deformation potentials and mobilities in nonpolar crystals," *Phys. Rev.*, vol. 80, pp. 72–80; 1950.

and germanium, respectively. Approximately the latter value has been obtained for germanium by a number of experiments of other types as well.[10]

Chemical impurities in the lattice may act as donors or acceptors. Those which have been studied extensively are elements in column III of the periodic chart, which act as acceptors, e.g., boron, aluminum, gallium, and indium, and elements in column V, which act as donors, e.g., arsenic, phosphorus, and antimony. Since the electron or hole is bound loosely in a large hydrogen-like orbit, it is expected that the ionization energy should be about the same for all these elements, and this is in fact found. This ionization energy decreases with increasing impurity concentration, vanishing for concentrations of about 10^{18} or 10^{19}/cm^3. The values stated in Table I are for small impurity concentration, up to about 10^{15}/cm^3 perhaps.[11] Zinc, which has also been studied, acts as an acceptor with considerably greater ionization energy than the elements of columns III and V. This value is also listed in Table I.

It has been found that acceptors can be introduced in germanium by means of heat treatment at temperatures above about 600°C.[12] Equilibrium concentration of acceptors introduced for a given temperature of heat treatment increases exponentially with the temperature over a considerable range. With this process it is possible to convert n-type germanium into p-type provided the initial n-type resistivity is not too low. Prolonged heating at 500°C is found to reverse this conversion. The process is also characterized by diffusion of the acceptors from the surface into the interior at a rate which is large for diffusion in solids.[13] The ionization energy for these thermal acceptors has been measured and is listed in Table I. There has been considerable speculation about the nature of these acceptors. The effects described might, in principle, arise from the introduction by the heat treatment of either lattice defects[14] or chemical impurities. It has recently been established, however, that copper is one important agent, if not the sole agent, responsible for these phenomena. In one set of experiments it was found that germanium which was heat treated for several hours in a hydrogen atmosphere and in a system free from contaminants was not converted, whereas it was if copper were added.[15]

Furthermore, experiments with radioactive copper indicate that the diffusion rate and equilibrium concentration of copper in germanium are, within experimental error, the same as those measured for the thermal acceptors.[16] It has also been shown that copper acts as an acceptor in the germanium lattice.

The energy required to ionize a donor or acceptor can be supplied in a number of ways, the most usual being heat and light. Strong electric fields can also create hole-electron pairs. This is called the "Zener effect" and will be discussed later under that heading. The effect of light will be considered in the section on optical properties. The remainder of this section will be devoted to considering the variation of carrier density with temperature.

Theory indicates that for a specimen in thermal equilibrium the product of the electron concentration, denoted by n, and the hole concentration, denoted by p, is a function of temperature only, given by

$$np = 2.33 \times 10^{31} \left(\frac{m_n m_p}{m^2} \right)^{3/2} T^3 e^{-\epsilon_G/kT}, \qquad (2)$$

where m_n and m_p denote the effective masses of electrons and holes, respectively, m is the free electron mass, and k is Boltzmann's constant. The values found empirically are listed in Table II. When the number of carriers donated by impurities or lattice defects is small compared to the number from the valence band of the semiconductor, the conduction is called "intrinsic." In that case,

$$n \simeq p = n, \qquad \text{intrinsic conduction.} \qquad (3)$$

This enables the use of (2) or, speaking practically, of the empirical formulas of Table II[17] to compute the density of electrons or holes in intrinsic material at any temperature. The results are

$$n_i = 9.7 \times 10^{15} \, T^{3/2} e^{-4350/T} \text{ Ge,}$$
$$n_i = 2.8 \times 10^{16} \, T^{3/2} e^{-6450/T} \text{ Si.} \qquad (4)$$

The values of n_i for 300°K are listed in Table II. In Fig. 1 n_i is plotted as a function of temperature. In the temperature range where the conduction is not intrinsic, (2) still holds for a specimen in thermal equilibrium, but the densities of electrons and holes, respectively, depend on the concentrations of donors and acceptors and the ionization energy in a relatively complicated fashion. There are two practical procedures for determining n and p for a given extrinsic sample at a given temperature. One way involves measuring the resistivity. This is related to np and the mobilities of electrons and holes according to formula (8). (See section on conductivity.) The mobilities can be computed as

[10] For a discussion of, see J. Bardeen and W. Shockley, *op. cit.*

[11] For a study of the variation with impurity content in silicon, see G. L. Pearson and J. Bardeen, "Electrical properties of pure silicon and silicon alloys containing boron and phosphorus," *Phys. Rev.*, vol. 75, pp. 865–883; 1949.

[12] J. H. Scaff and H. C. Theuerer, "Effect of heat treatment on the electrical properties of germanium," *Jour. Met.*, vol. 191, pp. 59–63; 1951.

[13] C. S. Fuller, H. C. Theuerer, and W. van Roosbroeck, "Properties of thermally produced acceptors in germanium," *Phys. Rev.*, vol. 85, pp. 678–679; 1952.

[14] Evidence that disorder in the lattice produces acceptors and donors is provided by experiments in which germanium and silicon are bombarded by high-speed atomic particles. See the chapter entitled "Nucleon Bombarded Semiconductors," by K. Lark-Horovitz in "Semiconducting Materials," Butterworth's Scientific Publications Ltd., London; 1951.

[15] W. P. Slichter and E. D. Kolb, "Impurity effects in the thermal conversion of germanium," *Phys. Rev.*, vol. 87, pp. 527–528; August 1, 1952.

[16] C. S. Fuller and J. D. Struthers, "Copper as an acceptor element in germanium," *Phys. Rev.*, vol. 87, pp. 526–527; August 1, 1952.

[17] Since the effective masses are not known (with one exception) but are probably not very different from the free electron mass, it is customary to set the factor in (2) involving the masses equal to 1. The empirical value of np is then compared with (2) to obtain the values of β and E_G (0) listed in Table I.

TABLE II

QUANTITY	VALUE	
	GERMANIUM	SILICON
Electron density × hole density	$9.3 \times 10^{31} T^3 e^{-8700/T}$ per cm⁶ [18]	$7.8 \times 10^{32} T^3 e^{-12,900/T}$ per cm⁶ [18]
Density of carriers in intrinsic material at 300°K	2.5×10^{13}/cm³	6.8×10^{10}/cm³
Intrinsic resistivity at 300°K	47 ohm cm	63,600 ohm cm

[18] W. van Roosbroeck, private communication, Bell Telephone Laboratories, Inc., Murray Hill, N. J.

described in the next section. With numerical values for ρ, e, μ_n, and μ_p inserted in (8) it provides a relationship between n and p which can be used in conjunction with the empirical formula for np in Table II to determine these quantities. A second and more direct procedure to determine them is by measuring the Hall effect.[19]
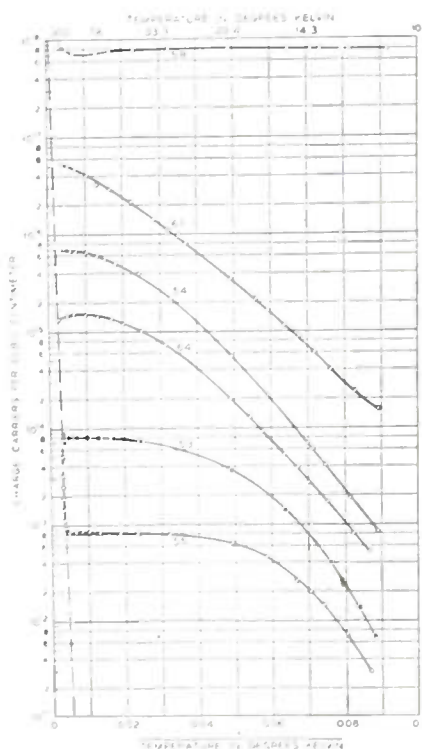
Fig. 1—Density of charge carriers versus the reciprocal of absolute temperature for a set of arsenic-doped germanium samples. The data were taken by Debye at Bell Laboratories. The numbers identify particular samples. The dashed line represents the density of intrinsic carriers.

Since the variation of carrier concentration with temperature for an extrinsic sample is not given here explicitly, it is useful to illustrate this for some particular cases. For this purpose there are shown in Fig. 1 experimental data for a set of n-type germanium samples with increasing amounts of arsenic (donor) impurity. These have been chosen to represent all parts of the resistivity range. Sample 55 is very pure, with a room-temperature resistivity of about 40 ohms cm, while sample 58 at the other end of the resistivity spectrum is 0.005 ohm cm. (Mobility and conductivity are plotted for these same

samples in Figs. 4 and 6, respectively.) Sample 58 has essentially the same number of carriers at all temperatures, or zero ionization energy. For all others, carrier concentration is relatively small at low temperatures, but increases rapidly with temperature because of the small ionization energy of the donors. For the pure and moderately impure ones, essentially all the electrons are freed from the donors somewhat below room temperature. The density then remains constant with increasing temperature[20] until temperatures are reached so that a substantial number of electrons in the valence band receive sufficient energy for the valence bonds to be broken, i.e., until the sample starts to become intrinsic. At temperatures higher than this the number of carriers is just the intrinsic number and is given by the dashed line.

For germanium doped with a column III impurity, i.e., p-type, the variation of carrier concentration with temperature would be very much the same. In the case of silicon the curves would be of this general type, but because of the higher ionization energy not all the electrons or holes would be free at room temperature. Also, because of the larger energy gap silicon would not become intrinsic until higher temperatures.

MOBILITY

Low Field Case

In the absence of an electric field, electrons and holes move in the crystal with random thermal energy. Their speeds will be distributed according to a Maxwell-Boltzmann distribution, with the average speed therefore proportional to $\sqrt{T}$. Application of an electric field accelerates electrons and holes. The low field case is defined as that in which the applied field is small enough so that the energy gain from the field is small compared to thermal energy and the Maxwell-Boltzmann distribution is not significantly disturbed by the field.

Acceleration of current carriers by the field results in the development of a net drift velocity in the direction of the force. The drift velocity, v_d, developed per unit of electric field intensity, E, is called the "mobility," i.e.,

$$v_d = \mu E. \qquad (5)$$

Since each time an electron or hole undergoes a collision its acceleration by the field is interrupted more or less

[19] See W. Shockley, *op. cit.*, p. 204.

[20] The small dip seen shortly before the intrinsic line in some of these samples, for example 64, does not indicate a decrease in the number of carriers. What is plotted in Fig. 1 is essentially the reciprocal of the Hall coefficient, which is related to the number of carriers by a proportionality factor which decreases slightly in this range.

effectively, the frequency and nature of the collisions determine the mobility. The collisions which are most important in semiconductors are those with thermal vibrations of the lattice and with ionized donors or acceptors. It is useful to speak of the lattice mobility, μ_L, and the impurity mobility, μ_I, defined as the respective mobilities a charge would develop if one collision or scattering process alone were operative.

For lattice scattering it can be shown that at a given temperature the average distance between collisions is the same for all particles regardless of speed.[21] Faster particles will therefore make more collisions per second, with the result that the mobility will decrease if the average speed of the particles increases. Increasing the temperature decreases the mobility for two reasons: First, the mean speed increases and, second, the average distance between collisions decreases because the lattice vibrations grow more vigorous. The theoretically predicted and experimentally verified result is essentially a $T^{-3/2}$ dependence of the lattice mobility.[22] The mobility measured at room temperature for samples with resistivity greater than a few ohm cm is the lattice mobility since impurity scattering is negligible in this case. Experimental values are listed in Table III. Combining these data and the temperature dependence mentioned above results in the empirical formulas for lattice mobility as a function of temperature which are shown in Table IV.

TABLE III
MOBILITIES IN CM²/VOLT SEC AT 300°K

CARRIER	GERMANIUM	SILICON
Electron	$3,600 \pm 180$[23]	$1,200 \pm 120$[24]
Hole	$1,700 \pm 90$[23]	250 ± 50[24]

[23] J. R. Haynes and W. Shockley, "Mobility and life of injected holes and electrons in germanium," *Phys. Rev.*, vol. 81, pp. 835–843; 1951.
[24] J. R. Haynes and W. C. Westphal, "Drift mobility of electrons in silicon," *Phys. Rev.*, vol. 85, p. 680; 1952.

TABLE IV
LATTICE MOBILITIES IN CM²/VOLT SEC

CARRIER	GERMANIUM	SILICON
Electron	$19 \times 10^6 T^{-3/2}$	$6.2 \times 10^6 T^{-3/2}$
Hole	$8.9 \times 10^6 T^{-3/2}$	$1.3 \times 10^6 T^{-3/2}$

Scattering by ionized donors or acceptors is coulomb scattering, which is less effective as the average speed of the carriers increases and thus as the temperature increases. The impurity mobility calculated theoretically for the case of scattering by ionized impurities is[25]

$$\mu_I = \frac{8\sqrt{2}\kappa^2 (kT)^{3/2}}{\pi^{3/2} N_I e^3 m_{eff}^{1/2} \ln\left[1 + \left(\dfrac{3\kappa kT}{e^2 N_I^{1/3}}\right)^2\right]}, \quad (6)$$

[21] See, for example, W. Shockley, *op. cit.*
[22] Experimental data of Fig. 5 show small deviation from this.
[25] E. M. Conwell and V. F. Weisskopf, "Theory of impurity scattering in semiconductors," *Phys. Rev.*, vol. 77, pp. 388–390; 1950.

where κ is the dielectric constant, N_I is the total density of ionized impurities, e is the charge on the electron, and m_{eff} the effective mass of the electron or hole. A thorough experimental test of this formula has not as yet been possible. There is evidence over a considerable range of temperatures and impurity densities that it is of the right order of magnitude. It has recently been found that (6) gives correctly the variation of mobility with resistivity observed for electrons in germanium at room temperature, provided the effective mass of an electron is taken as one-quarter the free electron mass.[26] Since there is other evidence for this effective mass for electrons in germanium, this seems to provide a satisfactory verification of the formula at room temperature for this case. For holes the situation may be much more complicated.[27] Approximations made in the derivation of (6) are such that it cannot be expected to be valid at very low temperatures or very high impurity densities.

With the values of the dielectric constants taken as 16 for germanium and 12 for silicon,[28] (6) can be written

$$\mu_I = 8.5 \times 10^{17} \left(\frac{m}{m_{eff}}\right)^{1/2} \frac{T^{3/2}}{N_I \ln\left(1 + 8.3 \times 10^8 \dfrac{T^2}{N_I^{2/3}}\right)} \quad \text{Ge}$$

$$\mu_I = 4.7 \times 10^{17} \left(\frac{m}{m_{eff}}\right)^{1/2} \frac{T^{3/2}}{N_I \ln\left(1 + 4.5 \times 10^8 \dfrac{T^2}{N_I^{2/3}}\right)} \quad \text{Si.}$$

$$(7)$$

These mobilities are plotted as a function of temperature for several different impurity concentrations in Fig. 2. For this plot, $(m/m_{eff})^{1/2}$ has been set equal to 1. This plot can be used to obtain μ_I, or at least an approximate value thereof, at any temperature provided N_I is known. If the density of carriers is known, it is usually a good approximation to take N_I equal to this density.[29] For electrons in germanium a more accurate result will be obtained by multiplying μ_I shown in the graphs by 2, since for these $m_{eff} \cong m/4$. For convenience, there are also plotted in Fig. 2 the empirical lattice mobilities of Table IV.

To obtain the mobility, μ, from μ_L and μ_I the approximation of taking the reciprocal of the sum of the separate reciprocals is generally used. This frequently provides a very poor approximation, however. An exact calculation can be carried out very simply with the use

[26] P. P. Debye and E. M. Conwell, "Mobility of electrons in germanium," to be published in *Phys. Rev.*
[27] It seems likely that the edge of the valence band is degenerate with consequent complications.
[28] H. B. Briggs, "Optical effects in bulk silicon and germanium," *Phys. Rev.*, vol. 77, p. 287; 1950.
[29] To see that this is not exactly true, consider an *n*-type sample. Such a sample, apart from a density N_D of donors, will always have some density of acceptors, N_A, where, of course, $N_A < N_D$. These acceptors will trap electrons which would otherwise go into the conduction band. They will therefore be ionized and scatter electrons at all temperatures. If N_D is the density of ionized donors, $N_I = N_D^i + N_A$, while $n = N_D^i - N_A$.
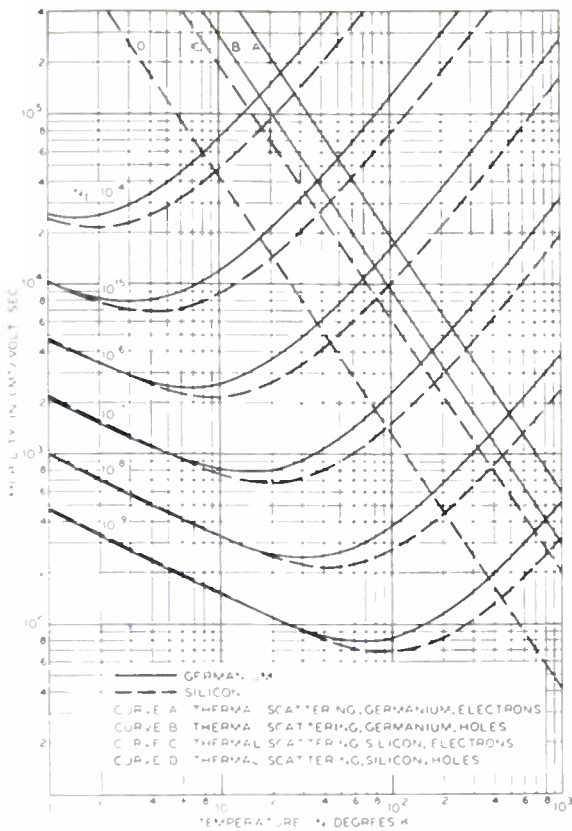
Fig. 2—Impurity and lattice (thermal) scattering versus temperature. The impurity scattering is taken from the theoretical formulas (7) with m/m_{eff} set equal to 1. To use for electrons in germanium, therefore, multiply μ_I by 2. The thermal scattering values are semiempirical, taken from Table IV.

of Fig. 3, which is a graph of the correct relationship between μ/μ_L and μ_L/μ_I. Given the value of μ_L/μ_I, the value of μ/μ_L can be obtained from the graph, and from this the desired value of μ.

Calculation of the mobility for any sample can be carried out as follows: For a relatively pure sample (resistivity greater than a few ohm cm) the mobility is essentially μ_L, and given by Table IV. For a less pure sample, knowledge of the resistivity enables at least an approximate calculation. Consider an n-type sample for definiteness. Provided that the approximation $N_I = n$ can be used, there is a unique relationship between resistivity and mobility at a given temperature, which can be calculated theoretically. If a value of N_I is assumed, μ_I and then μ can be calculated as previously described. For an n-type sample $1/\rho = ne\mu_n$ (see formula (8)), where e is the charge on the electron. Since $N_I = n$, ρ can then be computed. A set of such calculations for different values of n gives for n-type germanium at room temperature the curve shown in Fig. 4. In these calculations
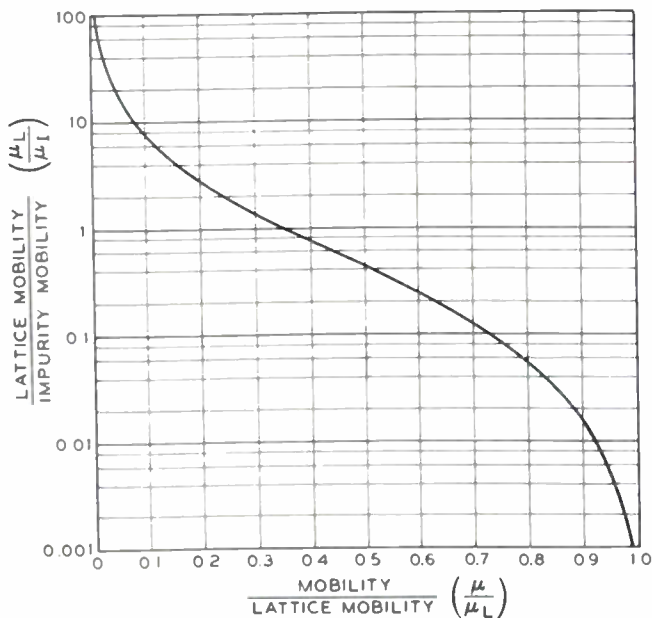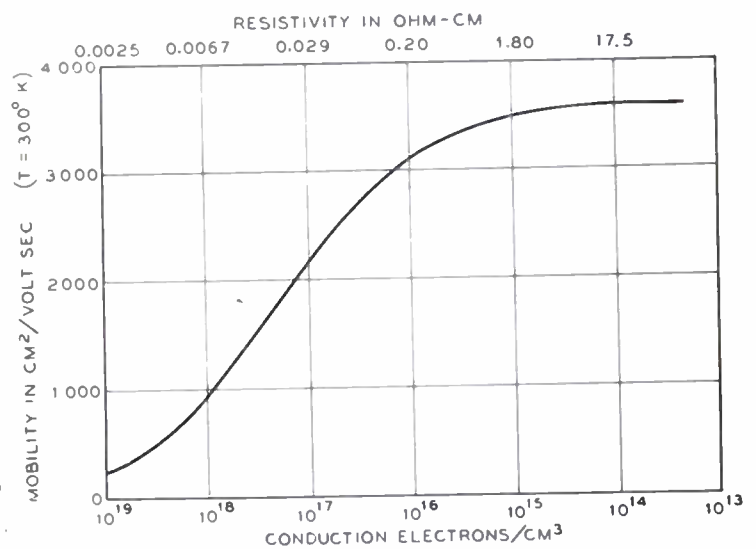
Fig. 4—Mobility versus carrier density and resistivity for n-type germanium at 300°K.

m_{eff} was taken equal to $m/4$. The lower scale shows N_I. This variation of mobility with resistivity has been experimentally verified. Below about 1 ohm cm impurity scattering cuts μ down significantly. This graph will enable immediate determination of the room-temperature mobility and N_I from the room-temperature resistivity. This value of N_I can then be used to compute μ at other temperatures. Similar graphs can be made for p-type germanium and silicon, with less probable accuracy, however, since the effective mass is not known.

We shall consider now, in the light of the previously discussed theory, some typical experimental data on mobility. In Fig. 5 the mobility is plotted on a log-log scale as a function of temperature for the same samples whose concentration data appears in Fig. 1.[30] A line representing a $-3/2$ slope, characteristic of the lattice

Fig. 3—Plot of the theoretical value of μ/μ_L versus μ_L/μ_I.

[30] What is plotted is actually Hall mobility, which is the mobility multiplied by a factor which varies from 1.18 for very pure samples to 1.93 for very impure ones.

mobility, is drawn for comparison. For the purest sample, 55, lattice scattering is dominant to the lowest temperatures at which measurements were made although impurity scattering affects the mobility below
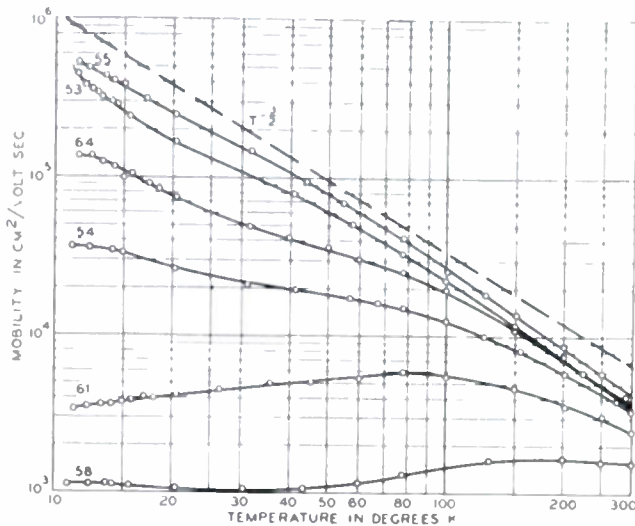
Fig. 5—Mobility versus temperature for the same set of arsenic-doped germanium samples. The data were taken by Debye at Bell Laboratories. The dashed line represents a −3/2 slope characteristic of lattice scattering.

about 70°K. As the impurity content of the samples increases, the mobility is decreased over the entire range and the departure from the $T^{-3/2}$ variation takes place at higher and higher temperatures. In the most impure samples the impurity content is sufficiently high to cause the mobility to start decreasing with temperature, which is characteristic of impurity scattering alone.

High Field Case

In the previously discussed low field case the carriers are able, in collisions with the lattice vibrations, to dissipate the power fed in by the field without their average speed being significantly changed from the thermal value. For this range of fields the mobility is independent of the field, as was discussed previously. When the electric field gets high enough, however, the average electron speed is increased, and the more so the higher the field. An increase of average speed in a sample in which lattice scattering is dominant (such as a fairly pure sample at room temperature) will, as discussed in the preceding section, lead to a decrease in mobility. In such samples, then, it is expected that beyond the range of constant mobility the mobility will decrease with increasing fields. This is found experimentally, as shown in Fig. 6. In this figure there are plotted data for two n-type germanium samples, the 298° curve taken for a 3 ohm cm (at low fields) sample, the other two curves for a sample which was 9 ohm cm at room temperature. The impurity concentration was small enough in both these cases so that lattice scattering was more important

than impurity scattering at 77°K as well as 298°K. Beyond the region of constant mobility the mobility decreases as $1/\sqrt{E}$, and at even higher fields, as $1/E$.[31] The latter dependence can be observed at 77°K at higher fields than those obtained in this particular experiment. In general, the departure from constant mobility occurs at lower fields as the temperature decreases. This is expected theoretically because collisions grow less frequent as the temperature goes down, and the rate at which the field supplies energy to the current carriers is higher.

At 20°K in the second sample shown in Fig. 6 impurity scattering was dominant. In this case increase of the average speed of a particle greatly decreases its deflection as it goes by an impurity ion, increasing its mobility. Beyond the region of constant mobility, therefore, the mobility at first rises with field. As a result of this increase in speed the impurity ions become relatively ineffective as scatterers, and collisions with the lattice vibrations become operative in determining the mobility. It is expected, then, that after this rise the $1/E$ variation characteristic of the lattice vibrations takes over, and this is also seen in Fig. 6.
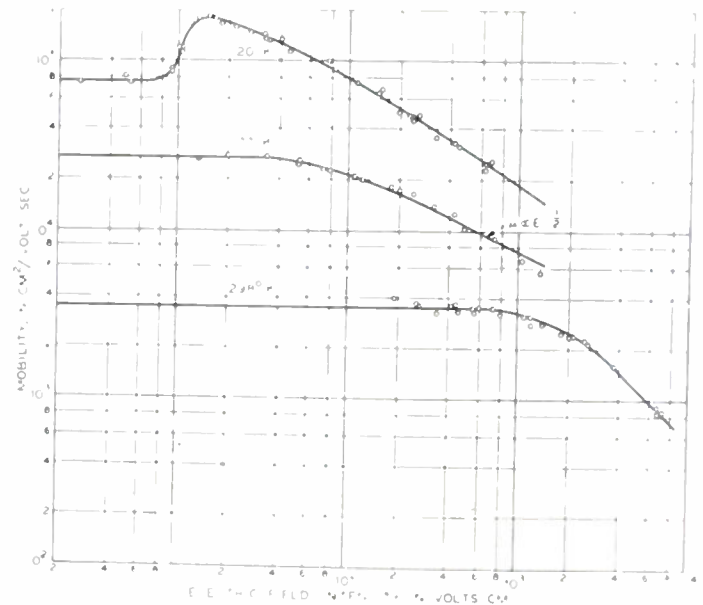
Fig. 6—Mobility versus electric field intensity. The data were taken by Ryder at Bell Laboratories.

These data are typical for fairly pure samples of n-type germanium. For less pure samples, the type of variation observed at 20°K would occur at a higher temperature. Approximate values for the field intensity at which departure from constant mobility takes place are listed in Table V. The value for n-type silicon is not available.

[31] For theoretical explanation, see W. Shockley, "Hot electrons in germanium and Ohm's law," *Bell Sys. Tech. Jour.*, vol. XXX, pp. 990–1034; October, 1951.

TABLE V

APPROXIMATE FIELD INTENSITY (VOLTS/CM) AT WHICH DEPARTURE FROM CONSTANT MOBILITY TAKES PLACE $(T = 300°K)$

Conductivity Type	Ge	Si
n	900[32]	
p	1,400[33]	7,500[33]

[32] E. J. Ryder and W. Shockley. "Mobilities of electrons in high electric fields," *Phys. Rev.*, vol. 81. pp. 139–140; 1951.
[33] E. J. Ryder, to be published in the *Phys. Rev.*

CONDUCTIVITY

The electrical conductivity, σ, and the resistivity, ρ, are related to the concentrations of electrons and holes and their respective mobilities by

$$\sigma = 1/\rho = ne\mu_n + pe\mu_p, \qquad (8)$$

where e is the charge on the electron. The conductivity will therefore be affected by any condition which affects carrier concentrations or mobilities. We shall consider explicitly only its variation with temperature.

For intrinsic material the carrier density and the mobilities can be taken from the empirical formulas given in (4) and Table IV, respectively. This gives

$$\sigma_i = 1/\rho_i = 4.3 \times 10^4\, e^{-4,350/T}\ \text{ohm}^{-1}\ \text{cm}^{-1}\ \text{Ge},$$
$$\sigma_i = 1/\rho_i = 3.4 \times 10^4\, e^{-6,450/T}\ \text{ohm}^{-1}\ \text{cm}^{-1}\ \text{Si}. \qquad (9)$$

The resistivities of intrinsic material at 300°K computed from these formulas are listed in Table II. Intrinsic conductivity is plotted versus the reciprocal of temperature in Fig. 7 as a dashed line.

Some typical experimental data for the case of extrinsic conductivity are also plotted in Fig. 7. The data shown are for the same set of arsenic-doped germanium samples whose concentration and mobility data are plotted in Figs. 1 and 5, respectively. For sample 58 σ does not vary much with temperature since n and μ do not. It is useful in discussing the other samples to consider various temperature ranges separately. At low temperature, variation of carrier density is the dominant effect. As the temperature rises and donors become ionized, n increases and therefore σ increases. It is interesting to note that in this temperature range samples 53, 64, 54, and 61, which have quite different impurity concentrations, do not differ greatly in conductivity. This can be understood as follows: For these samples impurity scattering is dominant in this range. Their mobility therefore varies approximately inversely with the density of ionized impurities. (See (5). The logarithmic variation with impurity concentration can be neglected.) Since the carrier density is approximately equal to this density, the product of n and μ, and therefore the conductivity, should be approximately the same for all of them.

Over the range of temperatures where the impurities are essentially all ionized and carrier concentration is constant, mobility variation dominates. In this range

lattice scattering is usually dominant, and the conductivity therefore decreases as the temperature increases. At still higher temperatures intrinsic conductivity, described by (9), sets in.
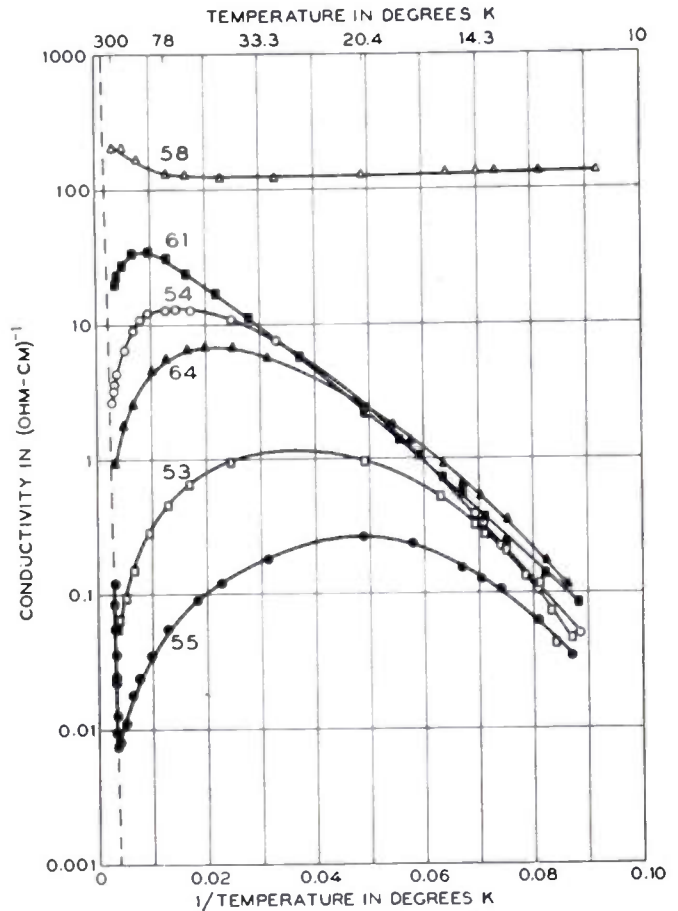
Fig. 7—Conductivity versus absolute temperature for the same set of arsenic-doped germanium samples. The data were taken by Debye at Bell Laboratories. The dashed line represents intrinsic conductivity.

ZENER CURRENT

Application of a strong electric field may enable electrons from the valence band to cross the energy gap which separates them from the conduction band. This involves a process of penetration of a potential barrier very similar to the penetration of the potential barrier at the surface of a metal when a high field is applied to produce cold emission. The resulting current, frequently called "Zener current," is very strongly field dependent. Fields of the strength required to produce a measurable Zener current without undue heating and other undesirable effects may be obtained in the space-charge region of a p-n junction biased in reverse.

A typical reverse current-voltage characteristic for a germanium p-n junction is shown in Fig. 8. For low fields the only carriers available which can cross the junction are the holes in the n-region and the electrons in the p-region. Since these are relatively few in number, this constitutes a small current and one which is practically independent of field since the carriers travel

mainly by diffusion. This accounts for the constant saturation current observed up to about 6 volts in this particular junction. At about 6 volts, however, an additional current starts to flow which increases very sharply with voltage. The field at which the transition takes place and the general behavior of the curve beyond the transition are approximately what is predicted theoretically from a simple picture of the Zener effect. The additional current has therefore been identified as Zener current.[34]
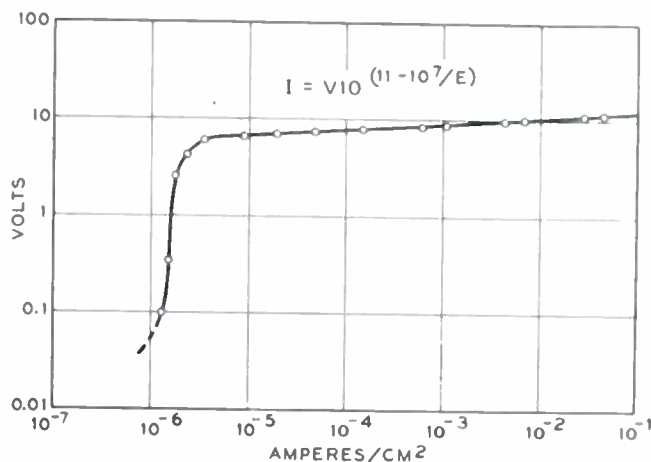
Fig. 8—Current-voltage characteristic for a *p-n* junction biased in reverse. The data were taken from McAfee, Ryder, Shockley, and Sparks.[34] The formula included is derived theoretically from a simple picture of Zener effect, and fits the Zener current for this particular junction quite well.

The field intensity at which the so-called Zener current becomes appreciable compared to the saturation current is called the "critical field." In germanium this has been found to depend on crystallographic orientation. The value obtained when the ||| direction of the crystal is perpendicular to the *p-n* interface is 2×10^5 volts/cm. Values found for other orientations in germanium differ by as much as a factor of 2.[35] In silicon there is no evidence for a dependence of the critical field on crystallographic direction. Values have been reported ranging from 2×10^5 volt/cm[36] to 2.5×10^5 volts/cm.[37]

It has also been found that the current-voltage characteristic changes somewhat with temperature. For the

junctions studied the voltage required to produce a particular current in the Zener range increases as the temperature increases at a rate of about 0.1 per cent per degree. This does not necessarily indicate that the critical field varies in this way.

LIFETIME

Departure from the thermal equilibrium concentrations of electrons and holes can be obtained by injecting carriers of the minority type, e.g., holes in *n*-type material, say, at a point contact.[38] To maintain electrical neutrality of the specimen an equal number of the majority carrier must be drawn in at the same time. (This process is of importance in many types of transistor action.) The carrier concentrations can also be changed from the thermal equilibrium values by, for example, the creation of hole-electron pairs by light, X-rays, or strong electric fields. When there is an excess of carriers present, the system tends to return to equilibrium by recombination of the excess electrons and holes.

Such recombination has been observed by Haynes and Shockley.[39] They find that a small pulse of holes injected into a filament decays exponentially, so that if N_0 is the original number of injected holes, the number existing after time t is $N_0 e^{-t/\tau_p}$. τ_p, called the lifetime of a hole, is thus the time required for the pulse to decay to $1/e$ of its original size.[40] Analogous results have been obtained for electrons injected into *p*-type material.

Experiments with filaments of varying cross section and surface treatment show that the recombination in general occurs both in the volume and on the surface of the material. The lifetime observed, τ, can be considered to be made up of separate lifetimes for body and surface recombination, τ_i and τ_s, respectively, according to

$$1/\tau = 1/\tau_i + 1/\tau_s. \qquad (10)$$

The factors affecting these separate lifetimes will now be considered.

Body lifetime varies very widely from crystal to crystal. It is also a function of temperature for any given crystal. The values to be quoted are for room temperature and small injected carrier density. In both *n*- and *p*-type germanium, values as high as 1,000 μsec have been observed.[41] Typical values for crystals grown

[34] K. B. McAfee, E. J. Ryder, W. Shockley, and M. Sparks, "Observations of Zener current in germanium *p-n* junctions," *Phys. Rev.*, vol. 83, pp. 650–651; 1951. In a number of samples measured since this publication, however, it has been found that the slope of the curve in the Zener region is quite different from the theoretical prediction. The effect observed may thus be more complicated.

[35] K. B. McAfee, private communication, Bell Telephone Laboratories, Inc., Murray Hill, N. J.

[36] K. B. McAfee and G. L. Pearson, abstract E6 in the program of the 1952 Washington meeting of A.P.S. These results were obtained for grown junctions. Also, "The electrical properties of silicon *p-n* junctions grown from the melt," *Phys. Rev.*, vol. 87, p. 190; July, 1952.

[37] G. L. Pearson and B. Sawyer, "Silicon *p-n* junction alloy diodes," PROC. I.R.E., vol. 40, pp. 1348–1352; this issue. These results were obtained for alloy process junctions.

[38] For references and a discussion of the literature see W. Shockley, *op. cit.*, chap. 3

[39] J. R. Haynes and W. Shockley, "Investigation of hole injection in transistor action," *Phys. Rev.*, vol. 75, p. 691; 1949; and vol. 81, p. 835; 1951.

[40] What is actually observed in these experiments and most lifetime-measuring experiments is a quantity called the "diffusion length" rather than the lifetime. Written for holes, the diffusion length, denoted by L_p, is defined by

$$L_p = \sqrt{D_p \tau_p},$$

and is the average distance along any given direction which a hole will diffuse before recombining.

[41] J. R. Haynes, private communication, Bell Telephone Laboratories, Inc., Murray Hill, N. J. However, much higher values have been rumored.

by the pulling method of Teal and Little[42] from originally purified germanium are a few hundred μsec for high-resistivity material (of the order of 30 ohm cm) and 100 μsec for material of about 5 ohm cm. Few data are available for silicon. The values reported are in general lower than those for germanium. As a general trend, the lower the resistivity of the samples the lower the lifetime, but there are frequent exceptions to this.

The great variations from crystal to crystal are due to the fact that body lifetime is quite sensitive to the presence of chemical impurities and imperfections in the lattice. In line with this it has been found that thermal conversion and bombardment both have drastic effects on the lifetime, being able to cut it down from several hundred μsec, for example, to 1 μsec or less. Crystals strained during solidification may have lifetimes as low as a μsec. Doping with column III and V elements in general decreases the lifetime somewhat.

The lifetime for surface recombination, τ_s, depends on the dimensions of the sample as well as on surface properties because the carriers must diffuse to the surface before they can recombine there. It is convenient to separate out the dependence on the surface itself by introducing the surface velocity of recombination, s, defined by

$$s = \frac{\text{number recombining per second per unit surface area}}{\text{excess density over thermal equilibrium value just below surface}}.$$

Values of s have been found ranging from less than 10^2 cm/sec to greater than 10^4 cm/sec, these being correlated with surface treatment. The high value stated is obtained for sandblasted surfaces. If the surface is polished smooth and etched with CP-4 etch to remove worked material, the value of s is decreased to 5×10^2 cm/sec. Chemical treatment of this polished and etched surface can reduce this to about 1×10^2 cm/sec, or increase it to values as high as those obtained for sandblasted surfaces.[43] Values of s less than 10^2 cm/sec have also been reported for etched surfaces.[44]

The relationship among τ_s, s, and the dimensions of the sample has been derived theoretically by Shockley for a rod or filament of rectangular cross section. The results can be expressed simply for the limiting cases of very large s and very small s. For a sample with transverse dimensions $2B \times 2C$, these are

$$\frac{1}{\tau_s} = \frac{\pi^2 D}{4}\left(\frac{1}{B^2} + \frac{1}{C^2}\right) \qquad s \to \infty$$

$$\frac{1}{\tau_s} = s\left(\frac{1}{B} + \frac{1}{C}\right) \qquad s \to 0, \tag{11}$$

where D is the appropriate diffusion constant.[45]

DIFFUSION OF CURRENT CARRIERS

An electric current can also arise from diffusion of electrons or holes. For specificity we shall speak of holes in n-type material, the modification for electrons in p-type material being obvious. Also, we shall assume there is no externally applied electric field. Consider a situation in which the density of holes varies with position, such as would be the case if holes were being injected at a point. For simplicity we will assume that this variation takes place in one direction only, and has a concentration gradient dp/dx. As a result of thermal agitation there will be a net flow of holes tending to smooth out the inequalities in concentration. In the case in which the injected hole density is much smaller than the electron concentration the holes will disturb the electron distribution very little and will diffuse like a neutral gas. The hole current density, j_p, arising from diffusion is proportional to the concentration gradient of the holes, and is directed oppositely. It can be written for this case of small injected density,

$$i_p = -eD_p \frac{dp}{dx}. \tag{12}$$

This defines D_p, the diffusion coefficient of holes. A diffusion coefficient for electrons, D_n, can be defined similarly.

For small injected hole density the diffusion coefficient is related to the mobility by the Einstein relation,[46]

$$\mu_p = \frac{e}{kT} D_p. \tag{13}$$

A corresponding relation of course exists for electrons. At 300°K for fairly pure germanium samples, the diffusion coefficients calculated from this are

$$D_n = 93 \text{ cm}^2/\text{sec}$$

Ge at 300°K.

$$D_p = 44 \text{ cm}^2/\text{sec}$$

Further numerical values will not be given since the Einstein relation plus knowledge of the mobility makes their calculation simple.

[42] G. K. Teal and J. B. Little, "Growth of germanium single crystals," *Phys. Rev.*, vol. 78, p. 647; 1950.

[43] See J. R. Haynes and W. Shockley, *op. cit.*, for composition of CP-4 etch and details of the chemical treatment.

[44] D. Navon, R. Bray, and H. Y. Fan, abstract E10 in the program of the 1952 Denver meeting of A.P.S. Also, "Lifetime of injected carriers in germanium," *Bull. Amer. Phys. Soc.*, vol. 27, pp. 14-15; June 30, 1952. Also, W. Brattain, private communication, Bell Telephone Laboratories, Inc., Murray Hill, N. J.

[45] For filaments of square cross section, the relationship between τ_s and s for all values of s is given graphically in W. Shockley, *op. cit.*, p. 324.

[46] A. Einstein, "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," *Ann. der Phys.*, vol. 17, pp. 549-560; 1905.

When a large density of holes is injected, the change in the electron distribution due to the holes cannot be neglected. To neutralize the space charge of the holes a correspondingly large density of electrons must accumulate in the neighborhood of the holes and must furthermore move at the same rate. The hole current (or electron current) is, in this case, also proportional to the concentration gradient, but the constant of proportionality is given by eD,[47] where

$$D = \frac{2D_n D_p}{D_n + D_p}, \qquad (14)$$

this being the same for electrons and holes. This result applies also to the diffusion of carriers injected into intrinsic material. At room temperature in germanium $D = 59$ cm²/sec.

MAGNETORESISTANCE

Experimental data on the change of resistance in a magnetic field have been obtained by Pearson and Suhl for n- and p-type germanium.[48] They investigated the dependence of the effect on magnetic field strength, temperature, and angle between current and magnetic field. In the case of p-type germanium the fractional change in resistance, $\Delta\rho/\rho$, was considerably larger for the field perpendicular to the current than for the field parallel to the current, while for n-type germanium the reverse was true. The size of the effect was found to increase as the square of the field at low fields and as a lower power of H at high fields. For any given field the effect was found to increase as the temperature was lowered. In the temperature range from 77°K to room temperature, $\Delta\rho/\rho$ was never larger than about 10 per cent for fields up to 1,000 gauss. Very considerable effects were observed, however, for higher fields. At room temperature in n-type material for a field of 100,000 gauss, $\Delta\rho/\rho$ was 1.1 for the field perpendicular to the current, 2.2 for the field parallel to the current. At 77°K $\Delta\rho/\rho$ for the latter case reached 2.0 at a field of 20,000 gauss.

OPTICAL PROPERTIES

We shall consider first the optical absorption of pure material. In Fig. 9, per cent transmission is plotted versus wavelength, λ, in microns (1 micron = 10^{-4} cm) for high resistivity specimens of germanium and silicon. Both absorb strongly in the visible and near infrared, and are relatively transparent over most of the range of longer infrared wavelengths. These qualitative features can be understood readily.

One process by which light can be absorbed in germanium or silicon is through excitation of electrons from the valence band to the conduction band, creating hole-electron pairs. The energy of a quantum required for this transition corresponds to light in the visible and near infrared regions of the spectrum. The lowest energy quantum which would enable an electron to make this transition would have to have energy equal to the energy gap. At the long wavelength limits of the absorption observed in germanium and silicon, the quantum energies agree quite well with the previously given values of the respective energy gaps. It has also been shown experimentally that this process has unit quantum efficiency, i.e., for every quantum absorbed in this range one hole-electron pair is created.[49]
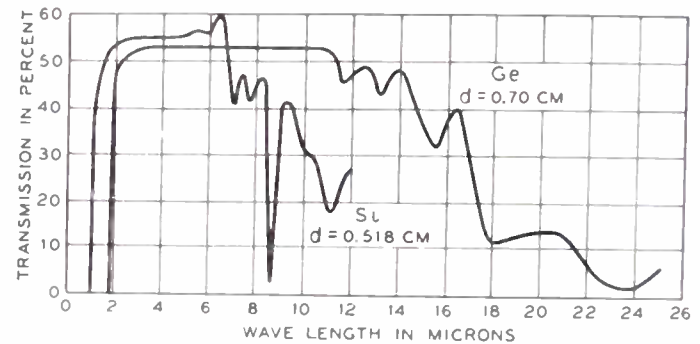
Fig. 9—Per cent transmission versus wavelength for high-resistivity germanium and silicon. The data were taken by Briggs at Bell Laboratories. Thickness of the sample is indicated in each case.

For wavelengths longer than the absorption limit, the energy cannot be absorbed by electrons in the valence band, and the crystals will be relatively transparent. Transmission is only about 50 per cent, however, because of reflection at the surface. The sharp peak at about 9 microns in silicon is probably due to absorption by the lattice.[50] The other absorption observed in the infrared may be due to the free carriers.

Impure samples of germanium and silicon show the same behavior in the visible and near infrared as the pure samples. Beyond the absorption limit, however, they show greater absorption than pure samples. This absorption increases with carrier concentration, and is presumably due to the absorption of the free carriers.

ACKNOWLEDGMENTS

[47] W. van Roosbroeck, "Theory of the flow of electrons and holes in germanium and other semiconductors," *Bell Sys. Tech. Jour.*, vol. 29, pp. 560–607; 1950.

[48] G. L. Pearson and H. Suhl, "Magneto-resistance effect in oriented single crystals of germanium," *Phys. Rev.*, vol. 83, pp. 768–776; 1951.

[49] F. S. Goucher, "Photon yield of electron-hole pairs in germanium," *Phys. Rev.*, vol. 78, p. 816; 1950.

[50] H. B. Briggs, "Infra-red absorption in silicon," *Phys. Rev.*, vol. 77, pp. 727–728; 1950.

THERMAL PROPERTIES

THERMOELECTRIC POWER DEPENDS STRONGLY ON CARRIER DENSITY SO NO VALUES WILL BE GIVEN

TABLE VI

QUANTITY	GERMANIUM		SILICON	
	VALUE	TEMPERATURE	VALUE	TEMPERATURE
Linear thermal expansion coefficient	$6.1 \times 10^{-6}/°C$[51] $6.6 \times 10^{-6}/°C$[51]	0–300°C 300–650°C	$4.2 \times 10^{-6}/°C$[57]	10–50°C
Thermal conductivity	0.14 cal/sec cm°C[53] ∼0.11 cal/sec cm°C[63]	25°C 100°C	0.20 cal/sec cm°C[54]	20°C
Specific heat	0.074 cal/gm°C[52]	0–100°C	0.181[52]	18.2–99.1°C
Latent heat of fusion	8300 cal/mol[55]		9450 cal/mol[51]	
Melting point	936°C[56]		1420°C[52]	
Boiling point	2700°C[52]		2600°C[52]	

[51] W. L. Bond, private communication. Bell Telephone Laboratories, Inc., Murray Hill, N. J.
[52] "Handbook of Physics and Chemistry," Chem. Rubber Publishing Co., ed. 33.
[53] A. Greico and H. C. Montgomery. "Thermal conductivity of germanium." *Phys. Rev.*, vol. 86. p. 570; 1952.
[54] "Metals Handbook." American Society for Metals; 1948.
[55] See paper by L. Brewer in "Chem. and Met. of Misc. Materials," edited by L. L. Quill, McGraw-Hill Book Co., Inc.; 1950.
[56] E. S. Greiner, private communication. Bell Telephone Laboratories, Inc., Murray Hill, N. J.
[57] M. E. Straumanis and E. J. Aka. "Lattice parameters, coefficients of thermal expansion, and atomic weights of purest silicon and germanium," *Jour. Appl. Phys.* vol. 23, pp. 330–334; 1952.

MECHANICAL PROPERTIES

TABLE VII

QUANTITY	VALUE	
	GERMANIUM	SILICON
Elastic constant C_{11}	12.98×10^{11} dynes/cm² [58]	16.740×10^{11} dynes/cm² [57]
Elastic constant C_{12}	4.88×10^{11} dynes/cm² [58]	6.523×10^{11} dynes/cm² [59]
Elastic constant C_{11}	6.73×10^{11} dynes/cm² [58]	7.957×10^{11} dynes/cm² [59]
Volume compressibility	1.3×10^{-12} cm²/dyne[60]	0.98×10^{-12} cm²/dyne[61]

[58] W. L. Bond, W. P. Mason, H. J. McSkimin, K. M. Olsen, and G. K. Teal. "Elastic constants of germanium single crystals." *Phys. Rev.*, vol. 78, p. 176; 1950.
[59] H. J. McSkimin, W. L. Bond, E. Buehler, and G. K. Teal. "Measurement of the elastic constants of silicon single crystals and their thermal coefficients," *Phys. Rev.*, vol. 83, p. 1080; 1951.
[60] Calculated from the elastic constants.
[61] P. W. Bridgeman. "Linear compressions to 30,000 kg/cm² including relatively incompressible substances," *Proc. Amer. Acad. Arts and Sci.*, vol. 77, no. 6, pp. 187–234; 1949.

MISCELLANEOUS

TABLE VIII

QUANTITY	VALUE	
	GERMANIUM	SILICON
Atomic number	32	14
Atomic weight	72.60[57]	28.08[57]
Lattice constant	5.657×10^{-8} cm[57]	5.431×10^{-8} cm[57]
Density	5.323 gm/cm³[57]	2.328 gm/cm³[57]
Dielectric constant	16[26]	12[26]
Magnetic susceptibility	-0.12×10^{-6} cgs[52]	-0.13×10^{-6} cgs[52]
Debye temperature	290°K	

Preparation of Germanium Single Crystals*

LOUISE ROTH† AND W. E. TAYLOR‡

Summary—Single crystals of germanium have been prepared by two methods using a vacuum furnace ($\sim 5 \times 10^{-6}$ mm Hg pressure) with induction heating.

In the first method nucleation is induced at the upper surface of the melt by producing the necessary heat gradient, and a single crystal is grown in the crucible by controlling the rate of cooling. In the second method molten germanium is seeded with a single crystal and is withdrawn as it solidifies in the orientation of the seed crystal. Seeds orientated with the (111), (100), and (110) planes parallel to the surface of the melt have been used and well-developed crystal faces are obtained. Single crystals of pure germanium and Sb or In-doped germanium does not wet the crucible; Al-doped germanium adheres to the crucible.

Resistivity gradients may be greatly reduced by repeated recrystallizations. Pure germanium of about 55 ohm-cm resistivity can be obtained in this way.

INTRODUCTION

SINCE THE ELECTRICAL properties of semiconductors depend critically on crystal imperfections (impurities or lattice flaws), the preparation of pure single crystals is of importance. It is of historical interest that the first attempt to measure electrical resistivity of vacuum melts of silicon by Heraeus[1] and single crystals[2] by de Haen and Schuchardt led to the conclusion that silicon was a metal and not a semiconductor.

Seemann with Gudden[3] came to this conclusion because of the high conductivity of the single crystals—resistivity values between 0.3×10^{-3} and 1.5×10^{-3} ohm-cm (by two powers of 10 below those obtained by Koenigsberger[4] for polycrystalline silicon).[5] A plot[6] of the logarithm of the conductivity as a function of $1/T$

(Fig. 1), summarizing some of the old measurements on silicon, leads one to conclude that silicon is a semiconductor and that the energy gap (note in Fig. 1 the nearly identical (intrinsic) high-temperature slopes) is 1.1–1.2 electron volts.
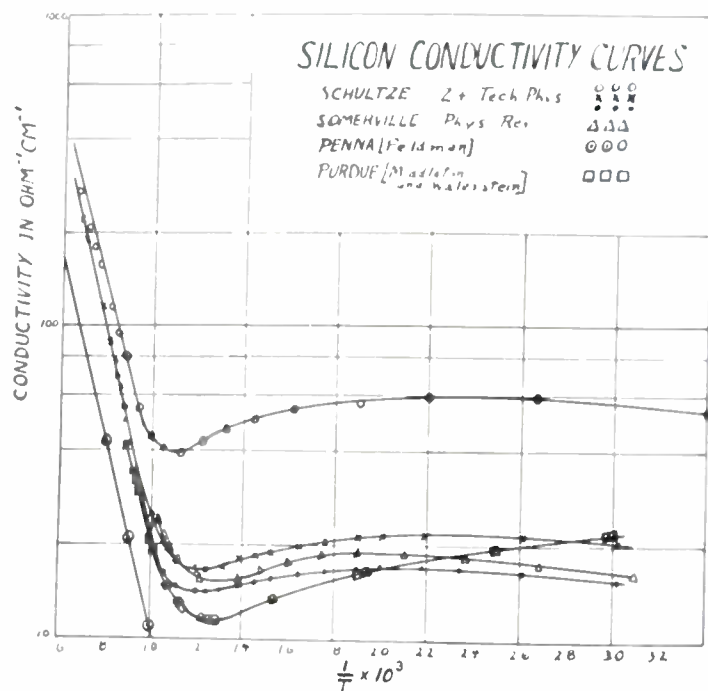
Fig. 1—The logarithm of the conductivity in mho-cm^{-1} is plotted as a function of $1/T$ for a number of silicon samples measured. High-temperature slopes indicate an energy gap of about 1.1–1.2 electron-volt. A. Schulze, "Untersuchung on Silizium," *Z. tech. Phys.*, vol. 11, p. 443; 1930. J. C. Somerville, "Temperature coefficients of electrical resistance III," *Phys. Rev.*, vol. 33, p. 77; 1911. A. E. Middleton and I. Walerstein, Purdue University, unpublished data; 1942–1943.

The reduction of SiCl$_4$ with zinc vapor, used during the war by Dupont[7] leads to single-crystal silicon needles orientated, as X-rays show[8] parallel to the [111] axis. Resistivity measurements by Walerstein[9] on these needles gave values from a few ohm-cm to 10,000 ohm-cm.[10] The only Hall value, obtained at room temperature on a needle having a conductivity of 0.0083 mho-cm^{-1}, was about -2.3×10^3.

[1] F. Lauster, "Zur Frage der Elektrische Leitfähigkeit des geschmolzenen Siliciums," *Z. f. Phys.*, vol. 59, p. 83; 1929; vol. 61, p. 578; 1930.

[2] H. J. Seemann, "Zur elektrische Leitfähigkeit des Siliziums," *Phys. Zeit.*, vol. 28, p. 765; 1927.

[3] B. Gudden, "Electrical Conductivity of Electronic Semiconductors," *Erg. exakt Naturw.*, vol. 13, p. 235; 1934.

[4] J. Koenigsberger and K. Schilling, "Über die elektrische Leitfähigkeit einiger fester Substanzen," *Phys. Zeit.*, vol. 9, p. 347; 1908.

[5] H. J. Seemann, "Zur elektrischen Leitfähigkeit des Siliziums Nachtrag," *Phys. Zeit.*, vol. 29, p. 94; 1928.

[6] K. Lark-Horovitz, Preparation of Semi-Conductor and Development of Crystal rectifiers; Final report NDRC 19-585, U. S. Dept of Commerce; 1945. "The New Electronics," A.A.A.S. Symposium, Section 2; 1949. We are indebted to Mr. Lark-Horovitz for use of this manuscript. This was later substantiated by careful measurements at Bell Telephone Laboratories. Correct value for the gap is 1–1.2.

G. L. Pearson and J. Bardeen, "Electrical properties of pure silicon and silicon alloys containing boron and phosphorus," *Phys. Rev.*, vol. 75, p. 865; 1949.

[7] A. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, N. Y., p. 67; 1948.

[8] K. Lark-Horovitz, Contractor's Final Report, NDRC-14-585, p. 24; 1945. (J. D. Orndoff; 1943.)

[9] Purdue University, personal communication; 1943.

[10] H. v. Wartenberg, "Über Silicium," *Z. f. anorg. Chem.*, vol. 265 p. 186; 1951.

Such experiments[11] clarify the necessity for producing single crystals of high purity in order to determine the true properties of silicon.

During the latter part of 1947 and early 1948 when R. E. Davis[12] was engaged in improving purification and metallographic techniques, in making germanium ingots he found higher mobilities in samples containing large crystallites than was observed in older material. It was concluded that grain boundaries may play a large part in the scattering of carriers and, therefore, a systematic program of single-crystal growing of germanium was started by W. E. Taylor at the suggestion of K. Lark-Horovitz.

The starting material for making high-purity germanium crystals is the highest purity germanium oxide obtainable from the Eagle-Picher Company. The oxide is reduced in a stream of hydrogen at 650° C. The resulting germanium powder is ready for vacuum melting without further processing.[13]

The reduction furnace used at the present time consists of a 2-inch quartz tube heated by a pair of hinged semicylindrical resistance units, which may be opened at the end of the heating period to hasten cooling. The brass fittings waxed to each end of the tube contain the hydrogen vents, and copper tubing for water-cooling the waxed joints. The sample is inserted at the forward end and the furnace is made gas tight by means of a rubber gasket between the end plate and the cylindrical part of the fitting, which are screwed together. The furnace is tilted slightly from front to back to keep the condensed moisture of the reaction from the charge. The excess hydrogen and water vapor pass from the furnace through a trap and a water bubbler and are allowed to escape to the outside.

The melting furnace consists of a 2-inch clear quartz tube sealed at one end and evacuated (Fig. 2). Heat is supplied by a movable high-frequency induction coil. Either graphite or glazed porcelain are satisfactory crucibles since pure germanium and most germanium alloys do not wet them. It has been found, however, that the interfacial surface tension in highly doped aluminum-germanium alloys is such that the melt adheres to the porcelain glaze.

The premelt is carried out in a graphite crucible. The germanium powder is slowly outgassed to prevent a rise of pressure above 10^{-5} mm Hg. When the material is molten, the power is turned off and a polycrystalline ingot is formed.

This ingot is etched for 10 to 15 minutes to remove debris, and is now ready to be made into a single crystal.

Single Crystals

There are several methods known for the production of large crystals,[14,15] but only two have been used by us. One of these makes use of a thermal gradient to promote the growth of large crystals from a spontaneously formed nucleus; the other combines a thermal gradient with artificial seeding of the melt with a previously obtained single crystal (Czochralski method). Both of these methods have been tried successfully at Purdue.

Thermal Gradient Method[16]

Germanium expands on solidification, thus causing crystals to float on top of the melt. By cooling the melt from the top it is possible to produce large crystals.
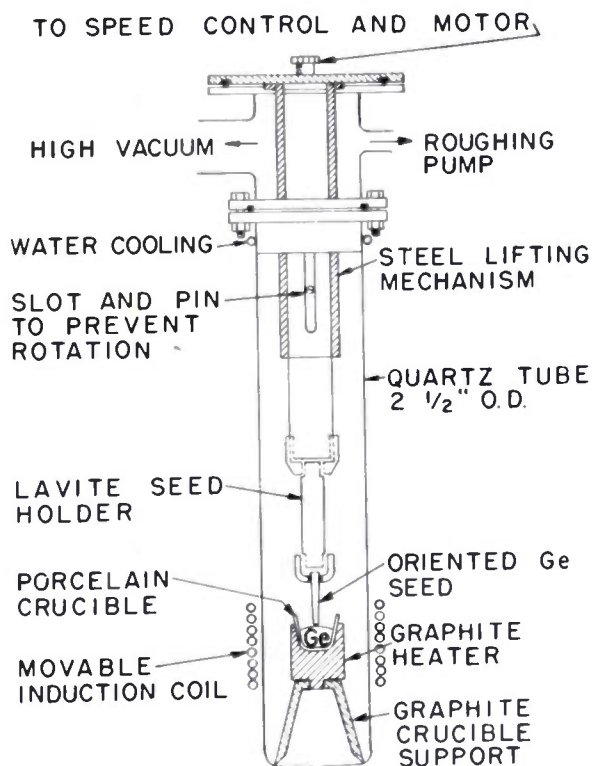
Fig. 2—High-vacuum melting furnace showing crystal pulling mechanism.

The premelt ingot, etched and rinsed, is placed in a porcelain crucible which is fitted into a graphite heater as shown in Fig. 2, and outgassing is carried out as in the case of the premelt ingot. After the charge is heated to about 1200° C for a period of 4 to 5 minutes, the power is lowered so that the temperature is decreased to about 965° C. The induction coil is then lowered at a rate of about 0.03 inch per minute until the first crystal forms at the top of the melt, and is maintained at this

[11] G. K. Teal and E. Buehler, "Growth of silicon single crystals and of single crystal silicon *p-n* junctions," *Bull. Am. Phys. Soc.*, vol. 27, no. 3, p. 14; 1952.

[12] Now at Westinghouse Research Laboratories.

[13] It is extremely important to maintain the highest standards of cleanliness at every step in the preparation of germanium since small amounts of impurities introduced into the material may produce changes in its electrical properties.

[14] F. Halla, "Kristallchemie and Kristallphysik Metallischer Werkestoffe," J. A. Barth, Leipzig; 1939.

[15] H. E. Buckley, "Crystal Growth," John Wiley and Sons, New York, N. Y.; 1951.

[16] W. E. Taylor, Ph.D. Thesis, Purdue University; 1950. Signal Corps Special Report W-36-039-SC-38151; 1950.

position until the top is solid. Then lowering is continued at the rate of 0.03 inch per minute over a distance equal to the height of the ingot. The pressure is maintained below 2×10^{-5} mm. Hg.

In principle this method of growing large crystals is quite simple. In actual practice a number of difficulties arise if one is really intent upon making a single crystal of the entire charge: 1. It is comparatively rare for only one crystal to appear and grow over the entire face of the melt. It may be necessary to melt back and start the coil-lowering process several times if a number of crystals appear simultaneously or in quick succession. 2. Even if one succeeds in producing a single large crystal, the bottom of the ingot (and often the edges) may contain small parasite crystals started at the surface of the crucible. In spite of these difficulties ingots consisting of from 1 to 3 large crystals and weighing up to 100 grams have been made by this method.

Such ingots show a resistivity gradient from top to bottom (the top showing the highest resistivity), as well as a radial gradient, and may contain both *N*- and *P*-type regions. The electrical properties of various regions of ingots made by this method have been discussed in detail by Taylor.

SEEDING THE MELT[17]

When the last portion of a germanium melt solidifies in the crucible, a single-crystal spike is frequently extruded through a break in the upper surface. Rotation X-ray photographs of such spikes show that the direction of growth is along the [111] axis of the crystal. Needles are also found in sublimed germanium, and again the X-ray diagram shows the same orientation and structure as for the silicon needles.[18] In view of the growth pattern of these single crystals, it was decided to prepare the first seed crystal to be introduced into the melt with the [111] axis perpendicular to the molten surface. Subsequently, [110] and [100] oriented seeds have been used with equal success.

A seed is prepared from a single crystal which has been alternately X-rayed and cut until the desired crystal face is exposed.[19] It is also possible to orient crystals by examining the etch patterns.[20] Fig. 3 shows the pattern observed on the (100) or cube face of a germanium crystal etched with H_2O_2.

The seed crystal is held in the lavite holder which is screwed into the base of the lifting mechanism[21] (Fig. 2). The driving motor (not shown) and Metron variable ratio speed changer are mounted on the top plate.

[17] G. K. Teal and J. B. Little, "Growth of germanium single crystals," *Phys. Rev.*, vol. 78, p. 647; 1950.
[18] However, a microphotograph of such a needle shows stacked octrahedra with the [111] axis along the needle (J. Radavich).
[19] We are indebted to I. G. Geib, who performed X-ray analysis.
[20] Such methods have recently been widely used at RCA and Westinghouse. Personal Communication.
[21] Designed by W. E. Taylor.

The crucible used for holding the charge is porcelain (Fig. 2). After the charge has been outgassed slowly, it is brought above the melting point as before and held there for about 5 minutes. Then the power is reduced to give a temperature of about 965° C. The induction coil is lowered 0.03 inch per minute until a crystal starts forming, then it is raised until it just melts. The seed is then lowered until it is in contact with the melt, and is allowed to remain stationary for a time sufficient to melt it at the interface. If the seed does not melt slightly, a poor bond is established between it and the molten germanium, resulting ingot will be polycrystalline.

Lifting is begun slowly, if necessary intermittently, to prevent pulling the seed from the melt which is drawn up by surface tension at the point of contact.
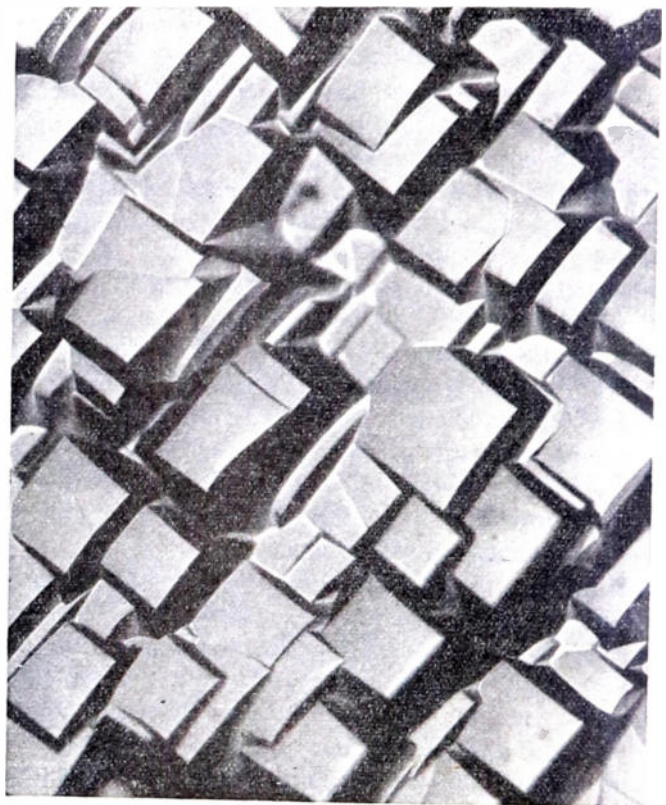
Fig. 3—Etch pattern brought out on (100) crystal face of germanium by 30-per cent hydrogen peroxide etch.

As lifting proceeds and heat is carried away from the molten material through the seed, the melt solidifies in the orientation of the seed crystal. The diameter of the ingot will usually increase as withdrawing proceeds. If it does not, it may be necessary to lower the heating coil slightly. When the desired diameter is reached, a constant speed of lifting will maintain it. The speed most frequently used in this laboratory is 3 inches an hour.

Pure germanium ingots of 175 grams have been made by this method. There is a vertical resistivity gradient, but horizontal layers are quite uniform and there are no barriers. Vertical gradients may be re-

Fig. 4—Pure germanium single-crystal drawn with growth axis in [110] direction. The faces shown are (111) crystal planes.

duced by cutting off and discarding the bottom of the ingot, which contains dissolved impurities, and by re-crystallizing the top. Material of 55 ohm-cm resistivity has been made in this way. Fig. 4 shows a crystal grown on the [110] axis. Laue diffraction patterns taken at intervals along the well-developed (111) planes show the orientation to be unchanged throughout the length.

Doped ingots have been prepared with antimony, indium, and aluminum. In all cases, single crystals have been obtained. These ingots too show a vertical concentration gradient, but horizontal uniformity.

Conclusion

Pure germanium powder, obtained by hydrogen reduction of the oxide, has been processed in vacuum to obtain single crystals of germanium and germanium alloys with predetermined orientation.

P-N Junctions by Impurity Introduction Through an Intermediate Metal Layer*

L. D. ARMSTRONG†

Summary—This paper describes an experimental method for making *p-n* junctions by alloying and diffusing indium into *n*-type germanium through an intermediate thin layer of some other metal, such as gold, which has been plated on the germanium. The junction characteristics are similar to those of junctions made by other methods, but the shape may be clearly defined and controlled. New possibilities of differentiation between alloying and diffusion are other advantages of the process. Applications have been made to rectifiers and transistors.

Introduction

A USEFUL METHOD for making junction semi-conductor devices consists of alloying with and diffusing into semi-conductors of one conductivity type, materials which, as impurities, give it conductivity of the opposite type.[1,2] For example, indium is an acceptor or *p*-type impurity in germanium and may be introduced into germanium containing an excess of donors (*n*-type impurities) so as to form a rectifying junction.

Control of the area of *p-n* junctions made by such combined alloying and diffusion of indium metal into *n*-type germanium is important to obtain uniformity of

results. For example, simple application of small pieces of indium metal on the surface of the germanium and subsequent heating to the desired temperature gives junctions whose size varies from unit to unit and whose location on the surface may also be somewhat indeterminate. In an attempt to define more clearly the junction area the indium has been electroplated onto the germanium. However, a smooth layer of any thickness greater than a few microns is not readily obtained; the indium tends to agglomerate when melted and the area finally coated is not the original plated region. Although other methods for control are possible, this paper describes a highly successful method using an intermediate metal.

Description of the Method

The germanium is first plated with a thin layer of a metal which alloys readily with indium and which does not subsequently affect the *p-n* junction. The plated area may be accurately controlled in size, shape, and position by customary plating techniques. The alloying metal, indium, is then placed on the plated area and the whole heated to melt the indium. It is found that the molten indium flows freely on the plated region, and it alone is covered. With further heating, diffusion takes place through the intermediate metal layer and into the germanium, forming the desired *p-n* junction with a well-defined area. Such junctions have been made using

[1] W. Shockley, U. S. Patent 2,569,347, applied for June 26, 1948.
[2] R. N. Hall and W. C. Dunlap, "P-n junctions prepared by impurity diffusion," *Phys. Rev.*, vol. 80, p. 467; May, 1950.

thin plated layers of the common metals which are easily "wet" by indium, i.e., silver, gold, copper, and nickel. Fig. 1, which is not drawn to scale, illustrates the principle.

INDIUM

PLATED REGION

GERMANIUM

BEFORE FIRING   AFTER FIRING
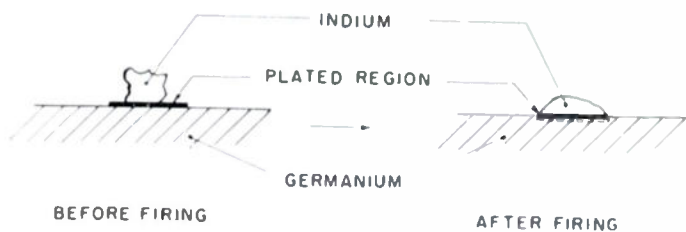
Fig. 1

Particularly satisfactory results have been obtained with plated gold. Although good junctions have been made with layers up to 2 mils in thickness, the best thickness is considerably thinner than this, perhaps of the order of a few hundred Angstrom units. After a first heating, which melts the indium so as to cover the gold and start the diffusion process, the assembly is chemically etched to dissolve excess plating metal which might otherwise cover the junction boundary. A second firing completes the diffusion process. The final etching is similar to that used in the case of junctions made without an intermediate metal layer.

ADVANTAGE OF THE PROCESS

To offset the more complicated firing and etching schedule, in comparison with devices made without the intermediate plated metal, there are several advantages which merit attention.

1. A properly chosen intermediate layer may permit wetting of the junction area at a lower temperature than if no layer is present. Indium will wet and flow on gold and copper platings at temperatures well below that required for wetting on germanium.

2. The shape of the junction area is readily varied. For example, rectangular or elongated junction areas may be made in this way for special applications.

3. Another controlled parameter is introduced into the alloying plus diffusion process. It is known from general metal studies that the presence of a small amount of a third metal in an original two-metal alloying region affects relative solubilities. It is possible to inhibit the solution of one metal into another by the addition of a small amount of a third metal. Experimental evidence indicates that the alloying of germanium into indium is retarded to a considerable extent by the plated layer. Formation temperatures may be raised as much as 50 per cent above those used in ordinary alloying and diffusion investigations without detrimental alloying effects. This retardation of alloying opens new possibilities in the formation of junctions by diffusion.

RESULTS

Rectification characteristics on junctions made as above, and one-half square millimeter in area, generally show of the order of an ampere forward current at 1 volt and only a few microamperes back current at the same voltage. AC impedance in the reverse direction at 1 to 2 volts is as high as 10 megohms in the best units. Transistors made with a junction on each side of the wafer have forward junction currents limited by the base resistance and yield results comparable with those obtained by other techniques.

Lifetime of Injected Carriers in Germanium*

D. NAVON†, R. BRAY†, AND H. Y. FAN†

Summary—A method is described for the determination of the lifetime of injected carriers in a semiconductor by measuring the variation of the sample conductance after a voltage pulse has produced excess carriers. Volume and surface recombinations are separated by varying the sample dimensions. It is shown that carefully etched surfaces give very small surface-recombination velocity. The lifetime in high-resistivity germanium does not correlate critically with the resistivity, but can be changed greatly by heat treatment. Quenching from temperatures above 500°C reduces the lifetime drastically before the resistivity is appreciably affected. Preliminary measurements on some *N*-type as well as *P*-type germanium samples at low temperatures indicate that the lifetime first decreases upon cooling from room temperature, but the *N*-type samples at liquid nitrogen temperatures show a rapid decay of excess conductance superimposed on a much longer decay.

INTRODUCTION

IN A SEMICONDUCTOR the electric current is carried by free electrons in the conduction band and holes in the highest filled band. An ideally intrinsic semiconductor has an equal concentration of free electrons and holes. Because of the presence of impurities a semiconductor may have electrons (*N*-type) or holes (*P*-type), prevalently; whichever of the carriers has the larger concentration is referred to as the majority carrier while the other is called the minority carrier. The equilibrium concentrations of electrons and holes at a given temperature will be called the "normal concentrations." When there is an excess due to injection or light excitation over the normal concentration, there will be a net rate of hole-electron recombination due to the processes restoring equilibrium. The excess

concentrations of holes and electrons are usually present in equal amounts, owing to the tendency of a conducting material to preserve electrical neutrality. If the excess concentrations are small compared to the normal concentration of the majority carriers, then the net rate of hole-electron recombination will be proportional to the excess concentration of the minority carriers, which will decay according to $\exp(-t/\tau)$. The factor τ is called the lifetime of the minority carriers, which is an important factor in the operation of semiconductor devices, in particular the transistor.

Measurement of the lifetime involves the production and the subsequent detection of an excess carrier concentration in the semiconductor. Excess carriers can be produced by injection at metal point contacts or p-n junctions and by light excitation. They are usually detected with the aid of a point-contact collector or a p-n junction.[1,2,3] In the experiments reported here the excess carriers are detected by simply measuring the change in conductance of the sample upon injection and following its decay in time. Special detection arrangements are dispensed with and their characteristics are not involved in the measurement. It was found early by one of the authors that hole injection into N-type germanium[4] as well as electron injection into P-type germanium[5] can be produced through large-area soldered contacts. (Injection can also be obtained in germanium samples with rhodium-plated contacts.) Thus for the sake of simplicity no special emitter is used. The sample to be measured has only to be provided with a lead soldered to each end or be plated at the two ends for the application of pressure contacts. The circuit can, however, be easily modified to provide for the injection by a special emitter to separate the sweeping field from the injection field. The following is a description of the circuit employed, followed by a discussion of some of the results obtained so far:

METHOD

Fig. 1 shows the circuit used. The pulse generator consists of a pulse-forming network keyed by a hydrogen thyratron. A rectifier X_1, biased with a battery B_1, is placed in the output to cut off the bottom of the pulse in order to eliminate ripples. The pulse is applied to the germanium sample which is in the form of a rectangular block. The resistance of the samples is usually of the order of 100 ohms. A pulse duration $T_p = 10$ μsec is used. This applied pulse injects and spreads excess

carriers into the sample. The pulse height V_p is usually adjusted so that the injected carriers drift the full length L of the sample; thus $E_p = V_p/L \sim L/\mu T_p$, where μ is the mobility of the injected carriers. For an N-type sample, $L \sim 1$cm, V_p is ~ 50 volts. The pulse repetition rate is kept low, 25 per second, in order to prevent heating the sample. Also measurements with different V_p, T_p and different repetition rates confirm that there is no appreciable temperature fluctuation during the on and off periods to affect the results of the measurements.

To measure the R of the resistance sample, a small direct current is maintained by battery B in series with a resistor $R_1 \gg R$ which keeps the current through the sample essentially constant, independent of small changes of R. The voltage across the sample is divided between the resistor $R_2 \gg R$ and biased rectifier tube
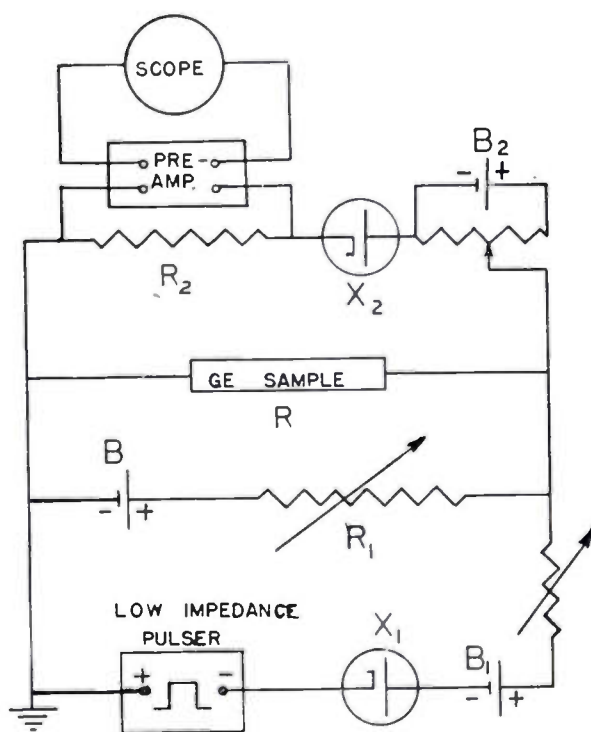
Fig. 1—Schematic diagram of the electronic circuit for the measurement of the time decay of excess conductance of a sample, produced by pulse-injected carriers.

X_2. During a pulse the large voltage across the sample appears mainly across X_2, whereas in the pulse-off period the small dc voltage V_{dc} across the sample appears across R_2. By a high-speed Tektronix oscilloscope one observes this latter voltage which, because of constancy of current, is proportional to the resistance of the sample. Figs. 2 and 3 show some typical pulse patterns appearing on the oscilloscope. By studying the variation of this resistance with time, one can determine the decay rate of excess carriers ΔN in the sample since it will be shown that $\Delta R \propto \Delta N$. A small dc voltage is always present across the sample. This produces an electric field which tends to sweep carriers out of the sample after the injecting pulse, which sweeps

[1] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., New York, N. Y., pp. 60–66; 1950.

[2] F. S. Goucher, "Measurement of hole diffusion in n-type germanium," *Phys. Rev.*, vol. 81, p. 475; 1951.

[3] 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, p. 637; 1951.

[4] R. Bray, "Dependence of resistivity of germanium on electric field," *Phys. Rev.*, vol. 76, p. 152; 1949.

[5] R. Bray, "The barrier layer on p-type germanium," *Phys. Rev.*, vol. 76, p. 458; 1949.
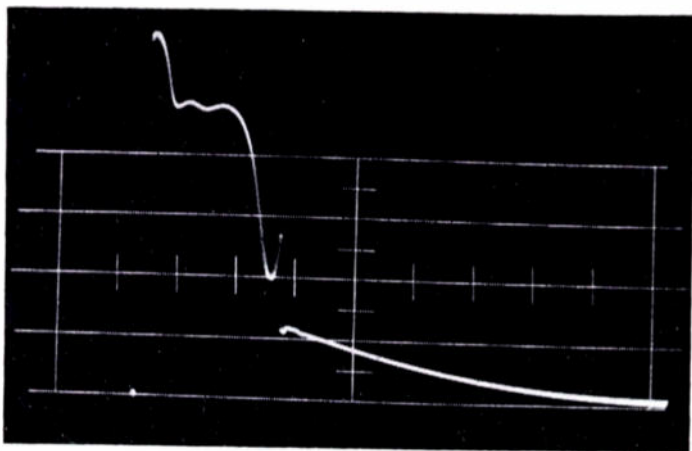
Fig. 2—Photograph of a pulse pattern appearing on the oscilloscope shown schematically in Fig. 1. The pulse at the left is due to the injecting pulse which is clipped by the biased rectifier tube X_2. The pips at the beginning and end of the pulse are due to the capacitance of the rectifier. The decay curve following the pulse shows voltage variation across the sample due to the decay of excess carriers. The sample used was single-crystal N-type germanium, $\rho = 7$ ohm-cm, $(0.10 \times 0.47 \times 1.12)$cm³, with ground surfaces. The time intervals indicated by the markers are 5μ sec/division. The measured lifetime of the sample is 15 μsec.

Fig. 3—Photograph of a pulse pattern for a sample of single-crystal N-type germanium, $\rho = 25$ ohm-cm, $(0.20 \times 0.22 \times 0.44)$ cm³, with etched surfaces. The time intervals here are 25 μ sec/division. The measured lifetime of the sample is 80 μ sec.

them in, falls to zero. Carriers are being lost not only due to recombination but also because they are continually being swept out; the resulting decay curves then observed are not strictly exponential. A formulation will be made to take into account this "sweep-out," enabling us to compute the recombination rate directly.

Before injection, the dc field in the germanium E_x is constant and given by V_0/L. At the end of the pulse some excess carriers are left in the sample and E_x may vary with x, the distance along the sample from the injecting end. If the excess carrier concentrations are everywhere small compared to normal concentration of the majority carrier, then change in resistance of the sample ΔR will now be shown to be proportional to the total number of excess carriers in the sample irrespective of their spatial distribution therein.

Neglecting the diffusion effect, the current density is always

$$i_x = e(\mu_e n_e + \mu_h n_h)E_x,$$

where e is the magnitude of the electron or hole charge, μ_e and μ_h are the mobilities, and n_e and n_h are the concentrations; subscripts e and h refer to electrons and holes, respectively.

Because of the presence of an excess concentration of holes Δn_h and electrons Δn_e over the equilibrium values, let

$$n_e(x, y, z) = n_{e0} + \Delta n_e$$
$$n_h(x, y, z) = n_{h0} + \Delta n_h$$
$$E_x(x, y, z) = E_{x0} + \Delta E_x = V_{dc}/L + \Delta E_x,$$

where V_{dc}/L is the average field.
One then obtains for the new current density

$$i_x = e\{\mu_e(n_{e0} + \Delta n_e) + \mu_h(n_{h0} + \Delta n_h)\}[E_{x0} + \Delta E_x].$$

Integrating this over the volume of the sample, we get

$$I_xL = e \int_\Omega \{\mu_e(n_{e0} + \Delta n_e) + \mu_h(n_{h0} + \Delta n_h)\}[E_{x0} + \Delta E_x]dxdydz,$$

where I_x is the current.
For $\Delta E_x \ll V_{dc}/L$ and $\Delta n_e = \Delta n_h = \Delta n \ll n_{e0}$, where n_{e0} is the normal concentration of the majority carrier (assuming N-type germanium), we get, by neglecting second-order terms,

$$I_xL = G_0V_{dc}L + \frac{eV_{dc}}{L}(\mu_e + \mu_h)\int_\Omega \Delta n dxdydz + e(\mu_e n_{e0} + \mu_h n_{h0})\int dy \int dz \left[\int_0^L dx\Delta E_x\right].$$

G_0 is the normal conductance of the sample. Thus since the integral $\int_0^L dx\Delta E_x$ vanishes

$$I_xL = G_0V_{dc}L + \frac{eV_{dc}}{L}(\mu_e + \mu_h)\Delta N,$$

from which we get for the change in conductance

$$\Delta G = \frac{e}{L^2}(\mu_e + \mu_h)\Delta N. \qquad (1)$$

Since for the case under consideration $\Delta G \ll G_0$, we have

$$\Delta R = -\left(\frac{1}{G_0}\right)^2 \Delta G$$

$$= \left[-e\left(\frac{1}{G_0L}\right)^2(\mu_e + \mu_h)\right]\Delta N, \qquad (2)$$

Thus the variation of V_{dc}, which is proportional to R, gives the decay rate of ΔN, the total number of excess holes or electrons in the sample.

As we previously stated, in order to determine the hole-electron recombination rate within the sample a formulation must be made which corrects for the sweeping out of excess carriers by the measuring current. Let us consider the case of the injection of holes into an N-type sample. At the end of the pulse of duration T_p the carriers have penetrated a distance $d = \mu_h E_p T_p$ into the sample, E_p being the field in the sample due to the pulse. In case d is larger than the length of the sample L, it should be replaced by L in the following consideration. The average concentration of excess holes over a cross section at the end of the pulse ($t = 0$) is given by

$$\overline{\Delta n} = \overline{\Delta n_0} \exp\left(- x/\mu_h E_p \tau\right), \qquad (3)$$

where $x/\mu_h E_p$ is the time elapsed since the injection of the holes found at the cross section x, and τ is the lifetime. $\overline{\Delta n}$ decays by recombination according to $\exp(-t/\tau)$, and at the same time the whole distribution of excess holes is moved back toward the injection end at a velocity $\mu_h E_{dc}$, E_{dc} being the field due to the dc bias supplying the current for measurement. Therefore, the total number of excess holes in the sample at time t, when the hole distribution is moved back by a distance $d' = \mu_h E_{dc} t$ and independently the excess carriers at any place have decayed exponentially in time, is given by

$$\Delta N(t) = \int_{d'}^{d} A \overline{\Delta n_0} \exp\left(- x/\mu_h E_p \tau\right) \exp(-t/\tau) dx \qquad (4)$$

$$= A \overline{\Delta n_0} \mu_h E_p \tau \left[\, \exp\left(- E_{dc} t/E_p \tau\right) \right.$$
$$\left. - \exp\left(- T_p/\tau\right)\right] \exp(-t/\tau),$$

A being the cross-sectional area of the sample. In practice $E_p \gg E_{dc}$, E_p being of the order of 50 volts/cm whereas E_{dc} is a fraction of 1 volt/cm. Furthermore, on account of the factor $\exp(-t/\tau)$, the values t of interest are not many times larger than τ. Therefore, $(E_{dc}t/E_p\tau) \ll 1$, and we have approximately

$$\Delta N(t) = A \overline{\Delta n_0} \mu_h E_p \tau \left[1 - (E_{dc}t/E_p\tau) \right.$$
$$\left. - \exp\left(- T_p/\tau\right)\right] \exp(-t/\tau).$$

For lifetimes $\tau \gg T_p = 10\ \mu\text{sec}$, the same approximation can be made for $\exp(-T_p/\tau)$, and we get

$$\frac{\Delta N(t)}{[1 - E_{dc}t/E_p T_p]} = \left[(\overline{\Delta n_0})(A\mu_h E_p T_p)\right] \exp(-t/\tau) \qquad (5)$$

$$= [N_0] \exp(-t/\tau).$$

N_0 represents the total number of excess holes in the sample at the termination of the injecting pulse if no appreciable recombination takes place during the pulse. On the other hand, for values of $\tau \gtrsim T_p$, $(E_{dc}t/E_p\tau)$ is completely insignificant compared to the time independent terms inside the bracket, and the approximation becomes

$$\Delta N(t) = \left\{\left[A\overline{\Delta n_0}\mu_h E_p\tau\right][1 - \exp(- T_p/\tau)]\right\} \exp(- t/\tau)$$

$$= \text{const} \exp(- t/\tau). \qquad (6)$$

We see in (5) that the decay of ΔN is not a simple exponential function of time if τ is $\gg T_p$. For long τ, many carriers are swept out of the sample during the time of measurement. The deviation due to the sweep-out effect is larger the greater (E_{dc}/E_p). This is borne out by the experimental observations. When different values of (E_{dc}/E_p) are used on large τ samples, the observed ΔV_{dc} decay curves take on different shapes, becoming clearly nonexponential as the ratio (E_{dc}/E_p) is increased. But all the curves yield the same value for τ when interpreted according to (5); straight lines with the same slope are obtained if $ln[\Delta V_{dc}/(1 - E_{dc}t/E_p T_p)]$ is plotted against t. For $\tau \gtrsim T_p$, the decay curves are observed to be exponential as predicted by (6).

Results

A. Surface-Volume Recombination

It has been shown[1] that excess carriers not only recombine within the volume of the sample but also at the surfaces with a rate dependent on the condition of the surfaces. It is then necessary to separate the two effects in order to study the factors influencing either one process. In order to do this the variation of the lifetime of injected carriers was studied systematically as a function of the dimensions of the sample for various surface conditions. The sample used was N-type, single-crystal germanium resistivity $\rho = 19$ ohm-cm at room temperature with dimensions $0.371 \times 0.737 \times 0.787\ \text{cm}^3$. A measurement of the lifetime of injected holes was made both with the sample surfaces roughly ground[6] and with the surfaces carefully etched.[7] The sample was then ground down to various dimensions and the lifetime was measured at each stage with ground and

TABLE I

$(2B \times 2C)$ (cm $\times$ cm)	τ_g (μ sec)	τ_e (μ sec)
0.371×0.737	144	280
0.297×0.730	120	270
0.202×0.716	78	340
0.124×0.708	24	300
0.100×0.705	16.5	290
0.080×0.6	11.0	310
0.075×0.48	10.3	
0.071×0.48	9.2	280
0.061×0.48	7.6	330
0.053×0.48	6.0	
0.036×0.48	3.1	235

etched surfaces. Table I lists the lifetime values obtained in this way, where τ_g represents the measured lifetime with the surfaces ground and τ_e with the surfaces etched. The injecting voltage was chosen so that the holes would drift over the full length of the sample.

[6] The surface was ground with 600 carborundum powder.
[7] Purdue etch (HNO$_3$+HF+Cu(NO$_3$)$_2$+H$_2$O) was used. The sample was washed in water and acetone after each etch.

The measured life, τ, can be written[1]

$$\frac{1}{\tau} = \frac{1}{\tau_s} + \frac{1}{\tau_v}, \qquad (7)$$

where $1/\tau_v$ is the bulk recombination rate and $1/\tau_s$ is the recombination rate due to surface effects. $1/\tau_s$ is determined by the cross-sectional dimensions of the sample $(2B \times 2C)$ and its surface condition, according to

$$\frac{1}{\tau_s} = 4D_h\left[\frac{\eta^2}{(2B)^2} + \frac{\zeta^2}{(2C)^2}\right], \qquad (8)$$

where D_h is the hole diffusion constant and η and ζ are given by

$$\frac{SB}{D_h} = \eta \tan \eta; \qquad \frac{SC}{D_h} = \zeta \tan \zeta. \qquad (9)$$

Here S is the so-called surface recombination velocity.

Fig. 4 gives the plot of $1/\tau$, measured with ground surface, against $[\eta^2/(2B)^2 + \zeta^2/(2C)^2]$ for two different
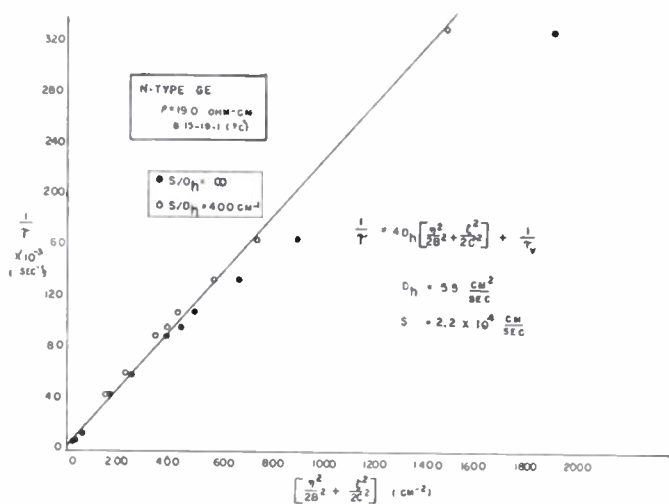
Fig. 4—Plot of the hole-electron recombination rate measured with ground surface against the factor $[\eta^2/(2B)^2+\zeta^2/(2C)^2]$. The black solid dots are for $S/D_h = \infty$, while the open circles are for $S/D_h = 400$ cm^{-1}. The latter points follow approximately the straight line shown. The slope of this line gives $D_h=55$ cm/sec^2 and $S=2.2\times10^4$ cm/sec for the ground surface.

assumed values of S/D_h. The points do not follow a straight line for $S=\infty$. A good fit is obtained for $S/D_h=400$ cm^{-1}. The slope of the straight line gives $D_h=55$ cm^2/sec which is somewhat higher than the usual values reported. This is likely due to the inaccuracies in such estimation. The S value for the ground surface is then 2.2×10^4 cm/sec. The ordinate intercept indicates a bulk lifetime of approximately 400 μsec.

From Table I it is apparent that the etched-surface lifetime does not vary with the dimensions of the sample until perhaps its small dimension is 0.05 cm or less.[8]

---

[8] Difficulty was experienced in getting nonrectifying contacts and a good etched surface on samples of thicknesses of 0.05 cm or less.

The scattering of the τ_e values is probably due to the experimental error involved in making these measurements. To insure that a good (low S) surface was obtained each time, the sample was re-etched (2 minutes per etch) a few times until a constant value of measured τ_e was reached. Since this measured lifetime is independent of the dimensions of the sample over a wide range of thicknesses, it must be the volume lifetime. The effect of surface recombination must be unimportant in this case. For small surface recombination velocity, (2) reduces to

$$\frac{1}{\tau_s} = 2S\left[\frac{1}{(2B)} + \frac{1}{(2C)}\right]. \qquad (8a)$$

From this it can be shown that the surface-recombination velocity for our etched surfaces cannot be greater than 100 cm/sec. This is much lower than values previously reported for etched surfaces.[1]

Using samples with carefully etched surfaces and adequate (based on the previous measurements) dimensions, the bulk lifetime of germanium of various resistivities and from different melts was studied. The results seemed to indicate that the volume recombination rate was not very sensitive to the resistivity of the sample. For example, samples from different melts of approximately the same resistivity had different bulk lifetimes. Also heat treatment of samples produced manifold changes in bulk lifetime while the resistivity hardly varied, as is discussed in the next section.

B. Heat-Treatment Studies

The mechanism of hole-electron recombination is not yet completely understood. Estimations on the basis of direct recombination with photon emission[9] indicate a lifetime of the order of one second. Bulk lifetimes no longer than several hundred microseconds have been observed in this laboratory at room temperature although lifetimes of a few thousand microseconds have been recorded in other laboratories. Quenching from a high temperature, suggested by Lark-Horovitz, had an appreciable effect on the volume recombination rate. A study was then begun on the variation of bulk lifetime with heat treatment.

The sample was placed in an evacuated furnace and held at a particular temperature for one hour. Then the sample was quenched by permitting it to fall into an oil bath at room temperature. Lifetime and resistivity were checked at room temperature following each quench. A piece of uniform n-type germanium single crystal, 23 ohm-cm with a bulk lifetime of 375 μsec, was quenched from temperatures ranging between 500 and 590°C. Fig. 5 shows a semilog plot of the bulk recombination rate of the quenched sample as a function of the reciprocal of the quenching temperature. Evidently additional recombination centers are introduced by

---

[9] H. Y. Fan, Purdue Semiconductor Research Progress Report, p. 6; March, 1949. See also footnote reference 1, p. 69.

such treatment which may be defects formed in the germanium lattice. Assuming that the equilibrium number of recombination centers at a heating temperature T are "frozen in" by the quench, the number of such centers is given by[10]

$$N_{rc} \propto \exp(-\Delta W/kT), \qquad (10)$$
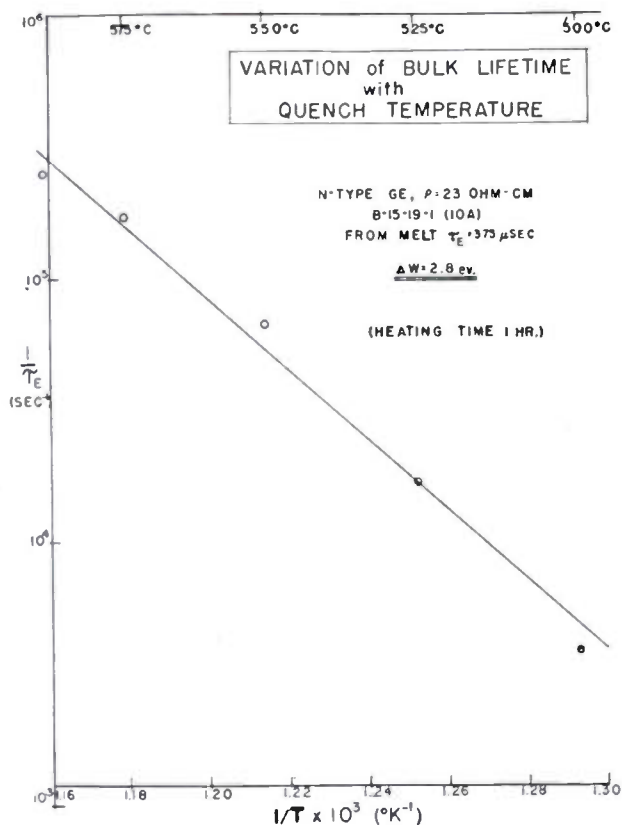
Fig. 5—The hole-electron recombination rate in a sample after quenching from different temperatures as measured with carefully etched surfaces plotted against the reciprocal of the quenching temperature expressed in degrees K. The slope of the straight line α gives $|\alpha k| = 2.8$ ev.

where k is the Boltzmann constant and ΔW is the activation energy necessary to produce one such recombination center. If we assume that recombination in a perfect lattice is very slow, the measured bulk recombination rate, $1/\tau_e$, should be proportional to N_{rc}. The slope α of $ln(1/\tau_e)$ plotted against $1/T$ should then be $-\Delta W/k$. Fig. 5 gives $|\alpha k| = 2.8$ ev. Actually, the time

[10] W. E. Taylor and K. Lark-Horovitz, Purdue Semiconductor Research Progress Report; December, 1949.

(one hour) for which the sample was kept at the various temperatures may not have been enough for the equilibrium to be reached, so that ΔW may be actually different from this value. Experiments with various heating times are being made. However, one conclusion can be drawn. Although the lifetime of the sample had changed due to these quenches, the resistivity did not vary appreciably.

If the introduction of recombination centers is a thermally reversible effect, one should be able to remove these by a proper annealing procedure. This was attempted by heating the quenched sample at 475°C for 24 hours and cooling it to room temperature in about 5 hours. This treatment only succeeded in bringing the lifetime back to 200 μsec as compared to the initial value 375 μsec. The reason for this is not as yet clear.

Sometimes the lifetime of a germanium sample cut from a grown crystal may be increased by annealing since the original melt may have been cooled too fast. A piece of high-resistivity n-type germanium was cut from a single crystal drawn from the melt and given a 40-hour anneal at 475°C. Measurements were made with etched surfaces. Before this treatment the lifetime was 430 μsec. After annealing, the lifetime had increased to 640 μsec with no appreciable change in the resistivity.

C. Low-Temperature Results

Preliminary studies have been made on the variation of lifetime with temperature. One p-type sample, $\rho = 17$ ohm-cm, and two n-type samples, $\rho = 4.5$ and 19.0 ohm-cm, were used. The lifetime as observed both with ground and etched surfaces in the p-type sample decreased monotonically as the temperature was reduced to 100°K. The lifetime in the n-type sample also decreased with decreasing temperature until about 250°K. However, at lower temperatures the decay of the excess carriers was no longer exponential. It appeared that a rapid decay was superimposed on a much longer decay. At 175°K and lower, the long tail on the decay curves indicated that there were excess carriers left in the sample long after they should have been swept out by the dc bias always present across the sample. Some photoconductive measurements were made on the n-type samples at different temperatures using pulsed white light. The lifetime deduced was in qualitative agreement with the above results. These investigations are being continued.

Silicon *P-N* Junction Alloy Diodes*

G. L. PEARSON† AND B. SAWYER†, ASSOCIATE, IRE

Summary—A new type of *p-n* junction silicon diode has been prepared by alloying acceptor or donor impurities with *n-* or *p*-type silicon. The unique features of this diode are: (a) reverse currents as low as 10^{-10} amperes, (b) rectification ratios as high as 10^8 at 1 volt, (c) a Zener characteristic in which $d(\log I)/d(\log V)$ may be as high as 1,500 over several decades of current, (d) a stable Zener voltage which may be fixed in the production process at values between 3 and 1,000 volts, and (e) ability to operate at ambient temperatures as high as 300°C.

I. INTRODUCTION

SILICON POINT-CONTACT DIODES[1] were developed some years ago for use in microwave radar receivers and are now in large-scale commercial production. This paper describes the preparation and properties of a new type of silicon diode, namely, the *p-n* junction diode prepared by alloying.[2] The unique features of this unit are: (a) extremely low reverse currents (of the order of 10^{-10} amperes), (b) large rectification ratios (as high as 10^8 at 1 volt), (c) ability to operate at high ambient temperatures (rectification ratios as high as 10^6 at 100°C, 10^4 at 200°C, and 10 at 300°C), (d) a flat Zener characteristic over several decades of current (for reverse voltages above a given critical value $I = KV^n$ with *n* as high as 1,500), and (e) ability to operate usefully at frequencies up to 20 mc, in contrast to the microwave frequencies for point contacts.
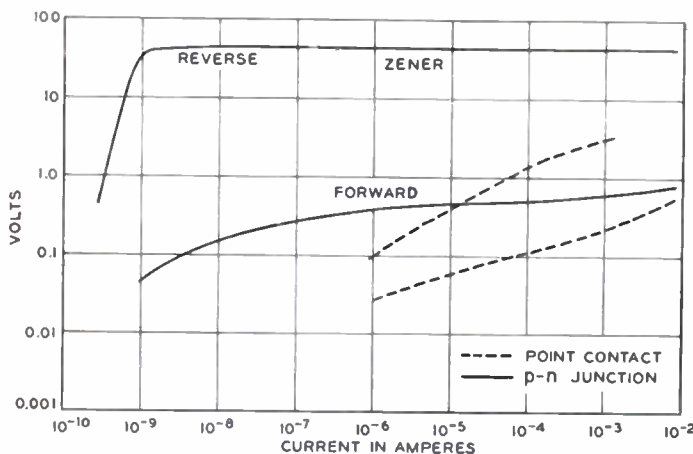
Fig. 1—Current versus voltage characteristics of a commercial silicon point contact diode and a silicon *p-n* junction diode prepared by alloying.

Fig. 1 compares the dc current voltage characteristics of a typical commercial point-contact silicon diode with those of an experimental model alloyed silicon *p-n* junction diode. Many of the unique features of the newly developed unit are clearly evident in this plot.

The growth of single-crystal silicon *p-n* junctions from

a melt has recently been described by Teal and Buehler.[3] The electrical properties of these diodes as reported by McAfee and Pearson[4] are quite similar to those described in this paper; *p-n* junction diodes prepared from single-crystal germanium as pulled from the melt have been described by Pietenpol[5] and germanium *p-n* junction diodes prepared by a diffusion process have been described by Hall and Dunlap.[6] Both types of germanium diodes have many useful properties, but the silicon units described here have many advantages which may be accounted for in part by the wider energy gap in silicon (1.12 electron volts as compared to 0.75 electron volts in germanium). In particular, silicon diodes have lower reverse currents, higher rectification ratios, and can operate at much higher ambient temperatures.

II. PREPARATION OF SILICON *p-n* JUNCTIONS BY ALLOYING

The *p-n* junctions are prepared by alloying donor- and acceptor-type electrodes to a homogeneous silicon crystal of either *n-* or *p*-type (for the present discussion we will assume an *n*-type crystal). The electrodes can be attached to the silicon crystal by heating the crystal and then bringing it in contact with the metals to which
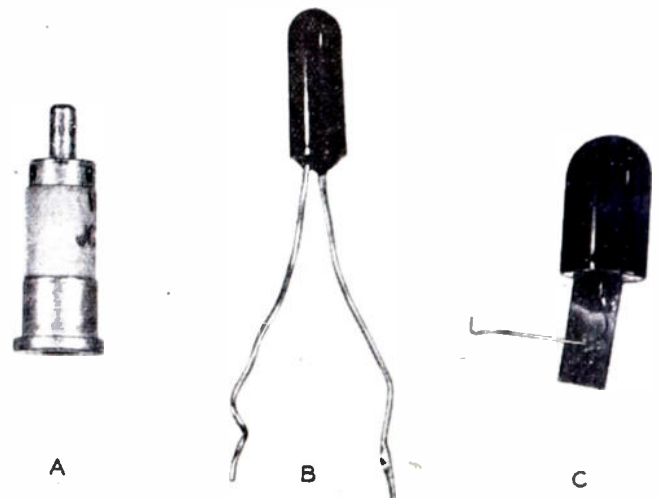
Fig. 2—Photograph of silicon diodes. (a) A commercial point-contact unit. (b) and (c) Developmental models of *p-n* junction diodes prepared by alloying.

it will alloy. A large variety of metals may be alloyed in this manner; but to prepare a *p-n* junction diode from *n*-type silicon we choose a metal such as aluminum from the third column of the periodic table. We believe that the *p-n* junction is grown by deposition from the

* Decimal classification: R282.12. Original manuscript received by the Institute, August 18, 1952.

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

[1] J. H. Scaff and R. S. Ohl, "Development of silicon crystal rectifiers for microwave radar receivers," *Bell Sys. Tech. Jour.*, vol. 26, p. 1; January, 1947.

[2] A preliminary account of the material described here was presented at the A.P.S. meeting in Washington, D. C., May, 1952. G. L. Pearson and P. W. Foy, "Silicon *p-n* junction diodes prepared by the alloying process," *Phys. Rev.*, vol. 87, p. 190; July, 1952.

[3] G. K. Teal and E. Buehler, "Growth of silicon single crystals and of single crystal silicon *p-n* junctions," *Phys. Rev.*, vol. 87, p. 190; July, 1952.

[4] K. B. McAfee and G. L. Pearson, "The electrical properties of silicon *p-n* junctions grown from the melt," *Phys. Rev.*, vol. 87, p. 190; July, 1952.

[5] W. J. Pietenpol, "*p-n* junction rectifier and photo-cell," *Phys. Rev.*, vol. 82, p. 120; April, 1951.

[6] R. N. Hall and W. C. Dunlap, "*p-n* junctions prepared by impurity diffusion," *Phys. Rev.*, vol. 80, p. 467; November, 1950.

molten-alloy phase during the cooling cycle and is situated at the interface between the unmelted silicon and the frozen-out primary solid solution. To form an ohmic or base contact, one should use a metal other than one from the third column of the periodic table.

Fig. 2 is a photograph of three silicon diodes. The unit on the left is a commercial point-contact unit manufactured by the Western Electric Company. The other two are alloyed *p-n* junctions. The unit in the center has wire electrodes and the one on the right is thermally coupled to a metal fin to permit greater heat dissipation.

III. Theory of the Step *p-n* Junction Diode

The theory of *p-n* junctions in semiconductors as developed by Shockley[7] may be applied to alloyed silicon diodes. The current-voltage characteristic at not too high voltages in either direction is given by

$$I = I_0(e^{-qV/kT} - 1). \qquad (1)$$

where I is the current in amperes per cm^2, V is the voltage across the junction (plus for current in the forward direction and minus in the reverse), q is the electronic charge, k is Boltzmann's constant, T is the absolute temperature, and I_0 is a constant determined by the properties of the silicon. It is found that (1) is correct for forward currents in silicon *p-n* junctions, but that the reverse currents do not saturate as expected. This has led McAfee to suggest[4] that because of the short lifetimes of the minority carriers in silicon, the reverse current due to carriers generated by traps within the space-charge region is greater than the saturation current of minority carriers diffusing into the space-charge region. This trap reverse current is proportional to the width of the space-charge region.

For units made from *n*-type silicon, the impurity concentrations are step functions (in the sense that the change from *p*- to *n*-type takes place abruptly within the space-charge layer) so that Poisson's equation becomes

$$\partial^2 V/\partial x^2 = 4\pi q N_0/K, \qquad (2)$$

where N_0 is the density of excess donor impurities in the silicon crystal and K the dielectric constant ($=12.1$ for silicon). The solution of (2) indicates that for reverse biases, the width of the space-charge region is proportional to the square root of the voltage and hence the capacity of the junction is inversely proportional to the square root of the voltage.

Fig. 3 is a plot of capacitance as a function of reverse bias obtained on a *p-n* junction prepared from 0.6-ohm cm *n*-type silicon by the methods described above. The open circles are experimental points, and they show that for reverse biases above 2 volts the relationship is an inverse square law, as predicted by theory. At low voltages the points fall below the theoretical line because of a built-in voltage at the junction of the opposite

polarity. If 0.75 volts are added to each experimental point, the solid circles are obtained and they continue on the square-law line. This, of course, is a measure of the built-in voltage at the *p-n* junction and agrees with the theoretical calculation for 0.6 ohm cm silicon.[8] The Zener voltage of this junction was 21.5 volts and is indicated by the horizontal line. The maximum field at this voltage is 250,000 volts per cm and is in fair agreement with Shockley's extension[9] of the formulas derived by Zener for the field at which tunneling between the valence and the conduction bands takes place. The measured value of $d(\log I)/d(\log V)$ in the Zener region, as shown in the following section, is much higher than the theoretical prediction,[9] and is not yet understood.
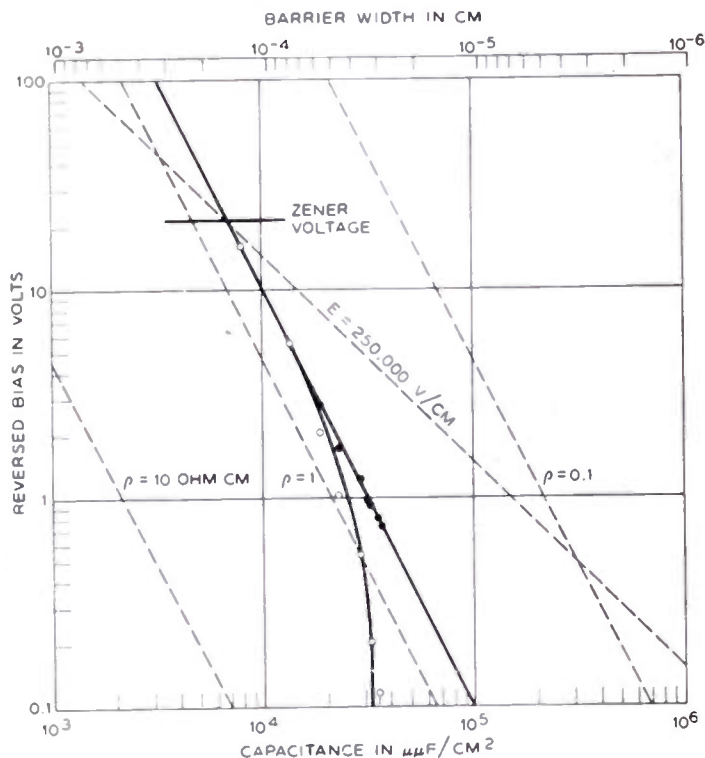
Fig. 3—Capacitance versus voltage characteristic for an alloyed silicon *p-n* junction. A dielectric constant of 12.1 and an electron mobility of 1,200 are assumed.

In Fig. 3 the dotted lines obtained from theory indicate the capacity of alloy diodes prepared from *n*-type silicon of various resistivities. Since the critical Zener field is a constant for any resistivity material, the Zener voltage is directly proportional to the resistivity and is given by the intersection of the constant ρ lines with the Zener field line. For *n*-type silicon, theory indicates that the Zener voltage is 39 times the resistivity ρ in ohm cm and that for *p*-type material the multiplying factor is 8. The scale at the top of Fig. 3 indicates the width of the *p-n* junction space-charge layer as a function of capacity.

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

[8] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Company, Inc., New York, N. Y., p. 240; 1950.
[9] K. B. McAfee, E. J. Ryder, W. Shockley, and M. Sparks, "Observations of Zener current in germanium *p-n* junctions," *Phys. Rev.*, vol. 83, p. 650; August, 1951.

IV. Electrical Properties

Fig. 4 is a log plot of the current-voltage relationships of two representative silicon junction diodes made by alloying. The junction areas of these units are estimated at 0.001 cm². The reverse characteristics show very low currents of the order of 10^{-9} amperes, and then an abrupt transition to the Zener behavior wherein
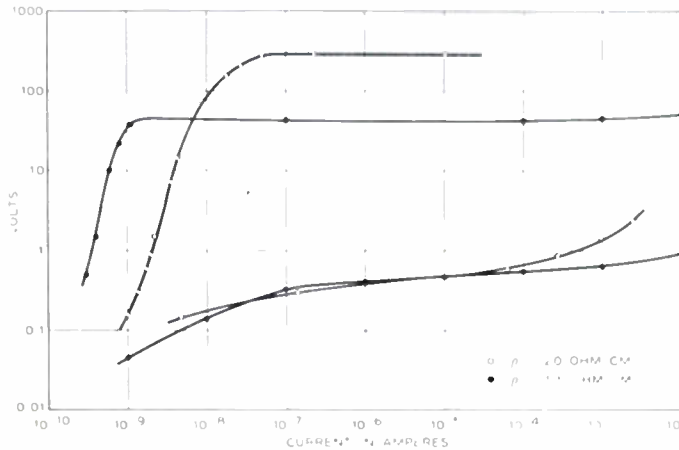
Fig. 4—Current versus voltage characteristics of silicon *p-n* junction diodes. A linear relationship between Zener voltage and silicon resistivity is shown.

the current increases at very nearly constant voltage over as many as six decades. The ac impedance changes from many megohms to values of from ten to a few hundred ohms, $d(\log I)/d(\log V)$ having values as large as 1,500. The Zener voltages of the units shown are at about 40 and 300 volts. The forward characteristics are also shown, making it easy to calculate the rectification ratios at 1 volt as about 10^7 and 10^5. Possible reasons for the differences between these units will be discussed in the next section under the heading of Design Theory.
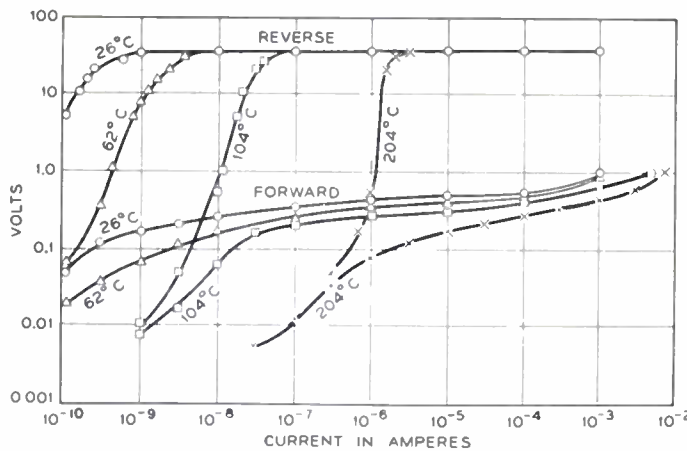
Fig. 5—Current versus voltage characteristics of a silicon *p-n* junction diode at various ambient temperatures.

Fig. 4 has presented evidence of the low reverse currents, high rectification ratios, and flat Zener characteristics of silicon *p-n* junctions. Fig. 5 shows characteristics of one silicon *p-n* junction diode at various temperatures. The greatest variation with temperature is seen to be in the reverse current which increases rapidly with increasing temperatures. Nevertheless, in

this diode the room temperature value of the reverse current is so low that, in spite of the rapid thermal increase, it only reaches about 1 μa at 200°C. Forward current in the diode also increases with increasing temperature, but not so rapidly as the reverse. Hence, the rectification ratio at 1 volt for this diode decreases from about 10^8 at room temperature to about 10^4 at 200°C.
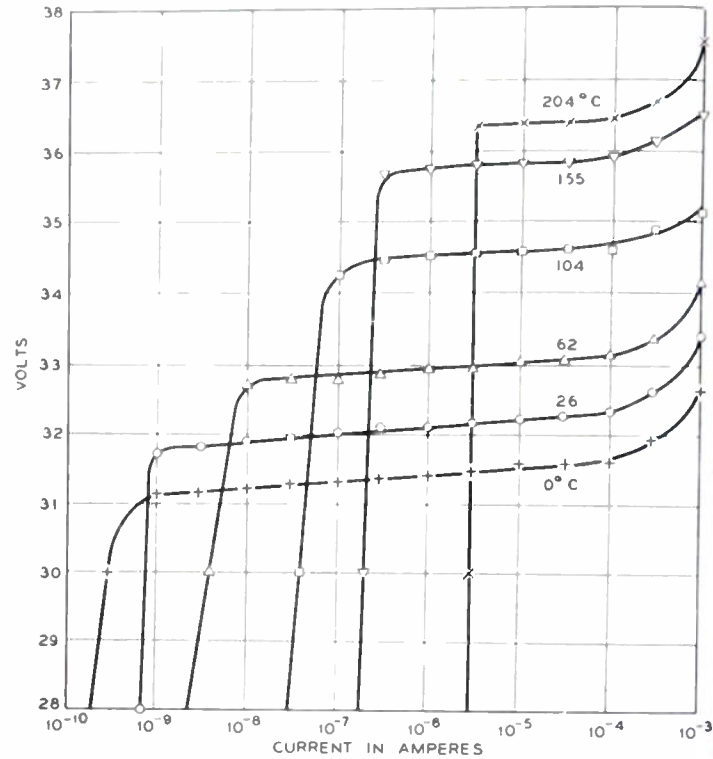
Fig. 6—Current versus voltage characteristics in the Zener region for a silicon *p-n* junction at various ambient temperatures.

The thermal effect on the Zener voltage cannot be seen in Fig. 5. Hence Fig. 6, a semilog plot of the same data arranged to magnify the region around 34 volts, is included. Here it can be seen that the Zener voltage
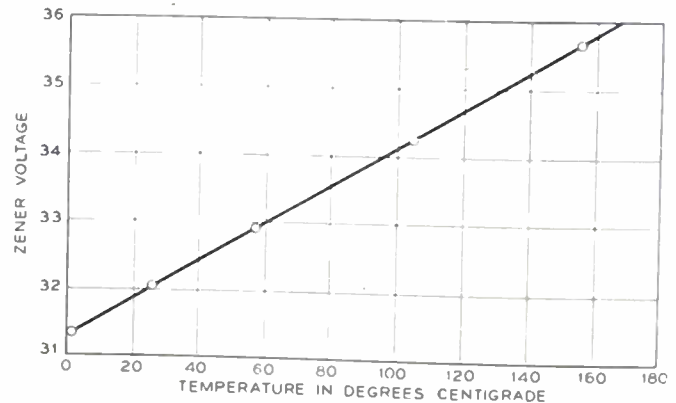
Fig. 7—Zener voltage versus ambient temperature for a silicon *p-n* junction. Current held constant at 1×10^{-6} amperes.

increases slowly with increasing temperature. Fig. 7 shows that the Zener voltage (taken arbitrarily as the voltage necessary to cause 1 μa to flow) varies linearly with temperature, and from this it can be determined that the change is at the rate of 0.084 per cent per °C, within the same order of magnitude as the value of 0.01 per cent per °C quoted for a VR75 electron tube.

The maximum power dissipation in units presently made is from 0.2 to 0.5 watts (viz., Fig: 2(b) and (c)), depending on the junction area and on the amount of provision for removal of heat. The high-frequency cut-off in units presently made ranges from about 1 to 20 mc, depending on the capacity in the junction.

A discussion of the electrical properties of silicon junction diodes would not be complete without some mention of the chief defects present in some units as presently made. These are two in number and might be called noise at the Zener knee and "softness" of the reverse characteristic, or a lack of saturation of the reverse current, as shown in Fig. 8. The noise is a phenomenon varying widely from unit to unit, and apparently not inherent in silicon junctions since some units had no noise detectable with an instrument sensitive to 1 mv. On the other hand, some noisy units produced peaks of noise as high as 3 volts (in addition to the Zener voltage). The noise is almost always limited to the low current region of the Zener characteristic, sometimes appear to be "clipped" at a fixed voltage, is generally white, and, in some units, is temperature dependent. Noise properties vary so much from unit to unit that many details of behavior in any one unit have not been considered significant. There is evidence that noise in some units may be associated with mechanical cracks at the junction.

Fig. 8—Current-voltage oscilloscope traces of silicon *p-n* junction diodes showing: (a) Typical good characteristic; (b) Noise at the Zener knee; (c) Softness of reverse characteristic.

"Softness" of the reverse characteristic means an unusual increase in the reverse current before the true Zener voltage is reached. This defect can also vary widely from unit to unit, and is sometimes improved after the units have been heated during testing. This defect may be due to leakage conduction around the junction, to minority carrier emission in an ohmic contact, or to various kinds of crystal lattice defects in the junction itself.

V. Design Theory

The design theory of silicon junction diodes is still in an early stage of development. Nevertheless, some remarks are possible largely on the basis of step *p-n* junction theory.

Probably the most interesting dependent variable to be fixed in designing a silicon junction diode is the Zener voltage. Theoretically, this voltage should be the voltage just sufficient to bring the maximum electric field within the junction region to the critical value of 250,000 volts per cm. Thus the relationship of the elec-

tric field inside the junction to the reverse voltage applied externally is critical in determining the Zener voltage for that unit. For the case of a "step" junction, wherein the transition from *n*-type to *p*-type silicon is abrupt, the relationship should depend only on the resistivities on each side of the junction so that the expected equation for Zener voltage becomes

$$V_Z = 39\rho_n + 8\rho_p. \tag{3}$$

In the alloy junction cases, the resistivity of the alloy doped side of the junction is negligible compared to the other, and that term in the equation drops out, leaving the theoretical Zener voltage to be a linear function of the initial silicon resistivity alone.

The relationship for *n*-type silicon with an alloy bearing *p*-dope, $V_Z = 39\rho_n$, has been experimentally tested. The general trend predicted is present as exemplified by the two diodes in Fig. 4; but when a plot of V_Z versus ρ_n is made, the points scatter, indicating that some other factors besides ρ must be affecting V_Z.

Nevertheless, fairly good reproducibility of V_Z is attainable, as evidenced by an experiment in which 12 alloy units were made from one slice of silicon ($\rho_n = 0.5$) under controlled conditions. The Zener voltages of all units were within the range of 20 ± 0.5 volts or ± 2.5 per cent.

Another important variable to control is the reverse current. This should vary directly with the number of traps in the space-charge region (or with the inverse square root of the lifetimes of minority carriers). The factors leading to "softness" of the reverse characteristic mentioned above may cause spurious increases in the reverse current.

The geometry of the junction diode can have an important bearing on some properties. The area of the junction itself is especially important in that all of the currents through the junction, forward, reverse, and Zener, will vary in proportion to this area. Unfortunately, the capacity of the junction which must be minimized if high-frequency service is desired will also vary in proportion to the junction area.

A series resistance due to the bulk silicon resistivity may become appreciable compared to the forward and Zener resistances of the junction, especially when resistivities are high. This is controlled by the areas of the ohmic contacts, their distances from the junction, and by the silicon resistivities.

Even though development work is still in an early stage and the present picture of silicon junction diodes is a preliminary one, the data at hand suggest that there should be many uses for these diodes. They will probably fall under the general headings of rectifiers, voltage regulators, switches, and photodiodes.

A Developmental Germanium *P-N-P* Junction Transistor*

R. R. LAW, Senior Member, IRE, C. W. MUELLER, Senior Member, IRE,
J. I. PANKOVE (PANTCHECHNIKOFF), Associate, IRE, AND L. D. ARMSTRONG†

Summary—A developmental germanium *p-n-p* junction transistor that may be readily made in the laboratory by alloying indium into opposite faces of a wafer of single-crystal *n*-type germanium is described. It is shown that this laboratory technique gives experimental transistors with desirable characteristics. Distribution curves of measured characteristics are given for a typical run of 118 units made and tested under similar conditions. Power gains up to 46 db, "alpha" up to 0.997, and noise factor (1 kc) as low as 6 db were achieved.

I. Introduction

THE MASS OF LITERATURE that has so far appeared describing the junction transistor is largely concerned with the theory of its operation and its circuit aspects. Very little has been written that might be of aid to the engineer interested in making experimental quantities of junction transistors in the laboratory. This paper is intended to help fill this gap. A *p-n-p* junction transistor that may readily be made in the laboratory is described; the techniques of making the junctions by alloying indium into opposite faces of a wafer of single-crystal *n*-type germanium are outlined; and a detailed account is given of the performance of transistors made in this manner.

The physics of the junction transistor has been described in detail by Shockley, Sparks, and Teal,[1] and by Shockley.[2] The circuit properties of the junction transistor have been outlined by Wallace and Pietenpol.[3] The use of diffusion in the preparation of *p-n* junctions has been described by Hall and Dunlap,[4] and *p-n-p* junction transistors made by diffusion have been described by Saby.[5] It will be assumed that the reader is familiar with this literature, and the analysis of the principles underlying transistor operation will not be repeated. The present work will begin with a description of a specific developmental germanium *p-n-p* junction transistor that has been found to have useful characteristics.

II. Description

The transistor here described is shown in two stages of assembly in Fig. 1. The "standard size" lead pencil

points to the unit before it is potted. The thin germanium wafer and one of the hemispherical indium-germanium alloy electrodes may be seen. The germanium wafer is mounted on the middle lead of the stem. Connections to the indium-germanium alloy emitter and collector electrodes are made by wires soldered to the alloy and spot-welded to the outer leads of the stem. For protection the completed unit is embedded in a light-proof water-resistant plastic.
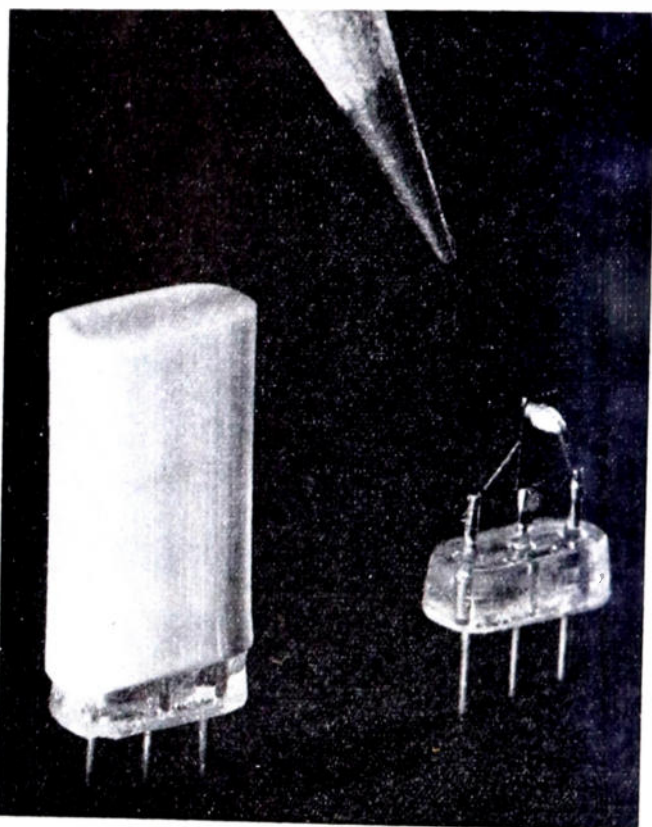
Fig. 1—Two stages in the assembly of a developmental germanium *p-n-p* junction transistor.

The junctions are made by alloying indium into opposite faces of the germanium wafer. The interaction of the indium with the germanium during processing at a temperature above the melting point of indium and below the melting point of germanium can be deduced from the indium-germanium phase diagram[6] shown in Fig. 2. As germanium dissolves in the molten indium, the melting temperature of the solution rises. Initially, the con-

* Decimal classification: R282.12. Original manuscript received by the Institute, August 4, 1952.
† RCA Laboratories Division, RCA, Princeton, N. J.
[1] W. Shockley, M. Sparks, and G. K. Teal, "*P-n* junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; 1951.
[2] W. Shockley, "Electrons and Holes in Semi-conductors," D. Van Nostrand Co., New York, N. Y.; 1950.
[3] R. L. Wallace and W. J. Pietenpol, "Some circuit properties and applications of *n-p-n* transistors," Proc. I.R.E., vol. 39, pp. 753–767; 1951.
[4] R. N. Hall and W. C. Dunlap, "*P-n* junctions prepared by impurity diffusion," *Phys. Rev.*, p. 467; 1950.
[5] J. S. Saby, "Recent developments in transistors and related devices," *Tele-Tech*, vol. 10, p. 32; December, 1951.

[6] W. Klemm, *et al.*, "Des Verhalten der Elemente der III. Gruppe Zueinander, und zu den Elementen der IV. Gruppe. "Zeit. für Anorganische Chemie, vol. 256, pp. 239–251; May, 1948.

entration of germanium in the indium is low and the liquid front penetrates rapidly. Later, as the concentration of germanium increases and the melting point rises,
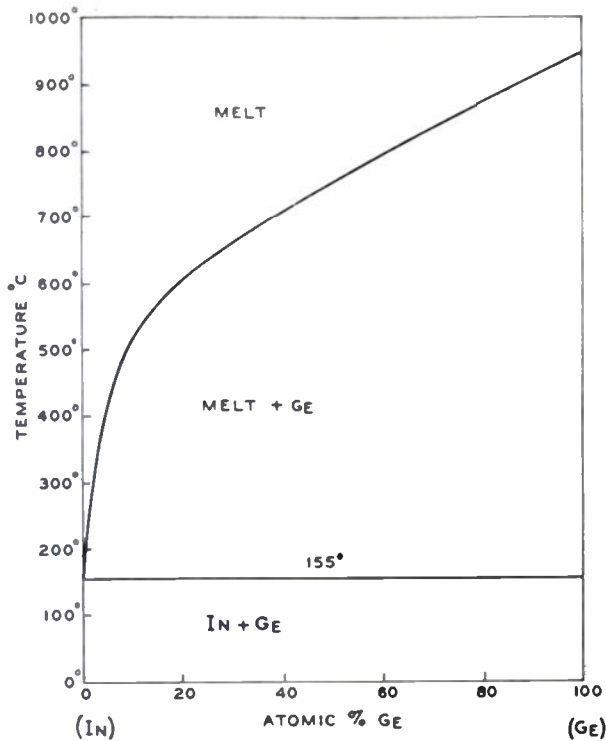
Fig. 2—Phase diagram of the indium-germanium system.

the liquid front penetrates more slowly. The end result is believed to be shown in the cross-section diagram of a transistor as shown in Fig. 3.
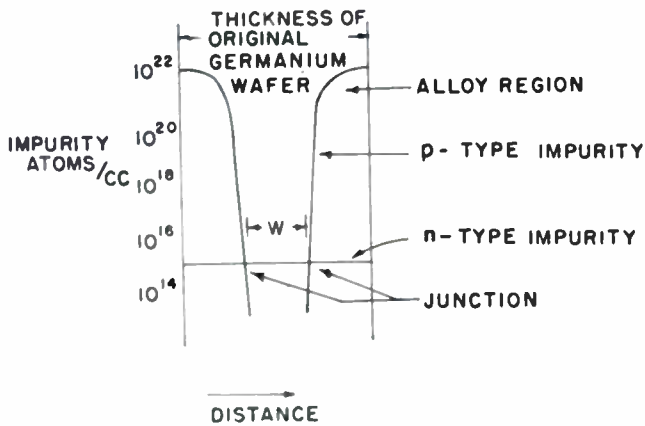
Fig. 3—Variation of impurity-atom concentration in developmental germanium *p-n-p* junction transistor cross section.

The vertical reference lines show the position of the original germanium surface. For clarity, the external portion of the indium is omitted. In the alloy region the concentration of indium is high and varies slowly with distance into the germanium. This alloy contains enough indium so that it behaves as a metal rather than a *p*-type semiconductor. Deeper in, beyond this alloy region, the concentration of indium or *p*-type impurity drops rapidly until it is balanced by the *n*-type impurity present in the original germanium crystal. The junctions occur where the concentrations of *n*-type and *p*-type impurities are equal.

The depth of penetration of the indium is dependent on the relative amount of indium as well as on the time and temperature. Elementary considerations indicate that the longer the time and the higher the temperature, the greater will be the penetration. That the relative amount of indium must also play a part will be evident from the following. If there is much indium, the germanium concentration in the indium melt will remain low and essentially unchanged during the process. The rate of penetration will be essentially characteristic of pure indium dissolving pure germanium at the temperature in question. If there is little indium, the germanium will rapidly saturate the indium and the rate of penetration must rapidly decrease. In general, excessive heating or too much indium causes the indium alloy to completely penetrate the germanium, thus causing an emitter to collector short circuit. Too little heating or too little indium results in widely spaced or inferior junctions and best transistor operation is not obtained. Between these extremes, excellent results are possible.

During the heat treatment some of the indium diffuses and evaporates over the surface of the germanium. This surface layer ordinarily bridges the junctions and would substantially short circuit them if it were not removed as by etching. When this surface layer is properly removed, the junctions exhibit the desired properties. With the mounting technique shown, the forward current is limited chiefly by the ohmic resistance of the base and the leads; the backward current is less than 10 μa at 6 volts. This dc conductance should not be confused with the ac conductance. Because of the change in slope of the current-voltage curve, the ac and dc properties may be quite different. At normal operating voltage the ac impedance may be several megohms in the reverse direction.

III. Results

Much information about the low-frequency properties of a transistor can be obtained from a suitable set of static characteristics. In the past it has been the practice to present static characteristics of the kind shown in Fig. 4. Such characteristics are useful for the point-contact transistor as indicated by the family to the left, but have very little value for the junction transistor as indicated by the family to the right. With a good junction transistor the collector current is very nearly equal to the emitter current and the slope of the curves becomes very small. For this reason the set of static characteristics shown in Fig. 5 is preferred. The reader should be reminded that the less sharp knee in the output characteristics in the new presentation is not
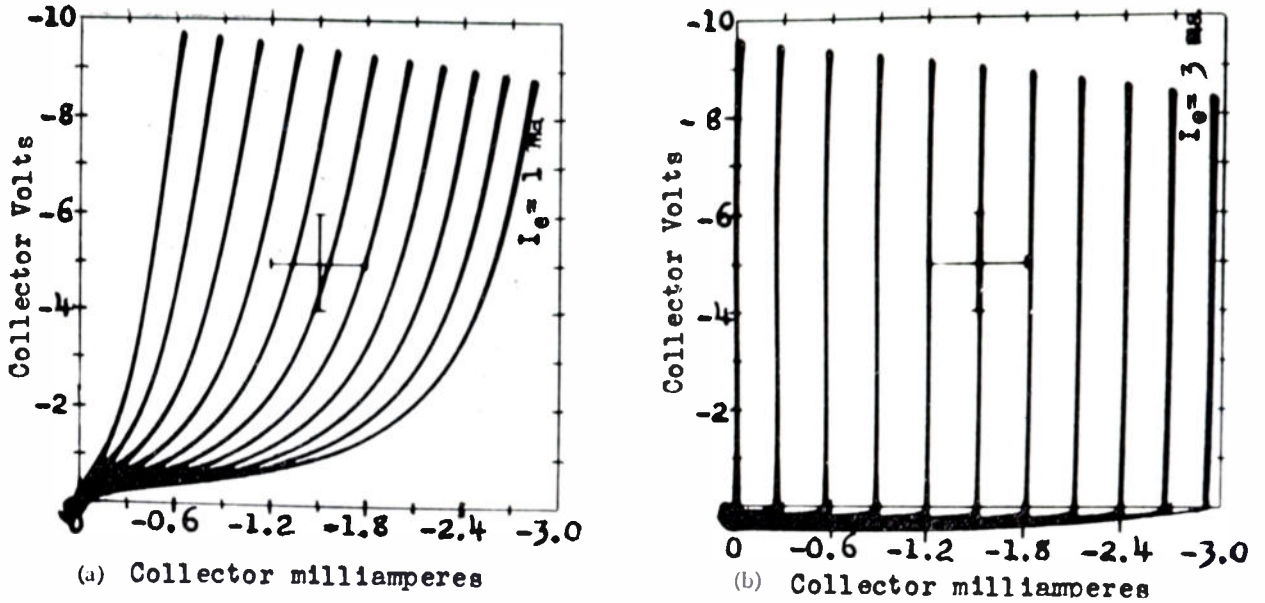
Fig. 4—Static characteristics of a point-contact transistor and a developmental germanium p-n-p junction transistor. I_e is the emitter current. (a) Curves for $I_e = 0$ to $I_e = 1$ ma in 10 equal steps. Characteristics of a typical point-contact transistor. (b) Curves for $I_e = 0$ to $I_e = 3$ ma in 10 equal steps. Characteristics of a typical junction transistor.
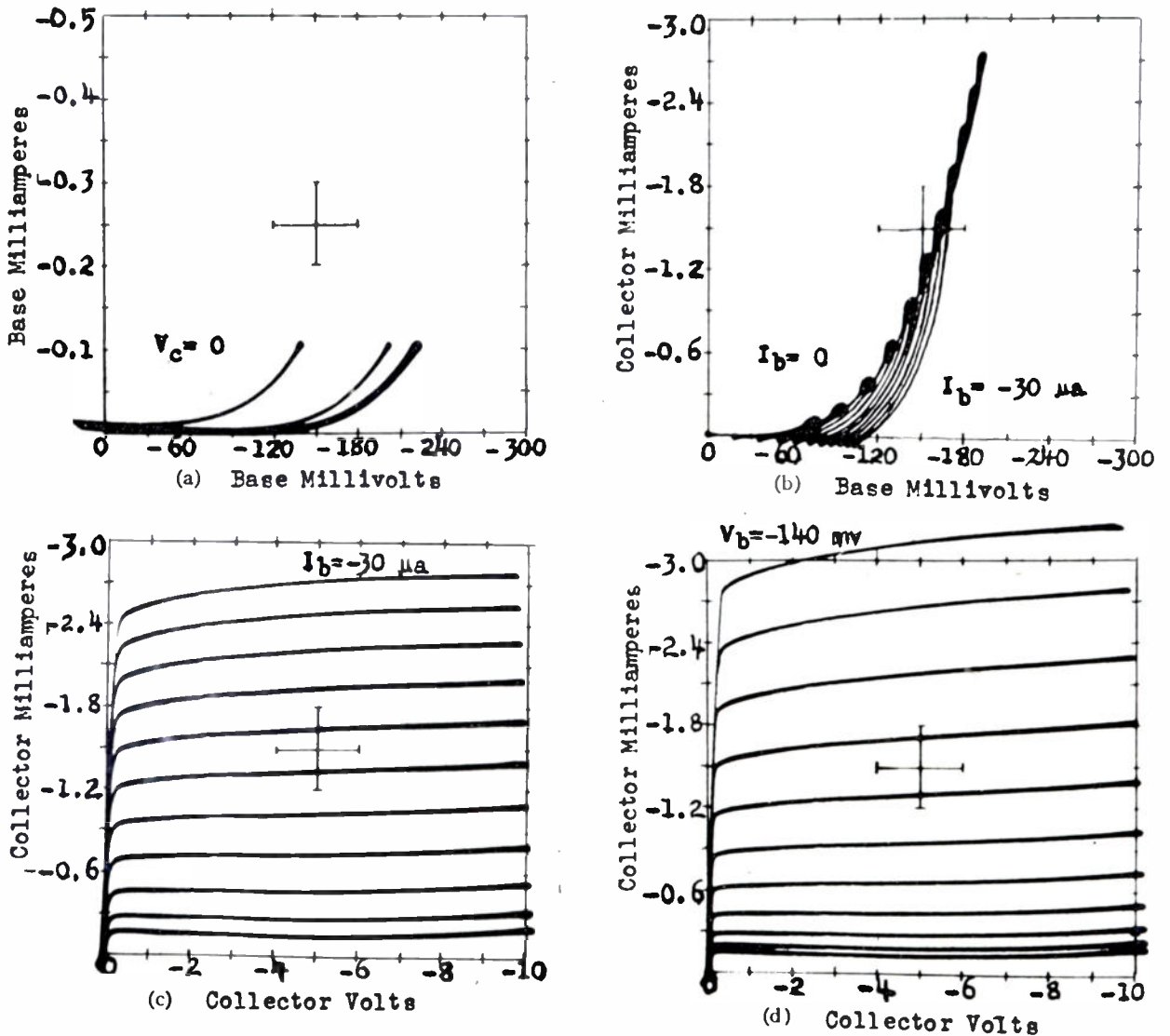
Fig. 5—Characteristics of a developmental germanium p-n-p junction transistor with base input connection. I_b and V_b are base current and base voltage, respectively. (a) Input characteristic. Curves from $V_c = 0$ to $V_c = -10$ volts in 10 equal steps. (Last 8 coincide.) (b) Transfer characteristics. Curves for $I_b = 0$ to $I_b = -30$ μa in 10 equal steps. (c) Output characteristics. Curves for $I_b = 0$ to $I_b = -30$ μa in 10 equal steps. (d) Output characteristics. Curves for $V_b = -40$ to -140 mv in 10 equal steps.

an indication of a poor transistor. Actually, the same junction transistor was used for both these presentations. The difference arises from the better inspection of their properties provided by the new presentation. Curves of this sort can be readily traced out on an oscilloscope and photographed with an automatic curve tracer, as was done for these figures. However, if the small-signal properties are to be determined accurately, it is necessary to measure them directly. The power gain of junction transistors, their circuit parameters and noise factor can be measured by apparatus of the same nature as used for point-contact transistors.

It is beyond the scope of this paper to describe the techniques for making these measurements but, because the properties of the transistor may vary with the operating point, it is appropriate to suggest practical test conditions and indicate the results to be expected.

For many applications the base-input (grounded-emitter) connection is desirable. Not only is the power gain high, but the interstage coupling problem is simplified by the fact that the input impedance more nearly matches the output impedance. For this reason the power gain versus collector voltage data for a typical developmental germanium *p-n-p* junction transistor shown in Fig. 6 is given for this connection. The choice
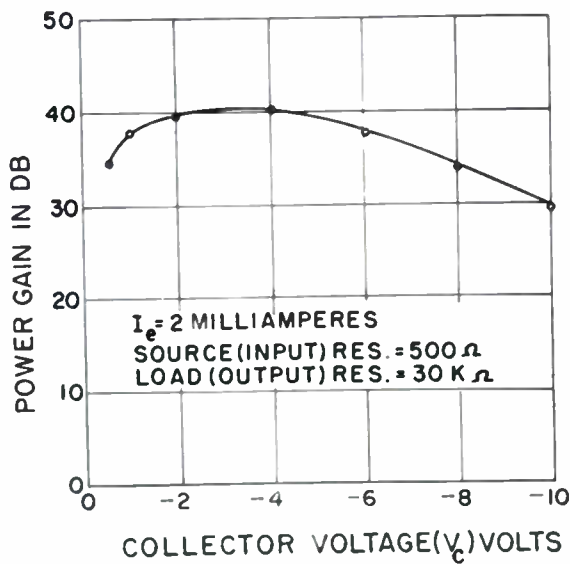
Fig. 6—Variation of power gain with collector voltage for a typical developmental germanium *p-n-p* junction transistor.

of I_e (emitter current) equal to 2.0 ma, a source (input) resistance of 500 ohms, and a load (output) resistance of 30,000 ohms is arbitrary. These values were selected only because they are near the optimum for the present transistor. It will be observed that the power gain has a broad maximum at a collector voltage of about −4 volts.

The collector-to-base current gain,

$$\alpha_{cb} = [-\partial I_c/\partial I_b]_{V_c=\text{constant}}$$

as a function of emitter current for a typical develop-

mental germanium *p-n-p* junction transistor, is shown in Fig. 7. The choice of V_c (collector voltage) equal to −6.0 volts (with respect to base) is also arbitrary. It will be observed that the collector-to-base current gain has a broad maximum at an emitter current of 1 to 3 ma.
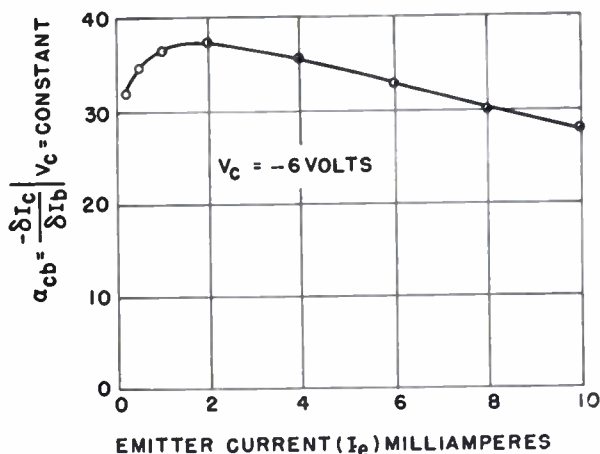
Fig. 7—Variation of α_{cb} (collector-to-base current gain) for a typical developmental germanium *p-n-p* junction transistor.

The variation of noise factor with collector voltage for a typical developmental germanium *p-n-p* junction transistor is shown in Fig. 8. It will be observed that
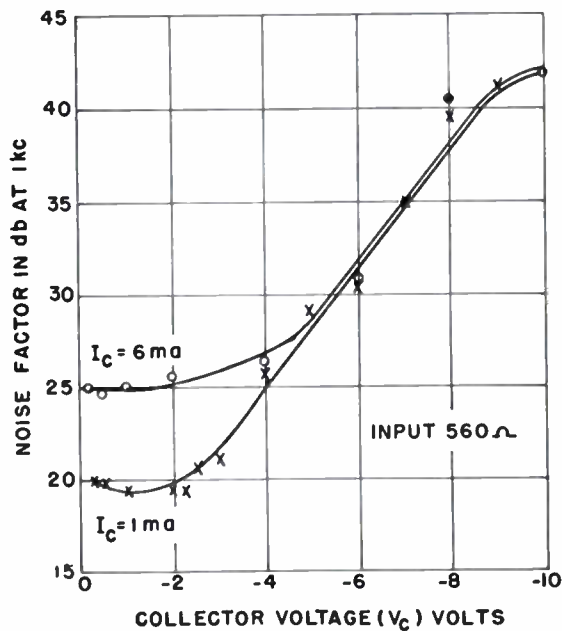
Fig. 8—Variation of noise factor with collector voltage for a typical developmental germanium *p-n-p* junction transistor.

$V_c = -1$ volt near the optimum collector voltage. Fig. 9 shows the variation of noise factor with collector current for the same transistor. It will be observed that $I_e = 1$ ma is near the optimum emitter current.

The present developmental germanium *p-n-p* junction transistor was not designed for high-frequency operation, and no attempt was made to control this parameter. In evaluating the high-frequency performance of

the transistor, it should be remembered that a number of cutoff phenomena[3] influence gain. An important factor which limits high-frequency performance may be

Fig. 9—Variation of noise factor with collector current for a typical developmental germanium *p-n-p* junction transistor.

evaluated by measuring the collector-to-emitter current gain, $\alpha_{ce} = [-\partial I_c/\partial I_e]_{V_c=\text{constant}}$ as a function of frequency. The results of this measurement for several units which indicate the spread are shown in Fig. 10.
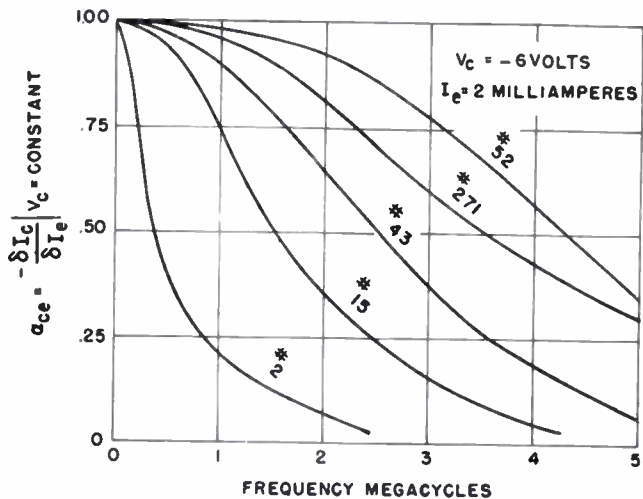
Fig. 10—Variation of α_{ce} with frequency for several developmental germanium *p-n-p* junction transistors.

The variation between units is large and typical values cannot be quoted as is possible for the other parameters so far discussed. It will be observed that some units exhibit usable α_{ce} at 4 mc. Selected units of this kind have been demonstrated to give useful gain over the broadcast frequency band.

It is of interest to see how other parameters of individual transistors made under similar conditions deviate from unit to unit. Although time has not been available to make a detailed study of the large number of laboratory units completed at the time of preparation of this paper, routine measurements have been made of power gain, noise factor at 1 kc, and collector-to-emitter

current gain. The variation of these parameters for a typical lot of 118 units made and tested under similar conditions are indicated in Figs. 11, 12, and 13.
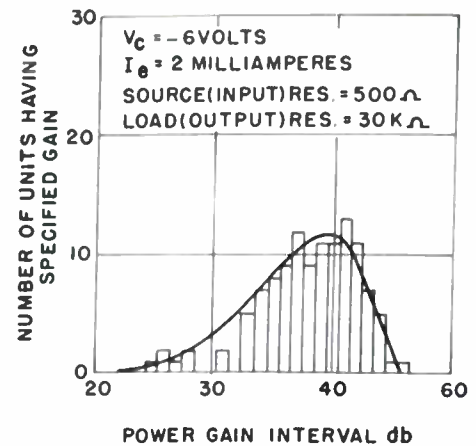
Fig. 11—Distribution of power gain for a typical lot of 118 developmental germanium *p-n-p* junction transistors made and tested under similar conditions.

Fig. 11 shows the distribution of power gain for the lot of 118 units made and tested under similar conditions. The test conditions are $V_c = -6$ volts, $I_e = 2.0$ ma, input resistance 500 ohms, and output resistance 30,000 ohms. This graph shows the number of units having a specified gain in each gain interval. Since the gain interval is 1 db, if gain is expressed to the nearest db,
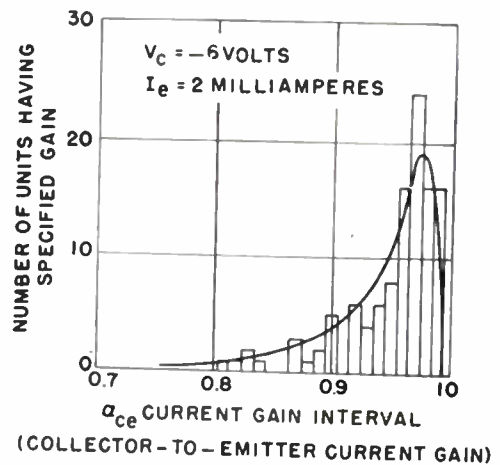
Fig. 12—Distribution of α_{ce} (collector-to-emitter current gain) for a typical lot of 118 developmental germanium *p-n-p* junction transistors made and tested under similar conditions.

the graph may be read directly. Thus, there were 12 units having a gain of 37 db, 9 units having a gain of 38 db, 11 units having a gain of 39 db, and so on. The solid curve smoothing out the irregularities represents the distribution to be expected if a large number of transistors were made and tested in the same manner. The peak of the curve occurs between 39 and 40 db. It will be observed that approximately 50 per cent of the units have a gain within the limits 40 ± 2 db.

Fig. 12 shows the distribution of current gain for the same 118 units. The current gain represented here is the collector-to-emitter current gain

$$\alpha_{ce} = |-\partial I_c/\partial I_e|_{V_c=\text{constant}}.$$

The test conditions are $V_c = -6$ volts and $I_e = 2.0$ ma. As in the preceding case, the height of the bar graph indicates the number of units that have a specified α_{ce} in each α_{ce} interval. Since the α_{ce} interval is 0.01, if α_{ce} is expressed to the nearest second significant figure, it may be said that there are 16 units that have an α_{ce} of 0.98, 24 units that have an α_{ce} of 0.97, etc. As in the preceding case, the curve smoothing out the irregularities represents the distribution to be expected if a large number of transistors were made and tested in the same manner. The peak of the curve occurs for α_{ce} between 0.97 to 0.98. It will be observed that approximately 50 per cent of the units have α's within the limits 0.98 ± 0.02. One unit of this series had an α of 0.997.

noise factor is read to the nearest db, it may be said that there are 9 units that have a noise factor of 20 db, 7 units that have a noise factor of 21 db, 12 units that have a noise factor of 22 db, and so on. As before, the curve smoothing out the irregularities represents the distribution to be expected if a large number of transistors were made and tested in the same manner. The peak of the curve occurs for a noise factor of 22 db. It will be observed that approximately 50 per cent of the units have noise factors within the limits 22 ± 5 db. The best unit in this lot had a noise factor of 6 db.

It is also of interest to see if there is a correlation between noise factor and power gain. Fig. 14 illustrates this relationship. In this diagram each of the 118 trans-

Fig. 13—Distribution of noise factor for a typical lot of 118 developmental germanium *p-n-p* junction transistors made and tested under similar conditions.
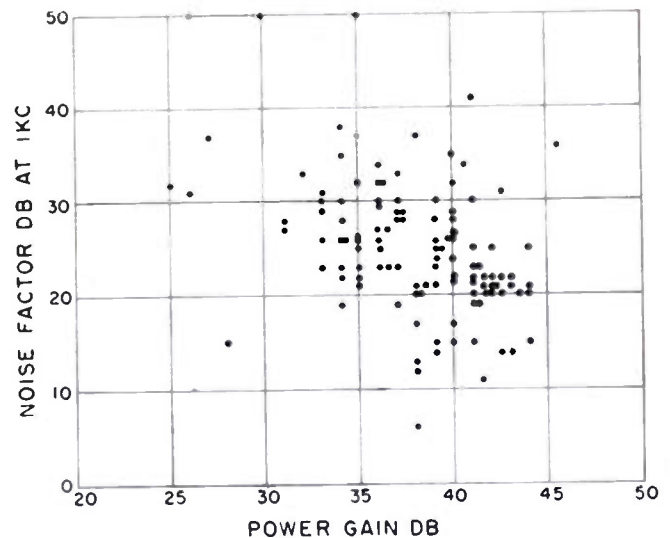
Fig. 14—Correlation of noise factor with power gain for a typical lot of 118 developmental germanium *p-n-p* junction transistors made and tested under similar conditions.

Fig. 13 shows the distribution of noise factor at 1 kc for the same 118 units. The test conditions are $V_c = -1$ volt, $I_e = 1.0$ ma, input resistance 560 ohms, and output resistance 33,000. The choice of different operating conditions for measuring noise and power gain is arbitrary. It is based on the fact that, for small signal operation where noise may be an important factor it may be advantageous to lower the collector voltage. As in the preceding cases, the height of the bar graph indicates the number of units that have a specified noise factor in each 1-db noise factor interval. If

istors of the series is represented by a dot located at the appropriate noise-factor and power-gain co-ordinate. The conditions used for measuring power gain are shown in Fig. 11. The conditions used for measuring noise factor are shown in Fig. 13. Although the correlation is not high, the noise factor evidently decreases as the gain increases.

Fused Impurity *P-N-P* Junction Transistors*

JOHN S. SABY†

Summary—Fused impurity *p-n-p* junction transistors have been made which exhibit the useful circuit characteristics of high-gain, low-noise figure and alpha slightly less than unity. Such transistors are well adapted for application involving high ambient temperatures or high dissipation levels since alpha remains very nearly constant up to 120 degrees C and decreases somewhat above this temperature, permitting their stable use in emitter-grounded circuits at high temperatures. This is in contrast to junction transistors having high-resistivity collector regions, for which alpha often rises above unity at high temperatures leading to circuit instability. Satisfactory operation at several watts collector dissipation has been achieved.

1. INTRODUCTION

AN ALLOY-DIFFUSION PROCESS for the creation of *p-n* junctions in semiconductors has been successfully applied[1] to the fabrication of junction transistors. The *p-n-p* fused-junction transistor consists of a thin wafer of *n*-type germanium in which *p*-type regions are created over controlled areas on the
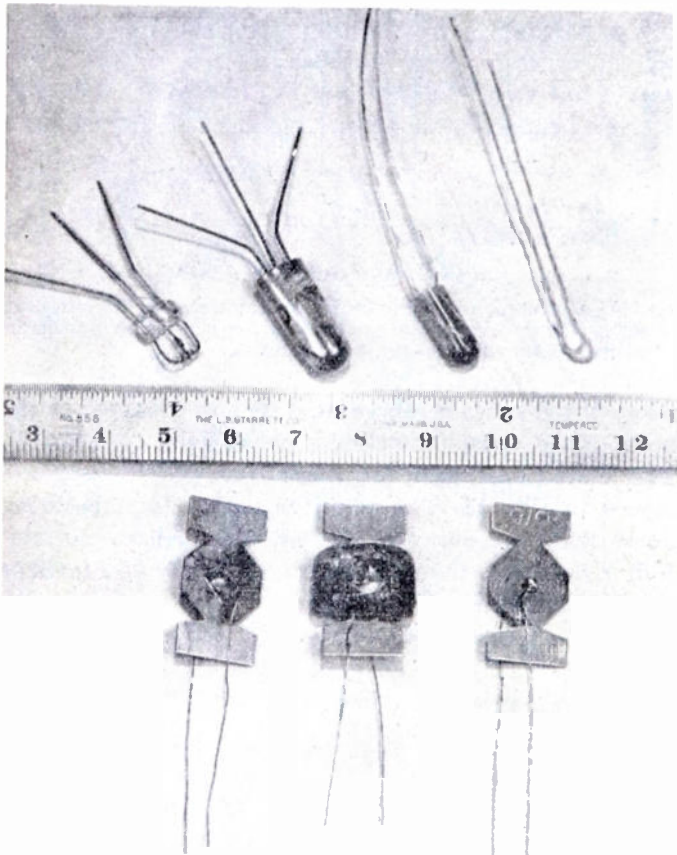
Fig. 1—Developmental forms of fused impurity *p-n-p* junction transistors. The power dissipation of one model (both sides shown) is aided by the metal-strip construction.

[1] Exploratory developmental transistors of this type were previously described by the author at the IRE-AIEE Electron Devices Conference at Durham, N. H., June, 1951.

opposite faces by a process described by Hall and Dunlap.[2] Ohmic contacts are made to the two *p*-regions which serve as emitter and collector, and to the *n*-type base region. Several physical forms of these transistors are shown in Fig. 1.

These *p-n-p* transistors have the general properties originally predicted[3] for this class of transistors. Slightly different boundary conditions apply to these junctions, and the design equations,[4] therefore, differ somewhat from Shockley's. The emitter is normally biased positively, the collector negatively, with respect to the base. Except for the reversed polarities, many of the circuit properties are similar to those of the *n-p-n* transistors previously described.[5,6] In addition, there are certain properties which are characteristic of the nature of the alloy contact junctions. By diffusion of donors into *p*-type germanium, it is also possible to make *n-p-n* transistors, but the data in this article were obtained exclusively with *p-n-p* transistors.

II. ALPHA

Values of alpha $(\partial I_c/\partial I_E)V_c$ less than unity can be obtained from 0.8 to greater than 0.99. The physical factors controlling alpha may be summarized by the relation[4,5,6] $\alpha = \gamma\beta\alpha^*$, where γ, the emitter injection efficiency, is improved by the extremely high acceptor density in the converted emitter region, so that γ can be very close to unity.

The factor β, the fraction of the emitted minority carriers which survive transport through the base region and are collected, is nearly unity (sech W/L_B), provided the base thickness W is much smaller than the normal diffusion L_B for the minority carriers in the base, for a one-dimensional transistor as assumed in the theory. The actual value of β in a transistor with projecting edges will be smaller by the fraction of the injected carriers which diffuse sideways and recombine in these edge regions. Obviously, germanium surfaces with high-recombinations velocity are to be avoided, and the collector junction should extend out as far as necessary to collect as many of these dispersed carriers as possible.

Asymmetrical transistors can be used with emitter and collector interchanged, but characteristics are poorer. However, it is possible to make fused-contact

[2] R. N. Hall and W. C. Dunlap, "*P-N* junctions prepared by impurity diffusion," *Phys. Rev.*, vol. 80, p. 467; 1950.
[3] W. Shockley, "Theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, p. 435; 1949.
[4] E. L. Steele, "Theory of alpha for *P-N-P* diffused junction transistors," *Proc. I.R.E.*, vol. 40, pp. 1424–1429; this issue.
[5] W. Shockley, M. Sparks, and G. K. Teal, "The *p-n* junction transistors," *Phys. Rev.*, vol. 83, p. 151; 1951.
[6] R. L. Wallace and W. J. Pietenpol, "Some circuit properties and applications of *n-p-n* transistors," *Proc. I.R.E.*, vol. 39, p. 753; 1951.

transistors which are symmetrical with good characteristics (alpha greater than 0.98) with either dot used as collector. The functions of emitter and collector in such transistors can thus be interchanged by merely switching bias voltages.

The quantity α^*, the intrinsic current multiplication at the collector, describes the fact that when holes pass from the base region into the collector some conduction electrons from the collector region will travel into the base region. The quantity α^* will be greater than unity by the ratio of the latter current to the former. The "backwards" collection increases with the number of minority carriers present in the collector, and is undesirable in a junction transistor wherever circuit stability requires that α remain less than unity.

The density of minority carriers in the collector region increases with temperature, but can be minimized by the use of very low-resistivity germanium in the collector region. This low-resistivity requirement is automatically met by the diffused impurity junction.

Temperature-alpha data are shown in Fig. 2 for three transistors of different collector resistivity. Due to the presence of other variables, these curves are intended only to indicate the general trend. Fused contact p-n-p transistors have been successfully operated above 140° C in circuits requiring alpha less than unity.

above 130° C, as shown in Fig. 3. Base resistance r_b ranges from 100 to 1,000 ohms, depending upon the resistivity of the germanium and the geometry of the ohmic base contacts. Emitter resistance r_e varies approximately inversely with emitter current. At 1 ma, a distribution of values of r_e from less than 10 ohms up to 50 ohms is observed.

Fig. 3—Collector resistance $(\partial V_c/\partial I_c)I_e$ decreases with increasing temperature. This curve is for a fused impurity p-n-p junction transistor.

IV. NOISE

The noise power per cycle bandwidth varies approximately inversely with frequency. Noise figures at 1 kc (observed at $I_c = 0.5$ ma, $V_c = 1$ volt) lie between 15 db and 30 db for most transistors, while noise figures below 10 db have been observed. The noise figure depends somewhat upon operating point.[7]

V. COLLECTOR CAPACITANCE

The collector capacitance varies approximately inversely as the square root of the dc collector voltage, as may be expected from the nature of the fused-contact p-n junction. Values of capacitance from 5 to 100 $\mu\mu f$ may be encountered. At low frequencies, the capacitance increases with increased current. The capacitance at zero-emitter current depends almost entirely upon the applied voltage and the resistivity of the germanium used for the base material, and is almost independent of the exact nature of the alloy-diffusion region.

Fig. 2—Alpha versus temperature for three transistors. The collector resistivity is known only approximately for the n-p-n transistors. For fused-contact transistors, alpha remains nearly constant up to temperatures above 100°C and decreases at higher temperatures.

III. SMALL-SIGNAL RESISTIVE PARAMETERS

Values of collector resistances, r_c, from several hundred thousand ohms to several megohms are obtained. Collector resistance decreases at high-current levels and at high temperatures, but may be usably high even

[7] E. Keonjian and J. S. Schaffner, An experimental investigation of transistor noise," PROC. I.R.E., vol. 40, pp. 1456–1461; this issue.

This is in contrast to collector capacitance behavior for previously described transistors,[8] for which the zero current capacitance depends principally upon the impurity gradient in the collection junction region and varies approximately as the cube root of the applied voltage.

VI. HIGH-FREQUENCY OPERATION

High-frequency operation depends upon several factors. Collector capacitance affects the output impedance in a manner analogous to the plate capacitance of vacuum tubes. There is a transit time effect of carriers across the base region which renders alpha complex.[8] The quantity $(1-\alpha)$ may thus increase noticeably before α has decreased very much, to the great detriment of grounded emitter circuits which require that $(1-\alpha)$ be small. These effects may become noticeable at audio frequencies.

Finally, the magnitude of alpha itself decreases rapidly with frequency. The 3-db alpha cutoff frequency may range from less than 100 kc to several megacycles, and is controlled principally by the base thickness. These transistors have been used successfully in oscillator circuits above 1.5 mc, and this does not represent a fundamental upper limit.

VII. HIGH-TEMPERATURE OPERATION

The operating temperature of the transistor affects its characteristics in several ways. The minimum collector current, I_{c0} (equal to the "saturation" inverse current for the collector junction), increases exponentially with temperature. At room temperature, I_{c0} may be less than 10 μa, but increases to several milliamperes at 100° C. Increasing the temperature thus results in additional dissipative losses which lower the efficiency of transistors and make low-level operation less attractive. In addition, there is more heat to get rid of, and less of the allowable dissipation can be usefully employed. Finally, the decrease in r_c lowers the power gain, although usable power gain can be obtained above 140° C, particularly if the grounded emitter connection is used, which is possible at these temperatures with fused-junction transistors because alpha remains less than unity.

VIII. HIGH-POWER OPERATION

Since these transistors can safely operate in circuits with collector voltages exceeding 150 volts, in some instances, the use of these transistors at high-dissipation levels is possible provided the high-temperature effects are kept in mind. The temperature difference between the active region of the transistor and the ambient medium must be minimized by proper design, otherwise the transistor will be operating under the handicap of unnecessarily high-junction temperature. In this connection, plastic encapsulation may sometimes cause difficulty.

Also, even though alpha remains less than unity, thermal instability can exist; for if I_{c0} increases, the dissipation likewise may increase, causing a temperature increase and a further increase of I_{c0}, leading to thermal runaway. Thus, some safety margin must be allowed. However, it is possible to increase the safe dissipation, provided suitable current stabilization circuitry is employed.[9]

Fused-contact transistors having active junction area of the order of 1 mm² have been operated at more than 8 watts with forced cooling. Within properly designed cases, these transistors have operated continuously with a good gain at more than 3-watts dissipation in an ambient temperature of 25° C. Dissipations of several hundred milliwatts are possible even in plastic imbedments. Higher-power transistors have been made with other arrangements of the fused junctions.[10]

High-power operation can be achieved with good efficiency by proper circuit design. A transistor megaphone[11] has been constructed using a carbon microphone and two transistors, operating upon self-contained batteries, whose output is one-half watt (electrical), power at an over-all efficiency of approximately 65 per cent. A Class-B push-pull circuit is used which consumes very little standby power with the push-to-talk switch closed in the absence of sound.

IX. CONCLUSIONS

The *p-n-p* fused-junction transistors have characteristics suitable for many applications. The characteristics are well suited for low-level operation. The insensitivity of alpha to temperature changes makes grounded emitter operation feasible above 120° C and permits operation at relatively high-dissipation levels.

ACKNOWLEDGMENTS

[8] R. L. Pritchard, "Frequency variation of current amplification factor in junction transistors," PROC. I.R.E., vol. 40, pp. 1476–1481; this issue.

[9] R. F. Shea, "Transistor operation stabilization: of operating points," PROC. I.R.E., vol. 40, pp. 1435–1437; this issue.
[10] R. N. Hall, "Power rectifiers and transistors," PROC. I.R.E., vol. 40, pp. 1512–1519; this issue.
[11] R. F. Shea, Paper at Conference on Reliable Electron Devices, Washington, D. C., May, 1952.

Four-Terminal *P-N-P-N* Transistors*

J. J. EBERS†, ASSOCIATE, IRE

Summary—The equivalent circuit of a *p-n-p-n* transistor is obtained. It is demonstrated that a *p-n-p* transistor and an *n-p-n* transistor can be connected so that the combination has the same equivalent circuit as the *p-n-p-n* structure. A simplified circuit is obtained which can be used when the *p-n-p-n* transistor is connected as a hook-collector transistor. A method of adjusting the current gain of *p-n-p-n* transistors by external means is given as well as experimental results.

INTRODUCTION

TRANSISTORS having a current multiplication in excess of unity which employ a hook-collector have been previously described by W. Shockley.[1] The hook mechanism has been verified with a transistor having a *p-n-p-n* structure similar in principle to that shown in Fig. 1(a).[2] The *p-n* junction on the left may be used as the emitter, the center junction would be the collector junction, and the junction on the right would be the hook. In a recent article[3] it has been shown that the presence of a hook junction increases the collector current by a factor $1/1 - \alpha_h$ where α_h is the alpha of the *n-p-n* structure on the right of Fig. 1(a). This alpha is equal to the electron-emission efficiency of the hook junction multiplied by the transport factor for electrons across the adjacent *p* region to the collector junction.

Fig. 1—A *p-n-p-n* transistor and its equivalent circuit.

The primary objective of this paper is to show how an equivalent circuit for this *p-n-p-n* structure can be obtained and to indicate how the components of the equiv-

[1] W. Shockley, "Electrons and Holes in Semiconductors," D. van Nostrand Co., Inc., pp. 112–113; 1950.
[2] Experimental data confirming the theory of hook-collector transistors have been obtained by M. Sparks and L. W. Hussey of Bell Telephone Laboratories.
[3] W. Shockley, M. Sparks, and G. K. Teal, "*p-n* junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; 1951.

alent circuit are related to the structure. By means of this circuit it is possible to demonstrate, from a circuit point of view, how the current multiplication is accomplished when the *p-n-p-n* structure is used as a hook-collector transistor. For this case, the equivalent circuit reduces to the one which is ordinarily used for transistors having a current multiplication of approximately one as well as those having a high-current gain.

THE EQUIVALENT CIRCUITS

The junction *p-n-p-n* transistor is shown in Fig. 1(a). This representation is conventional, except in the following:

1. Connections are made to all four regions, and
2. for reasons which will become apparent, both end regions have been termed emitters rather than calling one the emitter and the other the collector.

Briefly, then, the structure consists of a three-junction device having a hole-emitting junction J_{e1}, on the left, an electron-emitting junction, J_{e2}, on the right, and a reverse-biased, collector junction, J_c in the center.

A fraction γ_1 of I_{e1} is hole current which is injected into region 2, where the value of γ_1 is dependent on the conductivities of regions 1 and 2 as well as L_n and W_A. Because of recombination only a fraction β_1 of this hole current reaches the collector junction. Thus the hole current which crosses the collector junction is given by

$$I_1 = \gamma_1\beta_1 I_{e1} = a_1 I_{e1}. \tag{1}$$

Similarly, the electron current which crosses the collector junction is given by

$$I_2 = \gamma_2\beta_2 I_{e2} = a_2 I_{e2}. \tag{2}$$

This leads to the equivalent circuit shown in Fig. 1(b). In this circuit it has been assumed that the bulk resistivity of the semiconductor is negligible. Hence, no base resistances are shown and r_{e1}, r_{e2}, and r_c are dependent only upon the quiescent and saturation currents of the three junctions. First-order equations relating the components of the equivalent circuit to the structure are derived in the Appendix. A more detailed derivation is available in the literature.[4]

Shockley has pointed out[5] that the combination of a *p-n-p* transistor and an *n-p-n* transistor, properly connected, can be made to function in much the same way as a *p-n-p-n* structure. The equivalent circuit for a structure such as shown in Fig. 2(a) can be obtained by drawing the equivalent circuits of the separate struc-

[4] See footnote reference 3.
[5] Personal communication from W. Shockley.

tures and then connecting them together as shown in Fig. 2(b). In this equivalent circuit it is again assumed that the bulk resistivity of the semiconductor is zero.

Fig. 2—The combination of a *p-n-p* transistor and a *n-p-n* transistor.

It is observed that this equivalent circuit is the same as that shown in Fig. 1, except that now the two collector resistances are in parallel, giving an effective collector resistance

$$r_c = \frac{r_{c1}r_{c2}}{r_{c1} + r_{c2}}. \tag{3}$$

A more accurate equivalent circuit for a *p-n-p-n* structure is shown in Fig. 3, where r_A and r_B have been added. No attempt will be made here to relate these resistances to the geometry of the transistor, but their values can be measured for a given transistor.
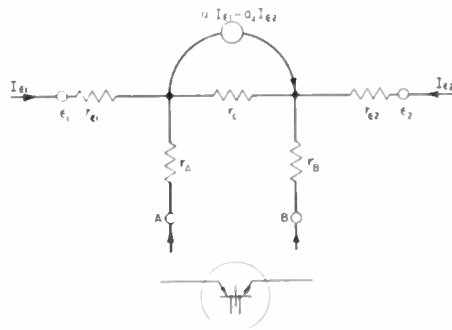
Fig. 3—Equivalent circuit of a *p-n-p-n* transistor.

THE HOOK-COLLECTOR TRANSISTOR

If the *p-n-p-n* transistor of Fig. 1 is used as a grounded-base, hook-collector transistor, a modified equivalent circuit can be obtained with a current generator in parallel with the collector resistance which is not dependent upon the collector current. For this case, terminal ϵ_2 could be used as the collector, terminal A would be the base, and I_B would be equal to zero. The voltage between terminal ϵ_2 and the base is then

$$V_{\epsilon 2} = I_{\epsilon 2}(r_{\epsilon 2} + r_c) + r_c(a_1 I_{\epsilon 1} - a_2 I_{\epsilon 2}), \tag{4}$$

which may be written as

$$V_{\epsilon 2} = \left[(1 - a_2)r_c \frac{a_1}{1 - a_2} \right] I_{\epsilon 1} + [r_{\epsilon 2} + (1 - a_2)r_c]I_{\epsilon 2}. \tag{5}$$

This equation leads to the equivalent circuit shown in Fig. 4, where

$$r_c' = (1 - a_2)r_c. \tag{6}$$

The factor $a_1/1 - a_2$ is approximately the alpha of the compound transistor.[6] A simple explanation of the effect of a hook is as follows: The holes injected across the collector junction by the first emitter produce a voltage
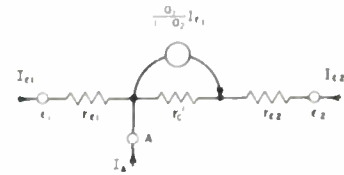
Fig. 4—Equivalent circuit of a hook-collector transistor.

drop across the second emitter junction which causes an electron current to be injected into region (3) by the second emitter, the value of which is dependent upon the γ of the second emitter junction. If the value of α is too large for a particular application, it can be decreased by adding a resistance between terminals B and ϵ_2 which effectively decreases the γ of the second emitter or hook junction. The equivalent circuit for this case is shown in Fig. 5, where

$$r_c'' = \left(1 - \frac{a_2}{1 + r_{\epsilon 2}/R_b} \right) r_c. \tag{7}$$

It is of interest to calculate the reverse collector current of a hook-collector transistor with zero emitter cur-
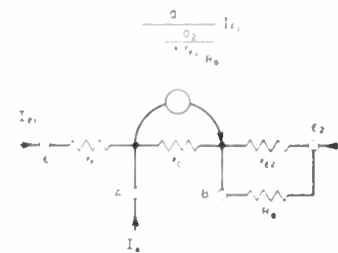
Fig. 5—Equivalent circuit with R_B added to decrease alpha.

rent. Using the circuit of Fig. 1 with $I_{\epsilon 1} = I_B = 0$ and assuming saturation current I_{cs} is flowing in r_c, then

$$I_{\epsilon 2} = I_{cs} + a_2 I_{\epsilon 2} \tag{8}$$

or

$$I_{\epsilon 2} = \frac{I_{cs}}{1 - a_2}. \tag{9}$$

Thus for values of a_2 on the order of 0.99, the saturation current of the collector might be increased by a factor of 100. This assumes that the value of a_2 is large for low emitter currents, which may not be true; however, increases in the saturation current by factors greater than ten have been observed.

[6] In the case of the combination of a *p-n-p* transistor and a *n-p-n* transistor the effective collector resistance is approximately $(1-a_2)$ $(r_{c1}r_{c2}/r_{c1}+r_{c2})$ and the effective base resistance is $(1-a_2)$ $(r_{b1}r_{c2}/r_{c1}+r_{c2})$, which indicates that the circuit is unconditionally stable. This fortunate circumstance is attributable to the fact that the current injected by the hook junction does not flow through r_{b1}. These relations will be derived in a forthcoming article by G. Raisbeck of Bell Telephone Laboratories.

EXPERIMENTAL RESULTS

The data to be presented here were obtained on a p-n-p-n transistor which was actually the combination of an n-p-n transistor having an α of approximately 0.984 and a p-n-p transistor having an α of approximately 0.920. The collector resistances of these transistors were sufficiently large for it to be assumed that the a's of the equivalent circuit are equal to the alpha's. Then, when the emitter junction of the p-n-p transistor is used as the emitter of the composite transistor, the resulting alpha is given by

$$\alpha \simeq \frac{0.920}{1 - 0.984} = 57.5.$$

As shown in Fig. 6 the value of the composite alpha is 71, which would correspond to an alpha for the n-p-n transistor of approximately 0.987. These values of α
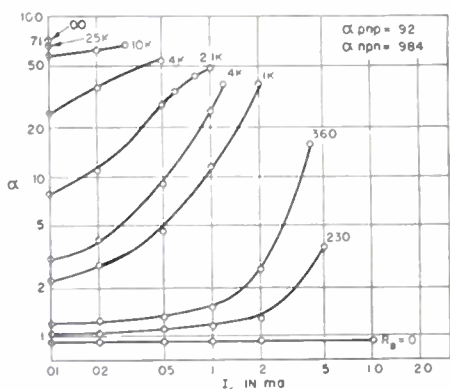
Fig. 6—Variation of alpha with emitter current.

were measured at different operating points, and therefore the discrepancy is not considered serious. Also shown in Fig. 6 are curves of alpha as a function of emitter current for various values of R_B connected as shown in Fig. 5. It is observed that as I_e increases the value of alpha increases slowly at first and then quite rapidly.
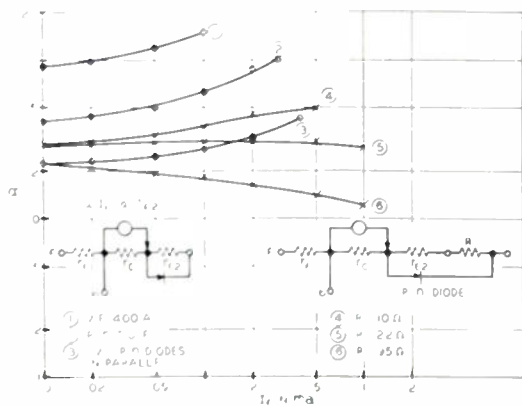
Fig. 7—Variation of alpha with emitter current.

This is attributable to the fact that I_{e2} increases rapidly with I_e and the value of r_{e2} decreases according to (A-16) of the Appendix.

Using circuits such as shown in Fig. 7 it is possible to obtain almost any α versus I_e characteristic that is de-

sired. As is shown, the addition of a diode in parallel with r_{e2} results in an almost flat characteristic; whereas curves in which alpha decreases with I_e can be obtained by also connecting a resistance in series with r_{e2}, as shown in the figure. Other nonlinear resistances such as thermistors could also be used.

The variation of alpha with R_B for a fixed emitter current is shown in Fig. 8. The value of I_{c0} (collector current with zero emitter current) as a function of R_B is also plotted in Fig. 8. An increase by a factor of 20 is observed for large values of R_B, which agrees semi-quantitatively with (9).
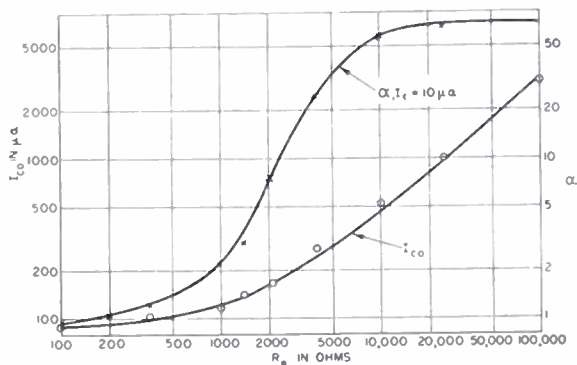
Fig. 8—Variation of alpha and collector current (zero emitter current) with R_B.

CONCLUSIONS

The fact that the combination of a p-n-p transistor and an n-p-n transistor can be used in many applications where a p-n-p-n structure is desired has been demonstrated. The dissimilarity in the way that current flows in the various regions of the two structures may well have a large effect on such factors as frequency response and stability.

The frequency response of hook-collector transistors is limited by several factors, the most important being the same effect that has been observed in the grounded emitter n-p-n of the p-n-p transistor amplifier. It has been pointed out that the factor $1/1-a$ decreases much more rapidly than would be expected because of a simple decrease in the magnitude of a since a has a negative phase angle associated with it that results from transit-time effects.[7] This becomes important when the magnitude of a is nearly unity. In addition, it is apparent that the response of a grounded-base, hook-collector amplifier stage would be the product of a grounded-base stage and a grounded-emitter stage. Because of the symmetry of both the p-n-p-n transistor and its equivalent circuit, it has been suggested that this device be called a conjugate-emitter transistor. This name becomes particularly appropriate when the device is used as a four-terminal transistor in applications other than a high-current gain amplifier. It has also been suggested that the symbol shown in Fig. 3 be adopted.

[7] D. E. Thomas, "Transistor amplifier—cutoff frequency," PROC. I.R.E., vol. 40, pp. 1481–1483; this issue.

Acknowledgment

Appendix

It is of interest to relate the components of the equivalent circuit of Fig. 1(b) to the geometry of the p-n-p-n structure. For this reason, the regions are numbered and subscripts on subsequent symbols will refer to these regions. The assumptions made and the symbols used are the same as in the Shockley, Sparks, and Teal article, except that, in addition, it is assumed that the bulk resistivity of the semiconductor is negligible in comparison to the junction resistances. It is also assumed that the semiconductor filament has a cross-sectional area A. To a first approximation, the current I_{e1} is given by

$$I_{e1} = \frac{\sigma_i^2 b A}{(1+b)^2}\left[\frac{1}{\sigma_2 W_A} + \frac{1}{\sigma_1 L_{n1}}\right]$$
$$\cdot\left[\exp\frac{q}{kT}(V_{e1} - V_A) - 1\right] \quad (A.1)$$

or

$$I_{e1} = I_{e1s}\left[\exp\frac{q}{kT}(V_{e1} - V_A) - 1\right], \quad (A.2)$$

where it is assumed that $W_A \ll L_{p2}$ and that the length of region (1) is much greater than L_{n1}. The first term on the right of (A.1) represents the hole current into region (2) and the second term represents the electron current into region (1). Usually the definition is made that

$$\gamma_1 = \frac{\text{hole current of emitter 1}}{\text{total current of emitter 1}} \quad (A.3)$$

or that, after substitution of the values from (A.1),

$$\gamma_1 = \frac{1}{1 + \dfrac{\sigma_2 W_A}{\sigma_1 L_{n1}}}. \quad (A.4)$$

For a good emitter, γ should be nearly unity.

Because of recombination, not all of the holes emitted by J_{e1} reach the collector junction. The fraction which reaches the collector junction can be shown to be given by

$$\beta_1 = \text{sech}\frac{W_A}{L_{p2}} \quad (A.5)$$

to a first approximation. The hole current which flows across the collector junction is

$$I_1 = \gamma_1\beta_1 I_{e1} = a_1 I_{e1}. \quad (A.6)$$

Similarly, for the right-hand emitter junction

$$I_{e2} = -\frac{\sigma_i^2 b A}{(1+b)^2}\left[\frac{1}{\sigma_3 W_B} + \frac{1}{\sigma_4 L_{n4}}\right]$$
$$\cdot\left[\exp\frac{q}{kT}(V_B - V_{e2}) - 1\right] \quad (A.7)$$

or

$$I_{e2} = -I_{e2s}\left[\exp\frac{q}{kT}(V_B - V_{e2}) - 1\right]. \quad (A.8)$$

As before, the definition is made that

$$\gamma_2 = \frac{\text{electron current of emitter 2}}{\text{total current of emitter 2}} \quad (A.9)$$

or that, after substitution of the values from (A.7),

$$\gamma_2 = \frac{1}{1 + \dfrac{\sigma_3 W_B}{\sigma_4 L_{p4}}}. \quad (A.10)$$

The fraction of the emitted electrons which reach the collector junction is given by

$$\beta_2 = \text{sech}\frac{W_B}{L_{n3}}, \quad (A.11)$$

and the electron current which flows across the collector junction is

$$I_2 = \gamma_2\beta_2 I_{e2}. \quad (A.12)$$

The current which flows in the collector junction due to the voltage across it is given by

$$I_c = \frac{\sigma_i^2 b A}{(1+b)^2}\left[\frac{1}{\sigma_2 W_A} + \frac{1}{\sigma_3 W_B}\right]$$
$$\cdot\left[\exp\frac{q}{kT}(V_B - V_A) - 1\right], \quad (A.13)$$

or

$$I_c = I_{cs}\left[\exp\frac{q}{kT}(V_B - V_A) - 1\right]. \quad (A.14)$$

The ac impedances r_{e1}, r_{e2}, and r_c can be obtained by taking the derivatives of (A.2), (A.8), and (A.14), respectively. They are

$$r_{e1} = \frac{kT}{q}\frac{1}{(I_{e1} + I_{e1s})}, \quad (A.15)$$

$$r_{e2} = \frac{kT}{q}\frac{1}{(I_{e2} + I_{e2s})}, \quad (A.16)$$

and

$$r_c = \frac{kT}{q}\frac{1}{(I_c + I_{cs})}, \quad (A.17)$$

where I_{e1}, I_{e2}, and I_c in the above three equations are the quiescent currents. For this first-order theory, the ac values of a_1 and a_2 are the same as the dc values.

A Unipolar "Field-Effect" Transistor*

W. SHOCKLEY†

Summary—The theory for a new form of transistor is presented. This transistor is of the "field-effect" type in which the conductivity of a layer of semiconductor is modulated by a transverse electric field.

Since the amplifying action involves currents carried predominantly by one kind of carrier, the name "unipolar" is proposed to distinguish these transistors from point-contact and junction types, which are "bipolar" in this sense.

Regarded as an analog for a vacuum-tube triode, the unipolar field-effect transistor may have a mμ of 10 or more, high output resistance, and a frequency response higher than bipolar transistors of comparable dimensions.

1. Unipolar Transistor and Field-Effect Modulation

PRIOR TO THE INVENTION by Bardeen and Brattain of the point-contact transistor at Bell Telephone Laboratories, the hope that purely electronic amplification might be obtained in semiconductors was strongly bolstered by the concept of the "field-effect" amplifier.[1] The field-effect transistor was conceived as consisting of a thin layer of semiconductor separated from another conducting electrode by a layer of dielectric. A potential applied across the dielectric would then charge the layer of semiconductor, and this change in charge would alter the number of holes and electrons available for conduction, thus altering the conductivity of the layer.

This effect was found to take place in experiments carried out by Pearson.[2] It was much less than that expected on the basis of an elementary theory, and the difference was attributed to surface states. These surface states were assumed, in a theory proposed by Bardeen,[3] to trap, in an immobile condition, the added charge induced on the layer of semiconductor.

It is the purpose of this article to show that transistor structures are possible in which the adverse effects of surface states are eliminated and to show how, in certain ways, these structures have advantages over previously described transistors.

One of the important characteristics which distinguishes the "field-effect" transistors described here from the more conventional types is that the working current is substantially "unipolar." In the point-contact, filamentary, and junction types an important process is the injection of minority carriers into regions having relatively high concentrations of majority carriers. In the region of injection, space-charge neutrality is maintained to a high degree by currents of majority carriers

which neutralize the space charge of the minority carriers. Since carriers of both signs are involved, these processes may be referred to as "bipolar." In a "field-effect" transistor, the current flow is carried by one type of carrier only. The changed conductance between input and output terminals results from changing numbers of carriers of this one type. For this reason the name "unipolar transistors" is proposed.

One feature which should be emphasized about field-effect transistors is the possibility of voltage gain in addition to current gain. We shall show that voltage gain is indeed possible and that mμ values of ten or higher can be obtained.

Another feature of importance is the efficiency for high-frequency response of these devices compared to bipolar types of comparable dimensions. The reason for this difference is that in the bipolar types, the electric field is either very low, as in the base layer of a junction transistor, or else is effective in producing a current of majority carriers, as in the filamentary and point-contact transistors. In the latter cases, the majority carrier current is an important source of power dissipation being produced by the electric field that gives the minority carriers a desirably high drift velocity. In the unipolar types, even higher drift velocities can be produced without similar disadvantages.

It should be remarked in passing that the unipolar transistors considered here are essentially different from devices of the "fieldistor" type such as that discussed by Stuetzer.[4] The theory proposed by Stuetzer does involve modulation by electric fields, but otherwise bears, at most, a distant relationship to the structures considered here.

2. Modulation of a Conducting Channel by Electric Fields

In Fig. 1 we represent an example of a unipolar field-effect transistor. It is a three-terminal device, and consists principally of a layer of p-type material sandwiched between two layers of n-type material, the doping being much stronger in the n-type material, which is designated as $n+$. The "working current" is carried by holes flowing in the x-direction in the p-layer. The terminals are inserts of heavily doped p-type material, designated $p+$. Under operating conditions, reverse bias is applied across the p-n junctions, and space-charge regions form in which the carrier concentration is negligible. As a consequence, the current flows in a *channel* of p-type material bounded by the space-charge regions.

If the reverse bias at one $p+$ terminal is larger than at the other, the channel will be narrower at the one

* Decimal classification: R282.12. Original manuscript received by the Institute, August 18, 1952.
† Bell Telephone Laboratories, Inc., Murray Hill, N. J.
[1] For a brief history see W. Shockley "Electrons and Holes in Semiconductors," D. van Nostrand, New York, N. Y., chap. 2; 1950.
[2] W. Shockley and G. L. Pearson, "Modulation of conductance of thin films of semiconductors by surface charges," *Phys. Rev.*, vol. 74, pp. 232–233; 1948.
[3] J. Bardeen, *Phys. Rev.*, "Surface states and rectification at metal semiconductor contact" vol. 71, pp. 717–727; 1947.

[4] O. T. Stuetzer, "A crystal amplifier with high input impedance," Proc. I.R.E., vol. 38, pp. 868–871; August, 1950.

terminal. How the channel varies in width between the terminals will be discussed in subsequent sections.

In passing, we shall point out that the unipolar transistor of Fig. 1 is in some ways a closer analog to a vacuum tube than are the bipolar transistors. If we imagine that a signal is applied between terminals 3 and ground, then the effect will be to widen and narrow the channel which carries current between 1 and 2. This is closely analogous to the action of a grid which controls the current flow in a "channel" of thermionic electrons flowing between the grid wires. As in the bipolar case, the control is exerted by flow of electrons by one mechanism on flow by another mechanism; in this case, flow by excess electrons in the $n+$ regions controls flow of holes in the channel. A difference from the bipolar case is that the controlling and controlled currents differ not only in conduction mechanism but also in spatial location.

In this section we shall consider only the case in which the reverse bias is almost the same on the two terminals so that the channel has substantially uniform width. We shall derive the dependence of conductance of the channel upon

$$W = \text{the magnitude of the reverse bias} \qquad (1)$$

applied across the p-n junctions.

In Fig. 1, the space-charge region is represented as lying entirely in the p-region. This is only an approximation; but it is a good one, as we shall show below if the $(n+)$-regions are much more heavily doped than the p-region. The doping of the regions defines two "chemical" charge densities

$$p\text{-region} \qquad \rho_0 = q(N_a - N_d) \qquad (2a)$$

$$(n+)\text{-region} \qquad \rho_n = q(N_d - N_a), \qquad (2b)$$
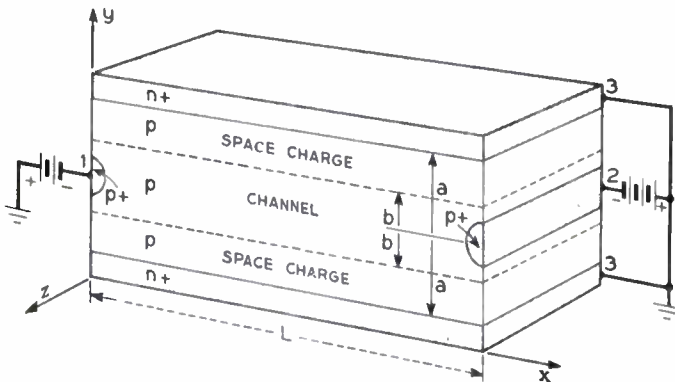
Fig. 1—Space-charge region and channel in a $(n+)p(n+)$ structure.

where q is the electronic charge, N_d and N_a densities of donors and acceptors. The charge density of the holes in the conducting part of the p-region is equal to ρ_0 and the conductivity is $\sigma_0 = \mu_0 \rho_0$, where μ_0 is the mobility of a hole. The choice of signs in (2a) and (2b) makes both ρ_0 and ρ_n positive so that ρ_n corresponds to the chemical charge density in the n-region while ρ_0 corresponds to the carrier charge density in the p-region.

The distribution of charge density in the space-charge region is represented in Fig. 2(a). The middle plane of the p-region is $y = 0$, the edge of the space-charge layer is at $y = b$, and the p-n junction at $y = a$. The

space-charge region is a dipole layer in which the charges add to zero. (If this were not the case, Gauss' theorem would require an electric field to exist beyond the boundaries of the layer.) Consequently, the shaded rectangles of Fig. 2(a) are equal in size and the depth of penetration is much less in the n-region.
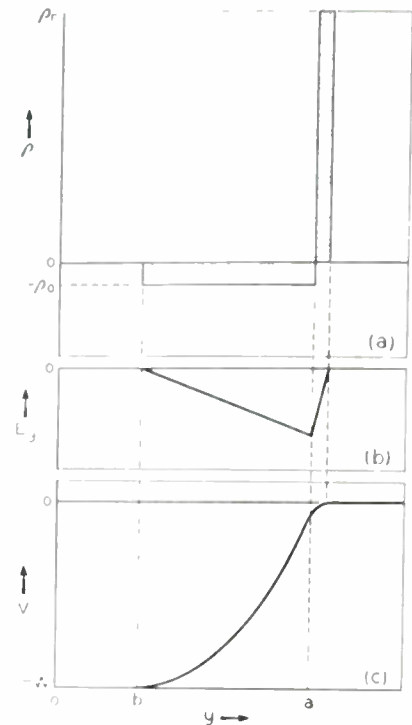
Fig. 2—Penetration of the space-charge region into the p and $n+$ regions. (a) Space charge. (b) Electric field. (c) Electrostatic potential.

Mathematically, this result may be deduced from Poisson's equation. If we let $V(y)$ represent the electrostatic potential, $E_y(y)$ the electric field, and y distance from the center of the p-layer, then Poisson's equation is

$$\kappa\epsilon_0 d^2V/dy^2 = -\kappa\epsilon_0 dE_y/dy = -\rho(y), \qquad (3)$$

where $\rho(y)$ is the charge density. For y less than b, E_y is zero. Near $y = b$ there is a transition region in which ρ changes from zero to $-\rho_0$. This transition region is about one "Debye Length" thick, and in it the potential changes by about kT/q, where k is Boltzmann's constant and T the absolute temperature.[5] These quantities are both very small compared to a and W, respectively, and may be neglected. Consequently, we assume that at the edge of the space-charge layer we have

$$E_y = 0 \quad \text{at} \quad y = b, \qquad (4)$$

and within the p region

$$dE_y/dy = \rho/\kappa\epsilon_0 = -\rho_0/\kappa\epsilon_0 \qquad y > b, \qquad (5)$$

so that

$$E_y = -\rho_0(y - b)/\kappa\epsilon_0. \qquad (6)$$

This leads to the variation of E_y shown in Fig. 2(b) for $y < a$. For $y > a$, the charge density changes from a small negative to a large positive value, ρ_n. Evidently, E_y will

[5] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, p. 441; 1949.

vanish when $(a-b)\rho_0 = \rho_n \times$ (thickness of space-charge layer in n-region). This equation requires the equating of areas shown in Fig. 2(a).

Since the average value of the electric field is the same in both regions, as may be seen by inspection of Fig. 2(b), the potential differences across the two regions will be proportional to the thickness of the space-charge layers. We shall neglect the thickness in the n-region and assume that the potential drop occurs wholly in the p-region. This leads to the value

$$V = -\int E_y dy$$
$$= (\rho_0/2K)[(y-b)^2 - (a-b)^2], \tag{7}$$

where the constant of integration has been chosen to make $V=0$ when $y=a$, corresponding to grounding the n-region, and we have introduced

$$K = \kappa\epsilon_0 \tag{8}$$

as the dielectric constant.

K may be conveniently expressed in farads/cm rather than farads/m. Since we shall make no use of magnetic quantities in this article, it is not necessary to have a magnetically consistent set of equations. The use of K in farads/cm permits the use of dimensions in cm, mobilities in cm²/volt sec, and conductivities in ohm/cm, while retaining currents and voltages in amp and volt. (These conclusions may be verified for any equations given here simply by deriving them in mks and then converting the length unit to cm.)

With $\kappa = 16$ and 12 for germanium and silicon, respectively, the values of K are

$$\text{Ge} \quad K = 1.42 \times 10^{-12} \text{ farads/cm} \tag{9}$$
$$\text{Si} \quad K = 1.06 \times 10^{-12} \text{ farads/cm}. \tag{10}$$

The potential in the channel is

$$V(b) = -(\rho_0/2K)(a-b)^2. \tag{11}$$

We shall find it advantageous to use W, a positive quantity representing this reverse bias,

$$W = -V(b) = [1 - (b/a)]^2 W_0, \tag{12}$$

where

$$W_0 \equiv \rho_0 a^2/2K. \tag{13}$$

W_0 is the magnitude of reverse bias required to make the space charge penetrate the entire p-region. We shall refer to it as the "pinch-off voltage" since it is the voltage that will reduce the channel to zero and pinch off the conducting path between terminals 1 and 2.

In Fig. 3, the potential distribution in the p-region is shown and related to the values of W and W_0. The dashed line shows the limiting case of $W = +W_0$. The solid line shows a case in which $W = W_0/4$.

We shall next apply these considerations to the treatment of the transistor structure represented in Fig. 1. We shall denote the reverse biases at terminals 1 and 2 by W and $W+\Delta W$, respectively. These biases cause a current I per unit length in the z-direction to flow from terminal 1 to terminal 2. We shall require ΔW to be

small compared to $W_0 - W$ so that the channel will be of substantially uniform width and can be treated simply as a layer of p-type semiconductor of conductivity

$$\sigma_0 = \mu_0 \rho_0, \tag{14}$$

where μ_0 is the mobility of holes. If the electric field in the layer is E_x, the current density will be $\sigma_0 E_x$ and the current per unit length in the z-direction will be

$$I = 2b\sigma_0 E_x = g(W)E_x, \tag{15}$$

where $g(W)$ is the conductance of a unit square of the layer $2b$ thick.
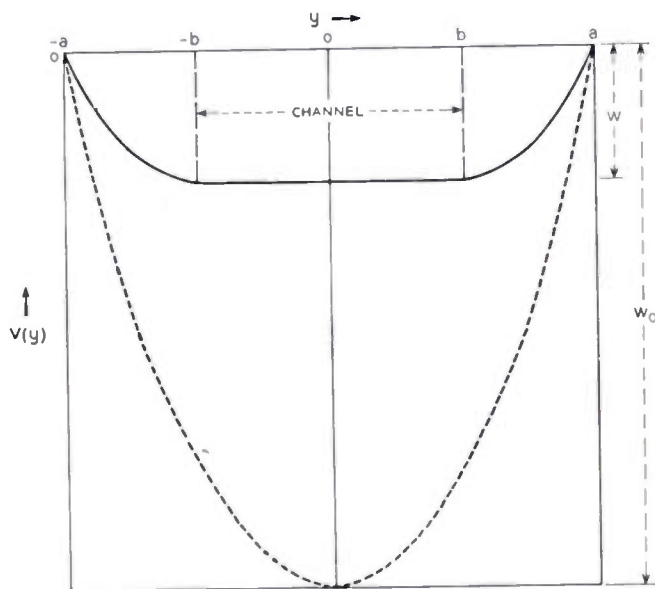
Fig. 3—Potential distribution for channel with $W = W_0/4$ (and $b = a/2$) and limiting case of $W = W_0$.

For the case in which ΔW is small compared to W and to $W_0 - W$, the channel will be of uniform width and the conductance per cm in the z-direction of a structure of length L in the x-direction will then be

$$G = g(W)/L. \tag{16}$$

However, if ΔW is not small compared to $W_0 - W$, then the channel will change its width appreciably, and this must be taken into account. We do this by considering the dependence of g upon W.

Equation (12) gives a relationship between W and b. This can be transformed into a relationship between g and W;

$$g(W) = 2\sigma_0 b(W) = [1 - (W/W_0)^{1/2}]g_0, \tag{17}$$

where

$$g_0 \equiv 2\sigma_0 a \tag{18}$$

is the conductance of a square of the p-region at zero bias. In Fig. 4, the dependence of $g(W)$ upon W is shown. In terms of this dependence it is possible to calculate the current flowing in the channel when there is a large voltage difference between the ends of the channel. This calculation is presented in Section 4.

3. Proposed Unipolar Transistor Terminology

As was stressed in the introductory section, the principles of operation of the unipolar transistor are sub-

stantially different from those of the bipolar types. For this reason, it seems appropriate to consider choosing a new set of names for the three terminals. An advantageous choice would be one which involves no overlapping of subscripts with the emitter, base, collector combination in use for bipolar transistors. The choice selected is "source" for the electrode through which the carriers flow into the channel, "drain" for the electrode into which the carriers flow out of the channel, and "gate" for the control electrodes that modulate the channel. One reason for selecting "gate" rather than something else is that the subscript "g" is reminiscent of "grid" and the analogy is close between the two. Voltages applied to and currents flowing into these electrodes are denoted by

$$V_s, V_g, V_d \quad \text{and} \quad I_s, I_g, I_d \tag{19}$$

for dc values, small letters being used for small-signal or ac values.
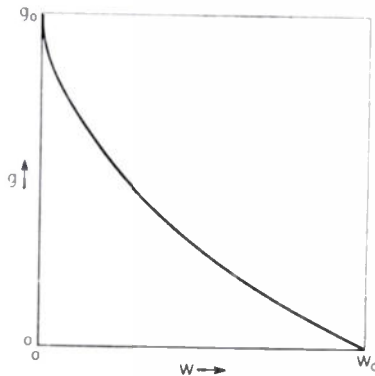
Fig. 4—Conductance of square of channel versus W.

A symbol for this type of transistor is shown in Fig. 5. As with the bipolar transistor, the arrow indicates the *direction of conventional current* flow, and is *located near source* where working carriers enter the structure.
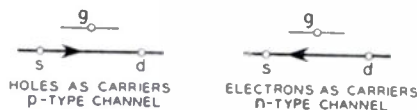
Fig. 5—Symbols for unipolar field-effect transistors.

When it is convenient to use numerical subscripts to indicate the three terminals, system will be that shown in Fig. 1, with 1 for source, 2 for drain, and 3 for gate.

In the treatment of this article, we shall, in general, deal with low frequencies and shall neglect the current to the gate both due to reverse currents and to charging currents and shall set I_g equal to zero. Consequently, I_s and I_d are equal and opposite. We shall use I to mean

$$I = I_s = - I_d, \tag{20}$$

these quantities being *currents per unit length* of structure in the z-direction.

The relationship between the V's and the W's is evidently

$$W_s = V_g - V_s \tag{21}$$

$$W_d = V_g - V_d, \tag{22}$$

and small changes in these quantities, which we shall treat as differentials, are

$$\delta W_s = v_g - v_s \tag{23}$$

$$\delta W_d = v_g - v_d \tag{24}$$

$$\delta I = i = i_s = - i_d. \tag{25}$$

We shall use these relationships in the next section in deriving an equivalent circuit for the unipolar transistor.

4. The Gradual Case

A. Introduction

When current flows in the plus x direction in the channel, an electric field with a component E_x must be present. This requires that the potential changes along the channel. Since the gate electrodes carry no current (we shall consider small reverse junction current later), they are equipotentials. Hence the reverse voltage between channel and $(n+)$-region varies with x and hence the channel width varies. In calculating the relationship between channel width b and reverse voltage W in Section 2, we assumed that

$$\partial^2 V / \partial x^2 = 0, \tag{26}$$

so that a one-dimensional Poisson equation could be used. Since when current flows,

$$E_x = - \partial V / \partial x \neq 0, \tag{27}$$

in general $\partial^2 V / \partial x^2$ will not vanish. However, if $\partial^2 V / \partial x^2$ is very small compared to ρ_0 / K, then the one-dimensional approximation can be used for $V(y)$ and the reverse potential $W(x)$ will be the same function of $b(x)$ as in the case of $I = 0$. The approximation that W and b are related in this way is valid if conditions along the channel vary gradually. We refer to this situation as the *gradual case*.
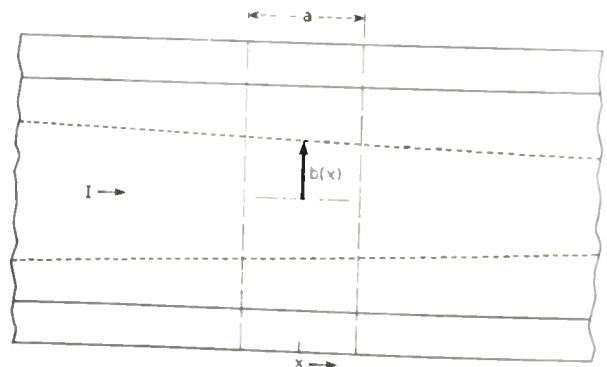
Fig. 6—Condition in a "gradual" channel.

Note: Interval marked a should be slightly wider than shown, so as to be one half the width of the p layer.

The situation that prevails in the gradual case is illustrated in Fig. 6. Current flows to the right so that in the $+x$ direction the potential becomes more negative, the reverse bias increases, and the channel narrows. Consider a particular value of x and a length a centered about x. In this length $b(x)$ is supposed to vary by a small fraction of itself. Now the potential difference be-

tween the points $(x, 0)$ and (x, a) will depend on the entire distribution of charge in the structure and the potentials applied to source and drain. However, the $(n+)$-regions exert a shielding action so that the potential at $(x, 0)$ is determined chiefly by charges lying within the interval a about $(x, 0)$, unless, of course, there are very much larger charge densities relatively near. If the channel changes relatively gradually, the potential will thus be that determined by the average value of b in the interval a, and this is substantially the same as the value of $b(x)$ at x. Hence the reverse voltage is

$$W = W[b(x)], \qquad (28)$$

where W has the functional dependence on b derived for $I = 0$.

In several parts of this section we shall take the potential on the gate as zero. This is especially convenient when considering how the potential varies along the channel.

One of the assumptions involved in establishing the function $W(b)$ in Section 2 was that $E_y = 0$ at $y = b$. This assumption is not exactly fulfilled for the gradual case, but is a good approximation. Inside the channel the electric field is predominantly in the x direction. At the edges there is a component in the y direction which produces the inward component of hole-drift velocity that corresponds to the narrowing of the channel. Thus at the edge of the channel

$$- E_y : E_z = db/dx. \qquad (29)$$

For the gradual approximation, db/dx must be much less than unity; if it were not, the channel would narrow abruptly. As will be shown below, E_z is also small compared to typical values of E_y in the region between the channel and the gate. Hence the value of E_y at the surface of the channel may be neglected so far as calculations of $W(b)$ are concerned. Since E_y is even smaller in the interior of the channel, where the current flow is more nearly parallel to the x-direction, the equipotential surfaces in the channel are well approximated by planes perpendicular to the x-axis; the electrostatic potential on these planes is $-W(x)$ since this is the potential difference between the $(n+)$-regions, which are equipotentials and taken to be at zero potential, and the channel.

We are now in a position to analyze the gradual case in terms of $W(x)$ and g. Since g is a function of b and since b and W are related as discussed in Section 2, g is the same function of W as in Section 2. This leads at once to the equation

$$I = g(W)dW/dx. \qquad (30)$$

This is the basic equation on which the theory of the gradual case is developed. It is applicable not only to the structure of Fig. 1 but to others in which there may be concentration gradients rather than abrupt transitions. For such other cases, new functions $g(W)$ can be derived. Once the appropriate $g(W)$ is known, the analysis is similar to that presented below.

Before discussing the consequences of (30) we shall point out that such an equation does not apply if the gradual condition is badly violated. We shall illustrate this by supposing that the channel width varies rapidly. Then the potential at x is not determined by $b(x)$ alone but by the much larger charges that lie to the left of x. Under these conditions, then, W cannot be expressed as a function of b alone and, consequently, b and hence g are not functions of W alone, but depend also on derivatives of W. Another case in which W and b are not directly related may occur if the potential at x is determined in large measure by the potential applied to the end electrode. In this case it is evident that there is no unique relationship between $b(x)$ and $W(x)$. This latter case will be examined in detail in subsequent sections in which we suppose that W exceeds W_0 near one electrode.

B. The Current-Voltage Relationships

Returning to the gradual case, we assume that (30) is valid and integrate it formally by multiplying by dx. We thus obtain

$$I \int dx = \int g(W)dW. \qquad (31)$$

Denoting the values of W at source and drain by W_s and W_d and letting L represent their separation, we obtain

$$I = (1/L) \int_{W_s}^{W_d} g(W)dW. \qquad (32)$$

The current voltage characteristics based on this equation may be conveniently expressed in terms of a function of one variable having the dimensions of a current

$$J(W) \equiv \int_0^W g(W)dW$$
$$= g_0 W[1 - (2/3)(W/W_0)^{1/2}], \qquad (33)$$

the last expression following from (17). In terms of J, the current per unit length I takes the form

$$I = [J(W_d) - J(W_s)]/L$$
$$= [J(V_g - V_d) - J(V_g - V_s)]/L. \qquad (34)$$

If I is plotted as a function of V_d with V_g as a parameter and $V_s = 0$ (analogous to a collector family for a bipolar transistor or a plate family for a vacuum-tube triode), then all of the curves have the same shape but are simply translated parallel to each other along the I and V_d axes. Such a plot is shown in Fig. 7, absolute values of V_d being used so the curves fall in the first quadrant.

These characteristics cannot be continued beyond

$$W_d = V_g - V_d = W_0 \qquad (35)$$

on the basis of (34) since for such potentials pinch-off occurs at the drain and the gradual approximation fails. The continuation shown in Fig. 7 is based on a treatment given in Section 5. The critical condition for pinch-off at the drain is indicated by the dashed line. This curve has the equation

$$I = J(W_0)/L - J(W_0 - |V_d|)/L. \tag{36}$$

From this it follows that the dashed curve is the same as the curve for $V_g = 0$, except that it is rotated through $180°$ and has its origin at $|V_d| = W_0$ and $I = J_0(W_0)/L$.

The small-signal behavior corresponding to (32) is readily obtained by making small changes v_s, v_d, and v_g in the V's. This leads to a small change i in I of

$$i = (g_d \delta W_d - g_s \delta W_s)/L$$
$$= (g_s/L)v_s - (g_d/L)v_d - (g_s - g_d)v_g/L, \tag{37}$$

where

$$g_s \equiv g(W_s) = g(V_g - V_s) \tag{38}$$

$$g_d \equiv g(W_d) = g(V_g - V_d). \tag{39}$$

Making use of the notation

$$G_{dg} = g_s/L \tag{40}$$

$$R_d = L/g_d \tag{41}$$

$$\mu_{dg} = G_{dg}R_d = g_s/g_d, \tag{42}$$

the current equation may be put in the form

$$i_d = -i = G_{dg}[(v_g - v_s) + (v_d - v_g)/\mu_{dg}], \tag{43}$$

which is the same as that for a vacuum-tube triode, provided we note the ac current into the drain from outside is $i_d = -i_s = -i$.
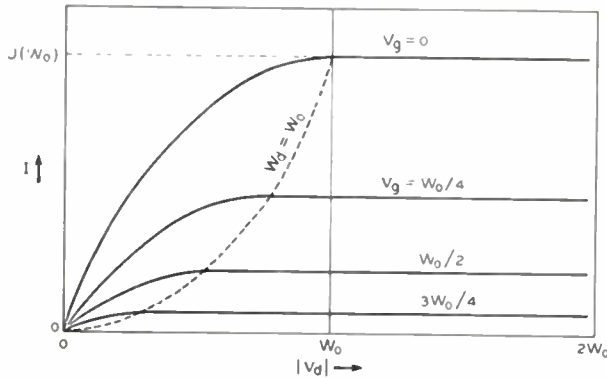
Fig. 7—Receiver family with source at zero potential.
(Note $W_2 = W_0$ line.)

The equivalent circuit corresponding to (43) is shown in Fig. 8, with the nonreciprocal element represented as a current generator.

It is evident that a high $m\mu$ will be obtained by operating with W_d nearly equal to W_0 so that the channel is nearly "pinched-off" at the drain and $g_d = g(W_d)$ is much less than g_s. We shall return to this subject in more detail in Section 5. The circuit is incomplete owing to the absence of capacitive elements. These omitted elements will play an important role at frequencies as high as those discussed in Section 7, and at these frequencies a lumped constant treatment becomes a poor approximation.

C. Natural Parameters

Although the circuit characteristics for the gradual case can be obtained from (32) as soon as $g(W)$ is known as a function of W, it is desirable to understand the behavior of the channel in greater detail. For this purpose,

we shall derive expressions for the dependence of W upon x along the channel and the shape of the channel in the x-y plane. It is convenient to use certain natural parameters associated with the constants of the structure.
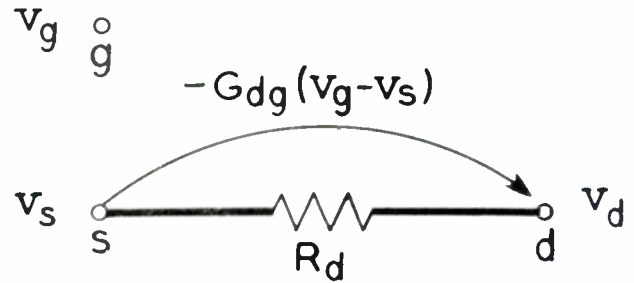
Fig. 8—Equivalent circuit for unipolar field-effect transistor.

The natural unit of voltage, W_0, has already been derived. The natural unit of field, E_0, is that at the plane $y = a$ when W_0 is applied. The natural unit of conductance is g_0. The natural unit for current per unit length is the current associated with g_0 and a field E_0. The natural unit of time, τ_0, is the time required for a hole to travel a distance a in a field E_0. The equations for these quantities are

$$W_0 = \rho_0 a^2/2K \tag{44}$$

$$E_0 = \rho_0 a/K = 2W_0/a \tag{45}$$

$$g_0 = 2\rho_0 \mu_0 a \tag{46}$$

$$I_0 = g_0 E_0 = 2\rho_0^2 \mu_0 a^2/K = 8\mu_0 K W_0^2/a^2 \tag{47}$$

$$\tau_0 = a/\mu_0 E_0 = K/\mu_0 \rho_0 = K/\sigma_0. \tag{48}$$

Several forms have been given for some of the quantities above. These are convenient for subsequent manipulations.

The expression K/σ_0 for τ_0 should be noted. This expression is simply the "dielectric relaxation time" for material characterized by σ_0 and K.

D. Expressions for the Solution

It is convenient not to solve for W as a function of x but instead for x as a function of b. Since the relationship between b and W has already been derived, this establishes the relationship between x and W. In order to find x as a function of b, we note that, from (12) and (13),

$$dW = -(a - b)\rho_0 db/K. \tag{49}$$

Using this equation to eliminate dW, we obtain

$$Idx = g(W) dW$$
$$= -g_0(a - b)(b\rho_0/aK)db$$
$$= -I_0 a(1 - u)u du, \tag{50}$$

where

$$u \equiv b/a \tag{51}$$

is the fraction of the p layer occupied by channel. It should be noted that dx/du is zero at $u = 0$ and $u = 1$. Formally, this means that the channel is changing its width with an infinite value of db/dx at those points.

When u is between 0 and 1, dx/du is symmetrical about $u = 0.5$. The equation for dx is readily integrated to give

$$x = -(aI_0/I)[(u^2/2) - (u^3/3)], \tag{52}$$

where the integration constant has been chosen so that $x = 0$ when $u = 0$. The resultant shape of the channel is shown in Fig. 9(a) for two values of I_0/I. In Fig. 9(b), the variation of potential along the channel is shown. These curves are obtained by computing values of the electrostatic potential $V(x, y)$, which may be denoted as $V(x, 0)$ in the channel,

$$V(x, 0) = -W(x) = -W_0(1 - u)^2, \tag{53}$$

for the same values of u as were used to compute x for Fig. 9(a) and then by plotting related V and x values.

The total length of the channel is

$$L(I) = (a/6)(I_0/I). \tag{54}$$
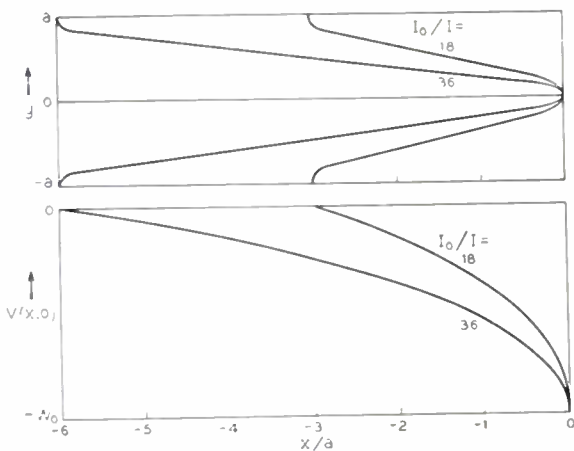
Fig. 9—The gradual approximation. (a) Shape of channel. (b) Electrostatic potential.

It is evident that the conditions for the gradual case can hardly be considered to hold unless $L(I)$ is about $3a$ or more, corresponding to I_0/I greater than 18. Even if $L(I)$ is equal to $3a$, the gradual condition will hold only in the middle of the channel. Where the channel narrows rapidly, the gradual condition will not hold.

As an approximate criterion for the limit of validity of the gradual approximation we shall require that over a distance a the fractional change in channel width shall be small. This leads to

$$a\,db/dx < b \tag{55}$$

or to

$$(a/b)db/dx < 1. \tag{56}$$

This seems a reasonable condition. If it holds, then db/dx is small and the arguments given early in this section for the relationship between b and W apply. On the other hand, if the condition fails, then b may differ by a factor of 2 or more in a distance a. Under these conditions it is not evident that the potential at a point x is determined by b at that value of x, since it will be strongly influenced by variations of b to both sides.

We shall use the criterion in form

$$1 > (a/b)(db/dx) = I/(1 - u)u^2I_0. \tag{57}$$

If I_0/I is 18, then the inequality is satisfied for values of u in the neighborhood of 0.5. For larger values of I_0/I and longer channels, the gradual condition will be satisfied over a wider range of channel widths. In all cases, however, the gradual condition will fail when u approaches unity or zero. The failure of the gradual condition near $u = 1$ is of little importance since the value of g is large there and the electric field E_x is small. Near $u = 0$, however, the electric field is high; furthermore, $u = 0$ will be approached in all cases in which W_d is greater than W_0.

Near $u = 0$, the term $1 - u$ is practically unity and the inequality may be written in the form

$$1 > I/u^2I_0. \tag{58}$$

In this range the u^3 term can be neglected in (52) for x so that

$$-x \doteq I_0au^2/2I. \tag{59}$$

Inserting the inequality here leads to

$$-x > a/2, \tag{60}$$

which implies that the gradual condition is applicable at distances greater than $a/2$ away from the end of the channel.

Actually the electric field does not approach infinity at $x = 0$ and the number of holes per unit length does not go to zero as implied by the gradual approximation. Thus the pinch-off condition at $x = 0$ does not really occur. Nevertheless, it is convenient to define a certain point as $x = 0$ and to refer to it as the *extrapolated pinch-off point*, or the *expop* for short. The expop is defined as the point at which u would vanish and W would equal W_0 if the gradual solution were continued beyond its range of validity.

In the next section we shall consider what occurs at and beyond the expop.

5. Conditions in the Expop Region

Before considering the conditions prevailing at the expop, we shall consider the conditions somewhat beyond it corresponding to $x > 0$. In this region, the charge per unit length due to holes will be very small compared to that due to $q(N_a - N_d)$. Thus that part of the electrostatic potential which is due to the charges may be well approximated by

$$V_0(x, y) = -W_0[1 - (y/a)^2], \tag{61}$$

which satisfied Poisson's equation and vanishes for $y = \pm a$. This potential is independent of x and produces no field which causes the holes to flow. The cause of the field E_x is the potential applied to the drain.

If we assume that the drain is located at a relatively large value of x and has a large negative charge, then we can approximate the field which it exerts near $x = 0$ by a relatively simple expression. This expression may be obtained by considering a Fourier expansion of poten-

tial on the plane $x = x_d$ at which the drain is located. If the potential is symmetrical about $y = 0$, it can be expanded in a cosine series with terms of the form

$$\cos [\pi(2n + 1)y/2a]; \quad n = 0, 1, 2. \quad (62)$$

Each term vanishes at $y = \pm a$. Each term can be combined with a function of x so as to obtain a solution of Laplace's equation, the complete terms being

$$T_n = \exp [\pi(2n + 1)x/2a] \cos [\pi(2n + 1)y/2a]. \quad (63)$$

Each such term satisfies

$$d^2T_n/dx^2 + d^2T_n/dy^2 = 0. \quad (64)$$

Terms with negative exponentials would satisfy Laplace's equation also, but would increase towards the left and would not represent properly a field due to the potential on the drain. The term with $n = 0$ decays by a factor of $\exp(-\pi/2) = 0.21$ in each interval a towards the left. The next term with $n = 1$ decays similarly but by a factor of $\exp(-3\pi/2) = 0.009$. From this it is evident that, except in the immediate neighborhood of the drain, the potential due to the drain can be represented by a single term

$$T_0 = (\exp \pi x/2a) \cos \pi y/2a. \quad (65)$$

This term gives rise to a set of equipotentials of the form represented in Fig. 10.
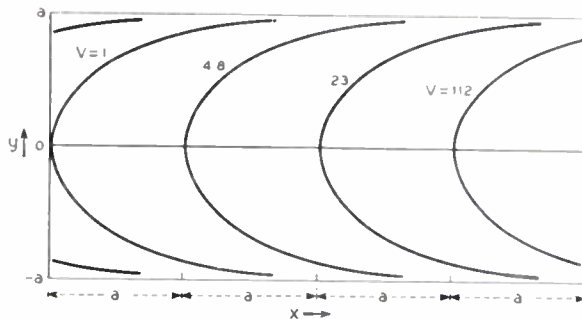
Fig. 10—Equipotentials for field satisfying Laplace's equation.

In the region in which the charge due to holes is small the potential can thus be well represented by

$$V(x, y) = V_0(x, y) - A T_0(x, y). \quad (66)$$

The coefficient A must be chosen so that $V = -W_d$ at the drain. Actually near the drain, other terms in the Fourier expansion should be included. However, if the drain has a shape somewhat like the equipotentials for T_0, these other terms may be quite small. We shall, therefore, neglect them and determine A by requiring that $V = -W_d$ at the point $(x_d, 0)$. This leads to

$$W_d = W_0 + A \exp (\pi x_d/2a), \quad (67)$$

or

$$A = (W_d - W_0) \exp (-\pi x_d/2a). \quad (68)$$

We shall next consider how operating conditions affect the value of x_d, the distance between the extrapolated pinch-off point and the drain.

In Fig. 11 we represent the dependence of $V(x, 0)$ upon x including a range of x in which the approximations that

$$x \doteq (I_0 a/2I)u^2 \quad (69)$$

and

$$W = W_0(1 - u)^2 \doteq W_0 - 2W_0 u$$
$$= W_0 - W_0(8Ix/I_0 a)^{1/2} \quad (70)$$

are valid. This solution will be valid for $x < -a$, where the gradual approximation holds. For positive values of x, it must join smoothly onto the solution in the space-charge region. The problem of joining is discussed more fully in an Appendix. It may be helpful to give an estimate of the coefficient A based on simpler reasoning.
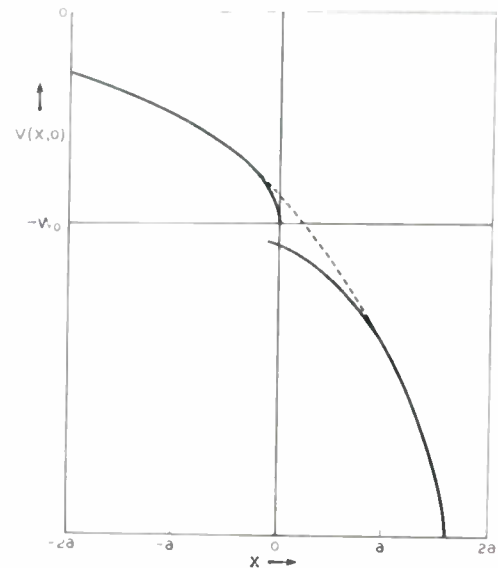
Fig. 11—Approximate joining of gradual solution to exponential.

In the gradual region, the longitudinal electric field is produced by the variation in hole-charge per unit length. This variation produces a changing potential along the channel. As the expop is approached, the hole charge becomes negligible and further variation from this cause is impossible. Consequently, the electric field must be furnished from the potential on the drain. Since the gradual solution begins to fail badly at $x = -a/2$, it is reasonable to assume that the T_0 term makes an appreciable contribution to the field at that point. As an approximate joining condition at $x = -a/2$ we shall let the contribution of T_0 be 25 per cent as large as that of the gradual solution. This leads to the approximation shown in Fig. 11 and to

$$dW/dx = W_0(8I/I_0 a)^{1/2}x^{-1/2}/2$$
$$= E_0(I/I_0)^{1/2} = 4A dT_0/dx$$
$$= 4(\pi A/2a) \exp (-\pi/4). \quad (71)$$

This leads to a value for A of

$$A = (W_0/\pi)(I/I_0)^{1/2} \exp (\pi/4)$$
$$= 0.7W_0(I/I_0)^{1/2}. \quad (72)$$

On the basis of this value of A, the distance between the expop and drain can be estimated as a function of W_d and I,

From (67) we find

$$x_d = (2a/\pi) \ln [(W_d - W_0)(I_0/I)^{1/2}/0.7W_0]. \quad (73)$$

As an example, we shall take $I_0/I = 18$, which gives a value of $3a$ for the maximum channel length, from full width to expop, and shall take W_d as $2W_0$. This leads to

$$x_d = (2a/\pi) \ln (4.25/0.7) = 1.1a. \quad (74)$$

Hence the drain lies approximately a distance a beyond the expop.

On the basis of the gradual solution and the expression for x_d it is possible to develop the circuit theory for operation beyond the pinch-off condition in much the same way as was done for operation entirely in the gradual condition. The following presents this theory.

6. Equivalent Circuit for Low Frequencies

In Section 2 we considered a structure of length L between the source and drain and introduced a function $J(W)$ defined by

$$J(W) = \int_0^W g(W)dW. \quad (75)$$

In terms of this function the current per unit length flowing into the source from outside is

$$I = [J(V_g - V_d) - J(V_g - V_s)]/L. \quad (76)$$

Curves based on this formula were shown on Fig. 7. For $V_d - V_g$ more negative than $-W_0$, the curves were continued with a constant value of

$$I = [J(W_0) - J(V_g - V_s)]/L. \quad (77)$$

We now examine the justification of this continuation.

If we suppose that W_s is sufficiently large compared to W_0, then we may use (73) to estimate x_d. This means that the effective length of the channel should be taken as L_s with

$$L_s = L - x_d. \quad (78)$$

In other words the channel adjusts its current until the expop falls short of the drain by a distance x_d. Under these conditions the equation

$$I dx = g(W)dW \quad (79)$$

holds along the gradual part of the channel. Furthermore, if the expression is integrated from W_s to W_0, then the integral of dx must simply be L_s, since by definition L_s is the distance from the source to the expop when (79) is integrated to W_0. This gives

$$IL_s = \int_{W_s}^{W_0} g(W)dW = J(W_0) - J(W_s). \quad (80)$$

This leads to

$$I = [J(W_0) - J(V_g - V_s)]/L_s. \quad (81)$$

It is evident that this current is larger than that used previously by the ratio

$$L/L_s = [1 - (x_d/L)]^{-1}. \quad (82)$$

This ratio will be nearly unity if L is 5 or more times a since x_d will in general be comparable to a. Thus the extrapolation used in Fig. 7 is seen to be reasonable.

We may use these expressions to obtain differential forms for small-signal effects. For this purpose we write the equation in the form

$$I(L - x_d) = \int_{W_s}^{W_0} g(W)dW. \quad (83)$$

We then take total differentials of both sides, letting

$$\delta I = i, \quad \delta W_s = v_g - v_s, \quad \delta W_d = v_s - v_d, \quad (84)$$

and note that

$$I/g(W_s) = E_{xs}, \quad (85)$$

the electric field at the source. Because of the exponential dependence of term T_0, we note also that the electric field at the drain is

$$E_{xd} \doteq \pi(W_d - W_0)/2a. \quad (86)$$

The contributions of $x_d(W_d, I)$ to the differential is

$$\delta x_d = - (ai/\pi I) + \delta W_d/E_{xd}. \quad (87)$$

These considerations lead to

$$i = I\left(\frac{v_s}{E_{xs}} - \frac{v_d}{E_{xd}} + v_g\left(\frac{1}{E_{xd}} - \frac{1}{E_{xs}}\right)\right)$$
$$\div [L - x_d + a/\pi]. \quad (88)$$

From these we see that the mμ is

$$\mu_g d = (E_{xd}/E_{xs}) - 1. \quad (89)$$

The meaning of this expression may be seen by thinking in terms of Fig. 11. Here we consider the curve of $V(x, 0)$ versus x for current I. A change in potentials which keeps the currents the same corresponds to changing W_s and W_d by amounts δW_s and δW_d while having L fixed. Thus L_s must decrease and x_d increase by equal amounts. The changes of δW_s and δW_d are evidently in the ratios of E_{xs} and E_{xd}. This leads to (89).

The dominant term in $\mu_g d$ is

$$E_{xd}/E_{xs} = [\pi(W_d - W_0)/2a]g(W_s)/I$$
$$= \frac{\pi(W_d - W_0)}{2a} \cdot \frac{I_0 b_s}{E_0 I a}$$
$$= \frac{\pi}{4}\left(\frac{W_d}{W_0} - 1\right)\frac{I_0}{I}\frac{b_s}{a}. \quad (90)$$

If W_d is $2W_0$ and $I_0/I = 18$ and $W_s = (1/4)W_0$ so that $b_s/a = 1/2$, the value of $\mu_g d$ is

$$\mu_g d = 7.1 - 1 = 6.1. \quad (91)$$

Higher values of mμ can be obtained by operating with smaller values of I/I_0 or larger values of W_d.

The transconductance is dominated by the term in E_{xs} and is negative with magnitude

$$G_g = \mu_g d g_s \div (L - x_d + a/\pi)(1 + \mu_g d)$$
$$\doteq 2\sigma_0 b_s \mu_g d (L - x_d + a/\pi)(1 + \mu_g d). \quad (92)$$

The value is substantially that of a p-layer $2b_s$ thick and L long.

The resistance looking into the drain is

$$R_d = 1/G_g \mu_g d, \tag{93}$$

and is thus substantially that of a p-layer $2b_s$ thick and $\mu_g d$ long.

7. Effects at High Frequency

Although the gradual case lends itself to analysis from the transient point of view, the analysis is somewhat tedious. We shall, therefore, discuss in simple terms the frequency at which the circuit constants deviate substantially from their dc values. For this purpose we shall consider a structure operated with the drain beyond pinch-off. We shall also assume that W_s is approximately $W_0/4$ so that $b_s = a/2$. The structure then may be regarded as consisting of a distributed resistance of total value somewhat greater than

$$R = L/2b_s\sigma_0 = L/a\sigma_0, \tag{94}$$

which charges a pair of capacitances of length L and spacing somewhat more than $a/2$. The value of the resultant capacity is thus less than

$$C = 2LK/(a/2). \tag{95}$$

If we consider that the charging current flows through about $R/2$, then the effective time constant of the combination is thus

$$\tau = RC/2 = 2(L/a)^2 K/\sigma_0 \tag{96}$$

$$= (I_c/I)^2 \tau_0/18. \tag{97}$$

If we imagine that an input signal at frequency $f = \tau/2\pi$ is applied to the gate, then the input power will be

$$i_g^2 R/2 \tag{98}$$

and the voltage will be

$$v_g = i_g R/2^{1/2}. \tag{99}$$

If we neglect transit-time effects, this voltage would produce a current to the drain, if grounded, of

$$i_d = G_g v_g = (a\sigma_0/L)v_g \tag{100}$$

and this output current generator can furnish power of magnitude

$$i_d^2 \mu_g d/4G_g \tag{101}$$

to a matched load of $\mu_g d/G_g$. The power gain is thus

$$\mu_g d G_g/4R = \mu_g d/4. \tag{102}$$

Thus power gain may still result even at frequencies as high as $1/2\pi\tau$ corresponding to

$$f = (2.9)(I/I_0)^2/\tau_0. \tag{103}$$

At lower frequencies, the power gain is higher since the input impedance is largely capacitive and increases with decreasing frequency so that for a given voltage there will be less current and less input power.

At higher frequencies there can still be power gain, but it is necessary to treat channel as a wave-transmitting medium rather than as a lumped constant circuit.

8. Some Quantitative Design Considerations

For purposes of example we shall consider a structure comprising a layer of n-type germanium bounded by p-type gates. Under operating conditions, the electric field in the specimen may well be as large as $2E_0$ in the neighborhood of the drain. One design consideration is that this field should not be as large as the "Zener field" at which appreciable field generation of hole electron pairs occur.[6]

In order for the theory developed in the previous sections to apply, the electric field along the channel should be less than E_c, the critical field at which Ohm's law fails due to inability of the electrons to dissipate their energy sufficiently rapidly. This field is about 900 volts/cm for electrons in n-type germanium at room temperature,[7] and for it the drift velocity is about

$$3.1 \times 10^6 \text{ cm/sec.} \tag{104}$$

Above this field, the drift velocity varies approximately as the square root of the electric field.

If we require that the constant mobility region extend to the point where the gradual approximation fails, we can write from (71)

$$dW/dx = E_c = E_0(I/I_0)^{1/2}$$

so that for $I_0/I = 18$, $E_0 \doteq 4,500$ volts/cm. This condition is considerably more stringent than the Zener field condition, which would permit E_0 values at least 10 times larger to be used.

Throughout the discussion we have neglected currents flowing to or from the gate. In the case of a p-layer structure a reverse current of thermally generated electrons will flow from the p-layer to the electrodes and a current of holes will flow from the n-regions into the p-layer. These currents will be of the same order of magnitude as those met with in p-n junctions biased reverse. The current of electrons will tend to be somewhat larger than a normal saturation current because of the large size of the space-charge region and enhanced generation therein,[8] and will tend to be somewhat smaller because usually the p-layer will be considerably less than one diffusion length thick. A typical reverse current in a p-n junction in germanium[9] may be of the order of 10^{-3} amp/cm².

We shall illustrate orders of magnitudes for an n-type unipolar transistor by taking n-type material having

[6] K. B. McAfee, E. J. Ryder, W. Shockley, and M. Sparks, "Observations of Zener current in germanium p-n junctions," *Phys. Rev.*, vol. 83, pp. 650–651; 1951.

[7] E. J. Ryder and W. Shockley, "Mobilities of electrons in high electric fields," *Phys. Rev.*, vol. 81, p. 139; 1951. Also, W. Shockley, "Hot electrons in germanium and Ohm's law," *Bell Sys. Tech. Jour.*, vol. 30, p. 990; 1951. See also E. Conwell, "Properties of silicon and germanium," PROC. I.R.E., vol. 40, pp. 1327–1338; this issue.

[8] W. Shockley and W. T. Read, Jr., "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, p. 835; 1952.

[9] F. S. Goucher, G. L. Pearson, M. Sparks, G. K. Teal, and W. Shockley, "p-n junction rectifier and photo-cell," *Phys. Rev.*, vol. 81, p. 637; 1951. Also, W. J. Pietenpol, *Phys. Rev.*, vol. 82, p. 120; 1951.

$\times 10^{13}$ electrons/cm³ and a charge density and conductivity of

$$\rho_0 = 1.6 \times 10^{-19} \times 4 \times 10^{13}$$
$$= 6.4 \times 10^{-6} \text{ coulombs/cm}^3 \qquad (105)$$
$$\sigma_0 = 3600\rho_0 = 2.3 \times 10^{-2} \text{ mho/cm.} \qquad (106)$$

We shall take a field E_0 of 10^4 volts/cm, but shall operate with $I_0/I = 60$ so that throughout most of the channel the mobility is constant. The resulting value of a is

$$a = KE_0/\rho_0 = 1.41 \times 10^{-12} \times 10^4/6.4 \times 10^{-6}$$
$$= 2.2 \times 10^{-3} \text{ cm} \qquad (107)$$

and

$$W_0 = aE_0/2 = 11 \text{ volts.} \qquad (108)$$

The characteristic current I_0 will be

$$I_0 = 2a\sigma_0 E_0 = 1.0 \text{ amp/cm.} \qquad (109)$$

If this structure is operated with a spacing L between source and drain of $5a$, and $W_s = W_0/4$ so $b_s = a/2$, and $V_d = 2W_0$, the following characteristics are found:

$$L = 1.1 \times 10^{-2} \text{ cm} \qquad (110)$$
$$W_s = 3 \text{ volts} \qquad (111)$$
$$W_d = 22 \text{ volts} \qquad (112)$$
$$I = I_0/60 = 17 \text{ ma/cm} \qquad (113)$$
$$\text{power dissipation} = 0.38 \text{ watts/cm} \qquad (114)$$
$$\mu_{gd} = 23 \qquad (115)$$
$$G_g = 4,600 \text{ micromhos/cm} \qquad (116)$$
$$G_g/I = 0.27 \text{ volts}^{-1} \qquad (117)$$
$$\text{output resistance} = 5,000 \text{ ohms cm} \qquad (118)$$
$$\tau_0 = 6.1 \times 10^{-11} \text{ sec} \qquad (119)$$
$$f = 9 \times 10^6 \text{ cps from (103)} \qquad (120)$$

The electric field at the source is 300 volts/cm and at $L/2$ short of the expop it is 1,300 volts/cm; hence, as stated above, the assumption of constant mobility in the channel is a good one.

The area of *p-n* junction per cm of length is $L = 1.1 \times 10^{-2}$. For a reverse current of 1 ma/cm² for these junctions, the reverse current would be about 10^{-2} ma/cm for the structure. This value would be negligible compared to the working current of 17 ma/cm.

These very superficial design considerations have been presented to illustrate the theory and to show that a favorable relationship between size and frequency applies for unipolar transistors.

Appendix

The problem of obtaining an accurate solution for the pattern of flow in a field-effect transistor appears to be a difficult one. The approximation of electrical neutrality within the channel itself can be shown to be seriously in error at a distance of about $a/2$ from the expop. The solution would thus require finding the charge density

for the carriers as a function of x and y, which vanished outside the boundaries of the channel.

It is possible to give physical reasoning, however, that shows that once the channel has become relatively narrow, one can eliminate one of the variables. Suppose that instead of allowing the carriers to flow over a channel of finite width we imagine them to be constrained to the plane $y = 0$, keeping the same number of carriers per unit area of the plane. The effect of this shift of charge on the potential is quite small. If we consider a case in which the channel is $a/4$ wide, then the value of W is $W_0(3/4)^2$ or $0.56W_0$. If the same number of carriers are compressed into the $y = 0$ plane, the potential is W_0 due to the fixed charge minus $aE_0/4$ due to the field $E_0/4$ of the carriers acting over a distance a. This results in a potential at the $y = 0$ plane of $0.50W_0$, or about 10 per cent less than the correct value. For narrower channels, the error will be proportionally smaller.

The advantage of assuming that the carriers all flow in one plane is that the charge density of the carriers can now be described by a function of x alone rather than by a function of both x and y.

If we assume that the mobile carriers are restricted to the central plane at $y = 0$, then it is possible to show that a very simple scaling law applies for the case of constant mobility. For this case we may advantageously write the potential in the form

$$V(I, x, y) = (\rho_0/2K)(y^2 - a^2) + V_m(I, x, y).$$

The ρ_0 term satisfies Poisson's equation for $0 < |y| < a$ and gives a uniform potential $-W_0$ at $y = 0$. The V_m term thus represents the fields due to the mobile charges and due to the potentials applied to the end electrodes. The simple result, which we shall now establish, is as follows: If

$$V_1(x, y) \equiv V_m(I_1, x, y)$$

is one function which satisfies the conditions of the problem for current I_1, then a solution that holds for current I_2 is obtained by replacing V_1 by

$$V_2(x, y) = (I_2/I_1)^{1/2}V_1(x, y).$$

This follows from the fact that if V_1 satisfies Laplace's equation for $0 < |y| < a$, then so does V_2. Furthermore, the mobile carrier charge density in the $y = 0$ plane is evidently

$$- 2K\partial V/\partial y,$$

and this is larger for V_2 than for V_1 by the ratio $(I_2/I_1)^{1/2}$. Furthermore, the longitudinal field is larger for V_2 by the same ratio. Hence the current is larger in the ratio of charge times field and hence is larger by the ratio I_2/I_1. Hence if V_1 satisfies boundary conditions and gives current I_1, then V_2 satisfies them also and gives current I_2.

We shall next consider what the solution is for this case in which the mobile carriers are assumed to move in the $y = 0$ plane with constant mobility. For these assumptions, the solution in the gradual region takes a very simple form. It also takes a very simple form in the region beyond the expop. We shall discuss both of these

cases and then present a simplified treatment of the transition of one solution to the other.

Solution is readily obtained for the gradual region. The electric field in a region where charge density per unit area is uniform with value P, which corresponds to $2bp_0$ in the earlier discussion, is simply

$$- \partial V/\partial y = E_y = P/2K.$$

This gives for the potential

$$V_m = (P/2K)(a - y) \equiv - W_m[1 - (y/a)],$$

where W_m is a quantity analogous to W as used in the text. The longitudinal field at $y = 0$

$$E_x = - \partial V/\partial x = - (a/2K)dP/dx = dW_m/dx.$$

The current equation is

$$I = \mu P E_x = (2K\mu/a)W_m dW_m/dx,$$

which integrates to give

$$W_m{}^2 = - (aI/\mu K)x$$

provided the integration constant is chosen so as to make the extrapolated pinch-off point come at $x = 0$.

The potential at any value of x arises chiefly from P values within a distance of a about that value of x. A condition for the validity of the gradual approximation will thus be that the value of $P(x)$ does not differ greatly from its average value between $x - a$ and $x + a$. Near some point x_0 we may take

$$P(x) = P(x_0) + P'(x - x_0) + (1/2)P''(x - x_0)^2.$$

Average value of $P(x)$ in the interval of $\pm a$ about x_0 is

$$\langle P \rangle = P(x_0) + (a^2/6)P''(x_0).$$

If this average is to differ only slightly from $P(x_0)$, then

$$1 < a^2 P''/6P = a^2/24x^2.$$

Hence the gradual approximation should hold quite well until x is less than a distant from the expop.

Well to the right of the expop, the potential due to the mobile carriers must be small compared to that due to the potentials applied to the drain electrode. This latter potential will be of the form

$$\cos (\pi y/2a) \exp (\pi x/2a),$$

and may be referred to as "charge-free" potential since it satisfies Laplace's equation for all values of x, y

In the neighborhood of the extrapolated pinch-off point, the potential will make a transition from the gradual case to the charge-free case. In Fig. 12, the limiting forms are shown, the scale of potential being selected so that unity comes at $W_m(a)$. A solution of the space-charge equations would be necessary in order to determine the coefficient of the charge-free solution which joins correctly to the gradual solution. Four charge-free curves are drawn so that they take values of -1, $-\frac{1}{2}$, $-\frac{1}{4}$ and $-\frac{1}{8}$ times $W_m(a)$ at the extrapolated pinch-off point. Since these curves are exponentials, the difference between them is equivalent to a shift by $(2a/\pi) \ln 2 = 0.45a$ along the x-axis. From inspection, it

is seen that joining can be accomplished most smoothly for the case of $- W_m(a)/4$. We shall accordingly assume that this curve does represent the correct joining relationship. An exact calculation would be necessary to determine the relationship accurately, but it seems unlikely that the result would correct the present guess by more than about $a/2$ for the translational position of the charge-free solution.

On the basis of the reasoning above, we take as an approximation for $W(x)$ the following forms:
For $x < 0$,

$$W(x) = W_0 - (ai/\mu K)^{1/2}(- x)^{1/2}.$$

For $x > 0$,

$$W(x) = W_0 + (1/4)(a^2 i/\mu K)^{1/2} \exp (\pi x/2a)$$
$$= W_0[1 + (i/2i_0)^{1/2} \exp (\pi x/2a)].$$

The latter expression leads to a value of x_d, as defined in the text, of

$$x_d = (2a/\pi) \ln [(W_2 - W_0)(2i_0/i)^{1/2}/W_0].$$
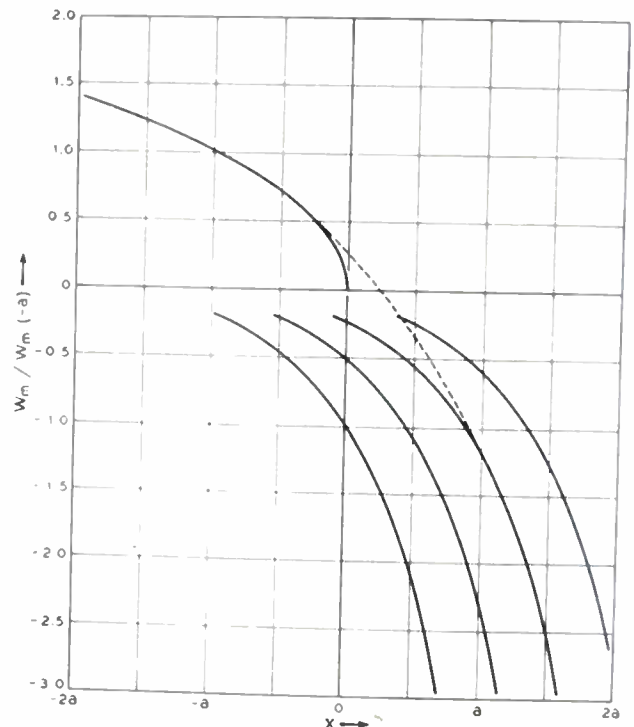
Fig. 12—Graphical fitting of gradual solution to an exponential solution.

This value differs by less than $0.1a$ from that given in the text. The reason for the good agreement is that the selection of $(1/4)$ for the ratio of the two fields used in the text was based on considerations of this appendix.

Junction Fieldistors[*]

O. M. STUETZER†, ASSOCIATE, IRE

Summary—A high-impedance, input-low impedance output amplifying device which utilizes surface conductivity control in the neighborhood of a *p-n* junction is described. The transconductances of the order of 1,000 micromhos can be reproduced at very low frequencies.

INTRODUCTION

A SCHEMATIC DIAGRAM (Fig. 1) shows the essentials of the experimental device we shall discuss. A *p-n* junction is biased in the reverse direction by a current I_a. A control electrode is attached in close proximity to the junction. The distance d is of

PN - JUNCTION FIELDISTOR

Fig. 1—Schematic diagram of arrangement.

the order of a few microns, the breadth B, of the order of 100 microns for practical devices. An ac signal can be applied to the control electrode on top of a variable dc bias $\pm U_g$. The circuit is balanced to keep the center of the junction electrically on zero and to measure all potentials versus this zero. For practical applications this is, of course, not necessary. Fig. 2 shows experimental models.

The arrangement is very similar to the point- or line-contact "fieldistor" described previously.[1] The voltage U_g exerts a controlling effect on the current I_a. As a quality figure for the device, we will choose the transconductance $G_m = (\partial I_a/\partial U_g)_{v_a}$.

The controlling effects depend on the state of the surface. The experiments to be reported were carried out with Ge junctions (Diodes M1470 and M1728, Bell Telephone Laboratories). They were cleaned with alcohol only; the surface, which comes heavily treated, was not changed.

In the arrangement of Fig. 1, this material produces devices with a G_m of several micromhos and a flat frequency response up to several hundred kc. Liquid dielectrics between control electrode and surface (f.i.

water free oils) increases the effect roughly in proportion with their dielectric constant.

However, if we introduce between surface and control electrode a liquid with a reasonably high polar moment, G_m changes sign and increases drastically. The output
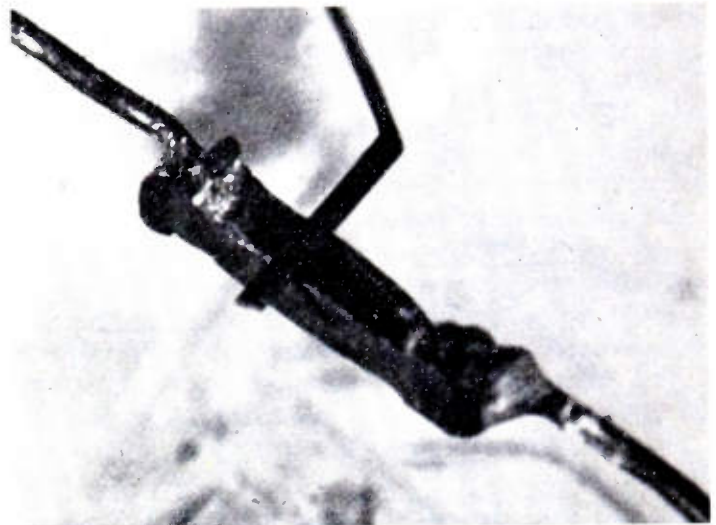
(a)
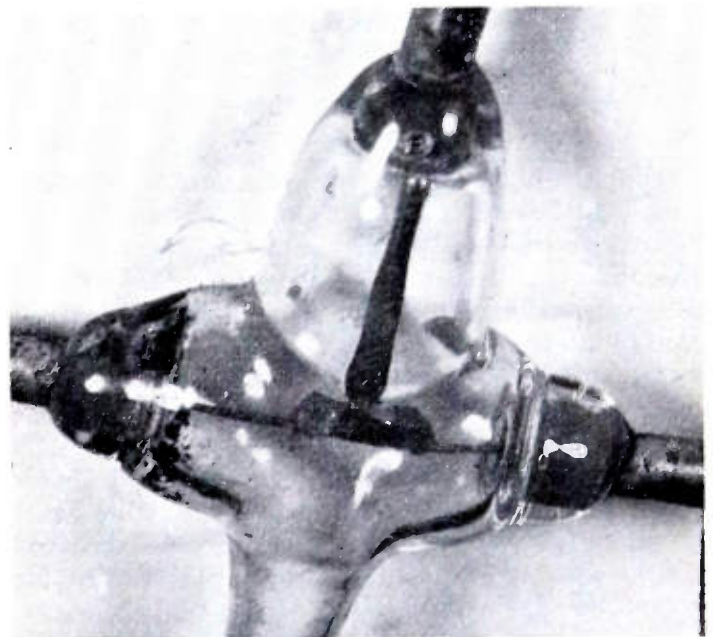
(b)

Fig. 2—Experimental models: (a) open, (b) sealed. Actual size of crystal: 0.3×0.08×0.08 cm³.

impedance of the device $\partial I_a/\partial U_a$ decreases and the frequency response shows a cutoff in the audio-frequency range. Devices of this kind, which are easy to reproduce and measure, shall be the subject of our discussion.

THE DC CHARACTERISTIC

The sign of G_m depends on the position S (see Fig. 1) of the control electrode. Fig. 3 illustrates the experi-

mental evidence. If the control electrode F is on the n-side of the junction, I_a increases pronouncedly if U_g becomes more negative. Positive increments of U_g are needed to produce a current increase if F is on the p-side of the junction. If F is positioned directly above the center, the two effects superimpose and a characteristic with a minimum occurs, as often observed with point-contact fieldistors.[1]

the polar moment of the liquid (see Figs. 6, 7, to be discussed later).

From Fig. 4 a considerable hysteresis effect becomes apparent, which is not surprising for a mechanism in which the orientation of polar molecules close to a surface with "surface states" and "traps" plays a role. I_a decreases more rapidly with decreasing control voltages than it increases with increasing control voltages.

Fig. 3—Controlling action versus position of field electrode F.
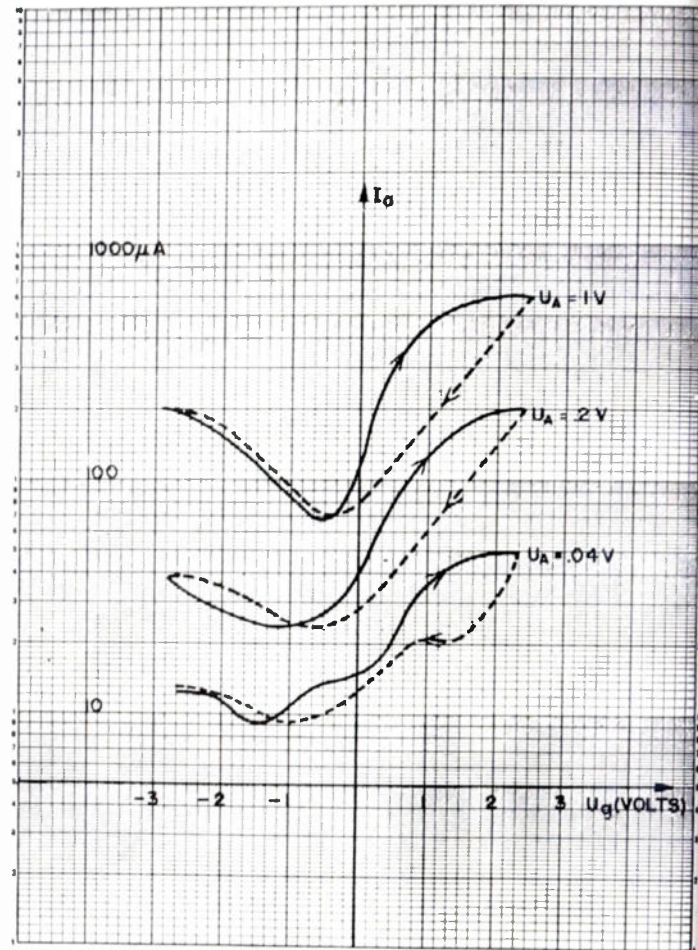
Fig. 4—I_a versus U_g diagrams for various "anode" voltages $U_a (L = 0.08, B = 0.003$ cm; $d = 0.0005$ cm).

The order of magnitude of the effect can be estimated from Fig. 4, a semilogarithmic plot of I_a versus U_g, with U_a as a parameter. The figure shows that for high positive and negative control voltages, I_a reaches different saturation values. These and the minimum current are plotted in Fig. 5 for a similar fieldistor in the form of a logarithmic I_a-U_a diagram. It shows that the dc impedance of the device can be changed by more than one order of magnitude.

The optimum G_m is directly proportional to the linear length L of the junction. The model measured in Fig. 4 gives four times the G_m values when, instead of a control electrode on one side ($L = 0.08$ cm), a ribbon extending around the crystal block ($L = 0.32$ cm) is used.

G_m is approximately inversely proportional to the electrode distance d. (See the low-frequency values in Fig. 8, which will be discussed later.) This means that the controlling effect is proportional to the controlling fieldstrength ("fieldistor").

G_m, furthermore, depends on the polar moment of the liquid introduced. In general, G_m is the larger, and the output (and input) impedance the smaller, the higher

Fig. 5—Junction voltage U_a versus junction current I_a, showing saturation values I_{sat+} for large positive and I_{sat-} for large negative biases as well as minimum current I_{min}.

[1] O. M. Stuetzer, "A crystal amplifier with high input impedance," PROC. I.R.E., vol. 38, pp. 868–871; August, 1950.

For high controlling (and driving) voltages, elec-
olytic action which causes irreproducible changes oc-
irs. If all voltages are kept below 1.5 volts, however,
irefully manufactured devices stay constant within
15 per cent over months.

In a junction without any control electrode attached,

each part has just the potential of the right sign to pro-
duce a current increase in the other part ("self-fieldis-
toring"). This results in marked changes of the back
characteristic (see Fig. 6). The plot indicates how im-
portant complete absence of polar surface molecules is
for high back-voltage rectifiers.

AC BEHAVIOR

With a suitable model the frequency response for an
ac signal over the dc bias was measured for various di-
electric liquids (Fig. 7). The curves could be reproduced
within 15 per cent after one month. The frequency re-
sponse depends, to a certain degree, on the dc bias ap-
plied, as is indicated for n-butyl-phthalate.

Fig. 8 shows the frequency response for a nitroben-
zene activated fieldistor with the distance d as a param-
eter. Figs. 7 and 8 are for measuring models only; in ac-
tual devices L was larger, leading to transconductance
values of several thousand μmhos at low frequencies.

The observed frequency behavior is reminiscent of
that of two RC combinations in series (see Fig. 9). That
the cutoff curves shown in Figs. 7 and 8 do not slope off
exactly inversely proportional to frequency can be ex-
plained by the fact that in reality we will have to as-
sume many of such combinations in parallel.
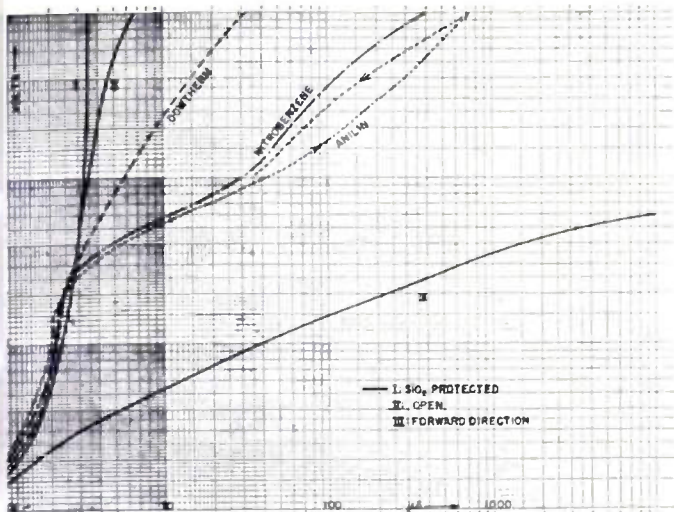
Fig. 6—Deterioration of rectifying junction characteristics
due to fieldistor effect.
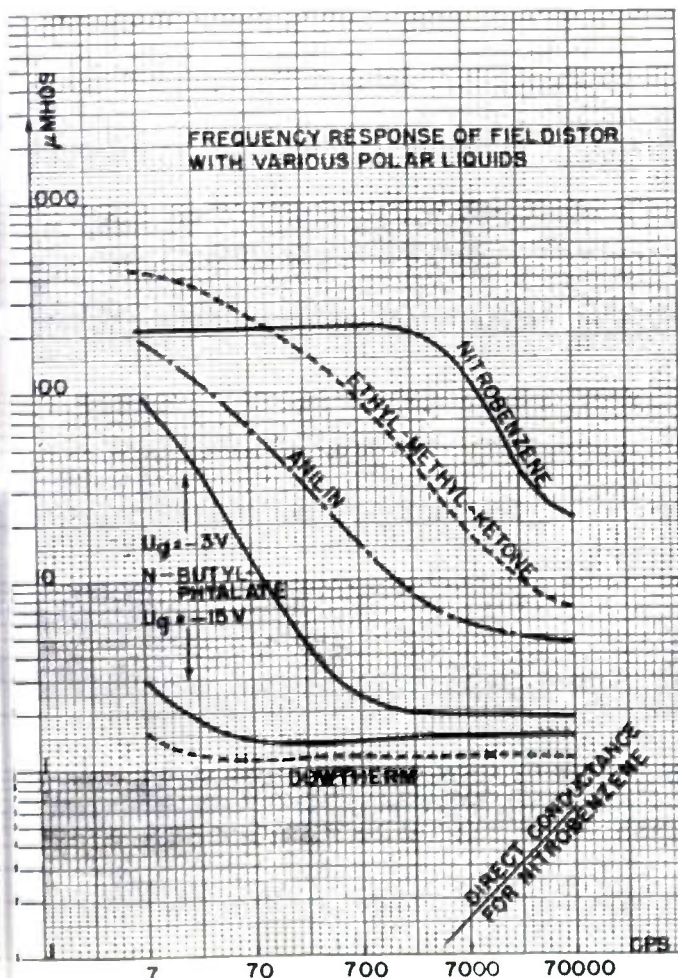
Fig. 7—Dependence of G_m on frequency for fieldistor model ($L=0.12$,
$B=0.003$, $d=0.0002$ cm, for various polar liquids.
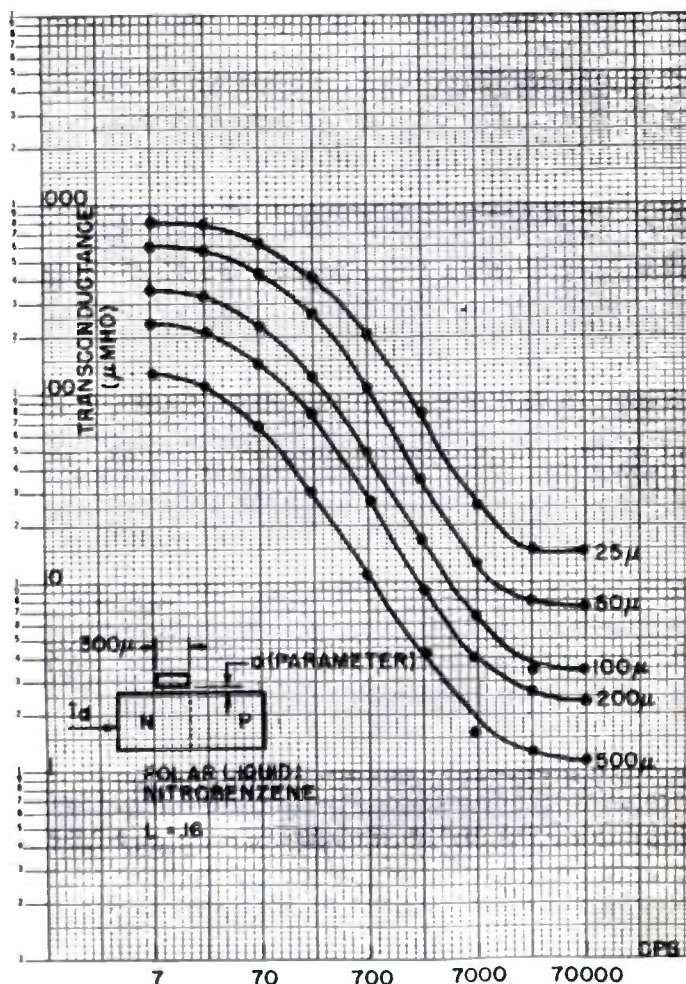
Fig. 8—Frequency response versus distance d.

The output impedance of the device is naturally dependent on the bias (compare Fig. 5). At points of maximum G_m, several thousand ohms are found with liquids of high polarity, several ten thousand with liquids of lower polar moment (compare Fig. 6).
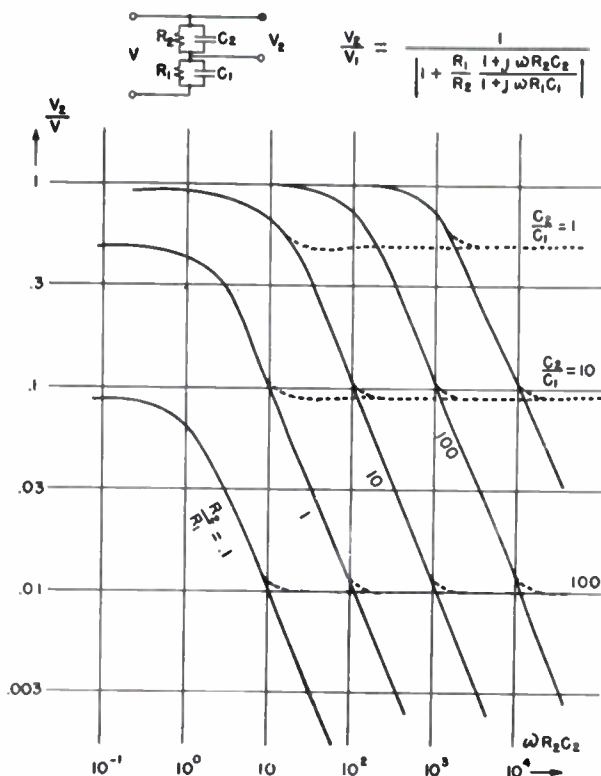
Fig. 9—Frequency characteristics of two RC combinations in series.

The input impedance of the devices is roughly a resistance of about 10 megohms paralleled by capacitance $\epsilon BL/d$, which turns out to be of the order of a few micromicrofarads. (See plot of "direct conductivity" for nitrobenzene sample in Fig. 7.) Leakages around the controlling system proper, and circuit and build-up capacitances, are mostly of higher magnitude, unless great care is taken. Input impedance is discussed more fully later in connection with the interpretation of Fig. 9.

Using the above figures, a power gain $R_{in}R_{out}G_m^2/4$ of about 40 db is obtained for low frequencies.

Only tentative data are available at present for the noise behavior. Unfortunately, the ideal noise properties of a junction are lost when a polar liquid is applied to its surface. Noise figures at operation points with high G_m are very similar to those of the point-contact diodes (60 to 80 db at 1000 cps.).

THEORY OF OPERATION

For a sketch of a theoretical explanation, we shall discuss the n-side of the junction. On the p-side, corresponding events happen, with all signs reversed.

On the surface of the junction we have to assume a layer of surface charges, (in our example, negative), which is counterbalanced by a space-charge layer of

ionized donors of about 10^{-4} cm depth. The left diagram in Fig. 10 represents that configuration and gives a rough plot of the electron energy ϕ_e in the conduction band and the hole energy ϕ_h in the filled band. The number of carriers present varies inversely to exp ϕ/kT.
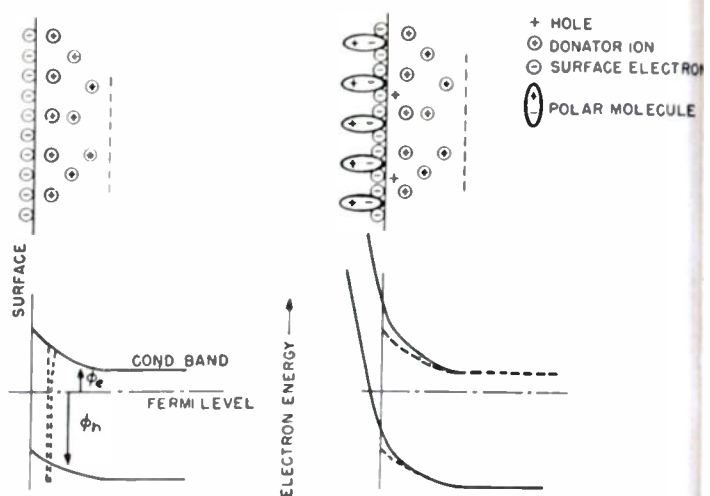
Fig. 10—Idealized model of surface on n-side of junction, with and without oriented dipole layer, showing electron and hole energies.

If a positive charge is induced into the system, that is, if the control electrode F (see Fig. 3) is negatively biased, the conduction electrons are repelled, the space-charge layer increased, and the energy bands raised a little.[2] A slight decrease in conductivity of the sample should be expected, and is observed with clean surfaces.

With a polar liquid present, however, a rather pronounced increase of conductivity is the case. We have to assume that the outside field raises the energy bands so much that, close to the surface, ϕ_h becomes smaller than ϕ_e, leading to the creation of excess holes. (Fig. 10 right.) The underlying physical mechanism is the repelling forces of the negative surface charges tearing valence electrons out of the lattice atoms. Experiments similar to the one described by Shockley, Pearson, and Haynes[3] indicate minority carrier conduction.

It is essential for our effect that the minority carriers originating in this way be produced on that side of the junction where the existing fields in the junction underneath the surface (see Fig. 3) help them cross the junction barrier. That is one reason why the arrangement works rather effectively.

Fig. 10 (right, top) shows a layer of oriented dipole molecules on a surface. The average potential rise in such a dipole layer is several volts[4] (about 5 volts for nitrobenzene). Close to one of the charged molecules, however, much higher rises are present, similar to the

[2] W. Shockley and G. L. Pearson, "Modulation of conduction of thin films of semiconductors," *Phys. Rev.*, vol. 74, p. 232; July, 1948.
[3] W. Shockley, G. L. Pearson, and J. R. Haynes, "Hole injection in germanium," *Bell Sys. Tech. Jour.*, vol. XXVIII, pp. 344–366; July, 1949.
[4] J. H. DeBoer, "Electron Emission and Adsorption Phenomena," Macmillan Co., New York, N. Y., p. 50, ff.; 1935.

tential fall indicated for a donor ion on the left side of Fig. 10. It is no surprise that the interaction with the surface atoms of the high fields creating these potentials raises the average potential close to the surface inside the crystal by a few tenths of a volt. (Potential rises of the order of magnitude were indicated in experiments by using a suitable probe arrangement and a high impedance electrometer.)

Fig. 11 shows the customary energy diagram for the junction in equilibrium. The carrier densities associated with the energies are plotted on the left side. It can be seen that a few tenths of potential rise brings substantial minority carrier density on both surfaces.

POTENTIAL AND CARRIER DENSITY DISTRIBUTION IN P-N JUNCTION
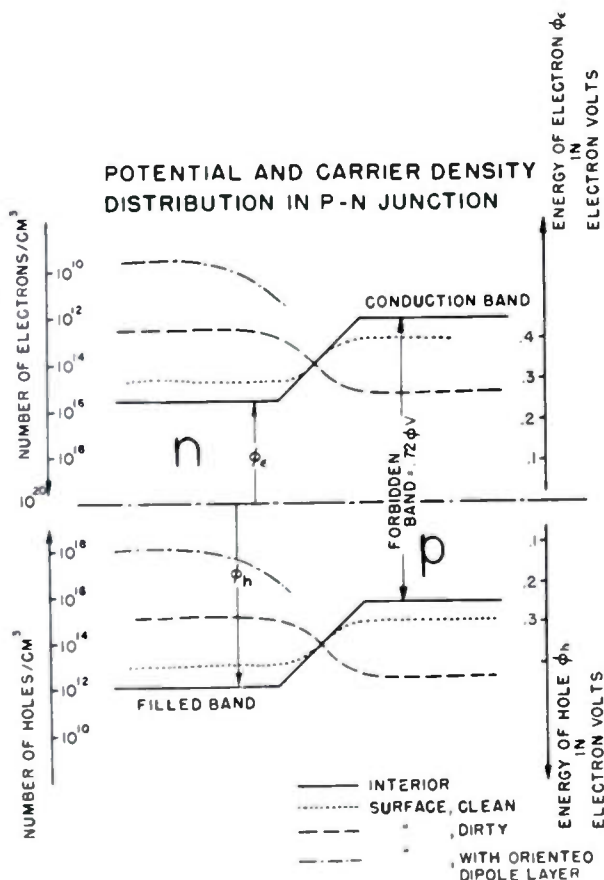
Fig. 11—Electron energy and carrier density distribution along junction. (Interior curves for conductivity of about 2 ohm⁻¹ cm⁻¹.)

The detailed theoretical description of the surface layer will be rather complex. The potential in perpendicular to the surface no longer varies parabolically, but increases much faster (until degeneracy occurs). The potential along the axis of the device can not be assumed to be independent of the distance from the surface. The layer of high generation and high recombination will affect the mobility close to the surface.

Roughly, a density of $10^{18}/cm^3$ directly on the surface, leading with an approximated potential distribution to an area density of $10^{12}/cm^2$, can be assumed. With a mobility of 10^3 cm²/volt sec and an accelerating field of 10 v/cm, the order of magnitude of the effect can be explained.

The high current densities (10^3 amp/cm²) on the surface will be partly responsible for the high noise of the device.

We can explain the saturation properties of the device (see Figs. 4 and 5) by assuming that all the microdipoles are then oriented.

To analyze the frequency behavior of the device, we shall use the circuit model of Fig. 9. V_2 is that part of the (ac) control voltage V, which induces the surface changes necessary for our control effect. Analysis shows that we have to attribute to C_2 a value of the order of 10^{-10} F. This points to the existence of a high capacitance layer on the surface—probably a gas layer similar to the one in electrolytic condensers—with a capacity of about $10^{-8} F/cm^2$. C_1, which is one to two orders of magnitude smaller, then represents the capacitance of the control electrode (in series with the capacitance of the conductive surface layer). R_2 in this picture would be the very big (several megohms) resistance of the (gas) layer and R_1 the resistance of the control liquid in series with that of the conductive surface layer. R_1 will certainly change with bias, and C_2 may very well do so. This explains the dependence of the frequency characteristic on dc bias.

CONCLUSIONS

A crystal amplifier with about 10-megohms input and several thousand ohms output impedance, a transconductance of about a thousand micromhos, and a noise figure of the order of 70 db was described. A frequency cutoff in the audio range will severely limit the application of the present form of the device.

The investigations add to our knowledge of surface phenomena. They explain why the presence of molecules with a polar moment on the surface of a diode damages its back voltage properties.

The Control of Frequency Response and Stability of Point-Contact Transistors*

B. N. SLADE†, MEMBER, IRE

Summary—The frequency response and stability of point-contact transistors are determined to a large degree by control of the point-contact spacing and germanium resistivity. Stability is particularly important in amplifiers in which the impedances of the emitter and collector circuits are very small in the frequency range in which the transistor is designed to operate. Satisfactory stability has been obtained with developmental transistors having a frequency cutoff (3-db drop in the current amplification factor, alpha) ranging from 10 to 30 mc. These transistors operate under approximately the same dc bias conditions used with lower-frequency transistors, and have an average power gain of approximately 20 db. By means of the methods outlined, transistors which oscillate at frequencies as high as 300 mc have been made.

INTRODUCTION

CHOICE OF RESISTIVITY and spacing can be discussed in terms of theoretical considerations involving the transit time of the holes, the variation of equivalent base resistance with point spacing, and the effect of equivalent base resistance on transistor stability.

Shockley[1] has expressed the transit time of the hole carriers which travel from emitter to collector as $\tau = S^3\sigma/\mu I_e$, where τ is the transit time in seconds, μ is the mobility of the holes in centimeters squared per volt-second, I_e is the emitter current in amperes, σ is the germanium conductivity in reciprocal ohm-centimeters, and S is the spacing in centimeters between the point contacts. The transit time has an approximately inverse relationship to frequency response. Thus, the frequency response would be expected to vary inversely as the cube of the point spacing if other conditions were maintained constant.

It has also been shown previously[2] that a decrease in the point spacing will increase the value of the equivalent base resistance, the resistance of that portion of the equivalent circuit of the transistor which is common to both the input and output circuits of the transistor. A large value of equivalent base resistance, however, contributes to transistor instability. The following expression[3] represents a criterion for transistor stability:

$$\frac{R_e}{R_b} + \frac{R_e}{R_c} + 1 > \frac{r_m}{R_c},$$

where $R_e = r_e +$ external emitter resistance, $R_b = r_b +$ external base resistance, and $R_c = r_c +$ external collector resistance. The terms r_e, r_b, r_c, and r_m represent the emitter, base, collector, and transfer resistances, respectively, as given in the T-network equivalent circuit of the transistor. The term r_m/r_c is equal to the current amplification factor, alpha.

For transistor stability, when no external resistance or impedances are in series with any of the three terminals of the transistor,

$$\left(\frac{r_e}{r_b} + \frac{r_e}{r_c} + 1 \right)$$

must be greater than r_m/r_c. A large value of equivalent base resistance, however, may cause the expression in parenthesis to become smaller than r_m/r_c, depending on the values of the other resistances, and may cause the transistor to become unstable and to oscillate, unless the value of alpha is unity or less. Most point-contact transistors have an alpha of 2 or greater.

MEASUREMENTS

The value of equivalent base resistance can be kept small at small point spacings, however, by decreasing the value of the germanium resistivity. Fig. 1 is a curve
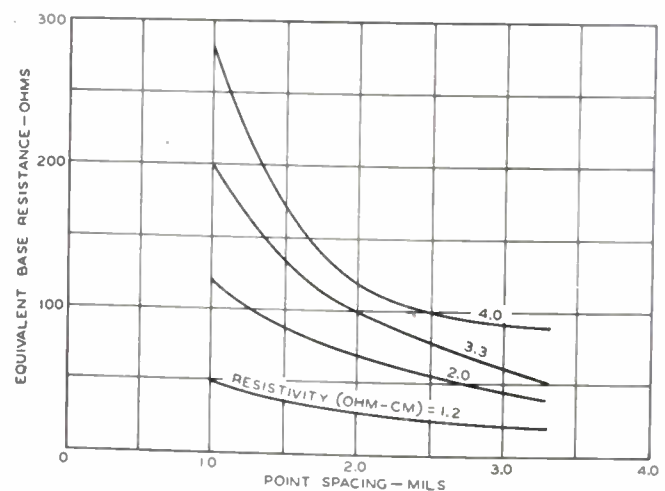
Fig. 1—Effect of variation in point spacing and germanium resistivity on equivalent base resistance.

showing the variation of the equivalent base resistance with point-contact spacing for several germanium specimens ranging in resistivity from 1.2 ohm-centimeters to 4.0 ohm-centimeters. At the higher values of resistivity, the equivalent base resistance increases rapidly at spacings less than 2.0 mils. For the specimens of germanium

[1] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y.; 1950.
[2] V. Bardeen and W. H. Brattain, "Physical principles involved in transistors action," *Phys. Rev.*, vol. 75, pp. 1208–1225; April 15, 1949.
[3] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, pp. 317–401; July, 1949.

having resistivities of 2 ohm-centimeters or less, the value of the equivalent base resistance approaches a more linear relationship with the point spacing. Furthermore, the value of the equivalent base resistance is much smaller at the low resistivity values for a given point spacing.

It would appear from the expression for transit time that, as the germanium resistivity increases, the fre-

manium resistivity, it is possible to achieve higher frequency responses without sacrificing stability.

By the use of the above methods, stable transistors having frequency cutoffs in the range of 30 mc with low-frequency power gains of 20 db and values of equivalent base resistance of 100 ohms and less have been made. A typical frequency response curve for such a transistor is given in Fig. 3.
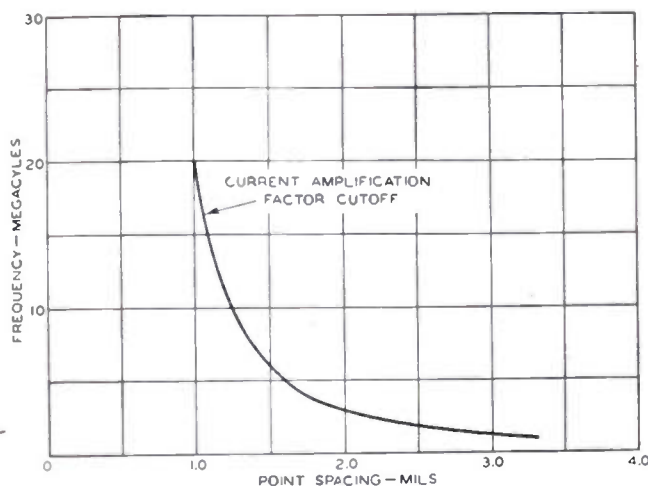
Fig. 2—Effect of variation in point spacing on frequency response.
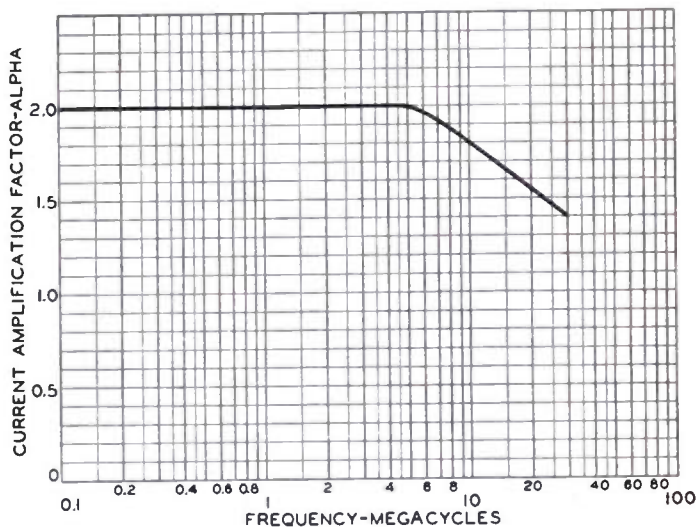
Fig. 3—Frequency-response characteristic of developmental point-contact transistor.

quency response also increases. Actual measurement, however, indicated that an increase in resistivity from 1.2 to 4 ohm-centimeters had little effect on the frequency response. Fig. 2 is a curve showing the variation of frequency cutoff (3 db down in the current amplification factor) with point spacing for germanium specimens having resistivities within the range of 1.2 to 4 ohm-centimeters. The curve shows that at spacings less than 2 mils, the high-frequency response improves rapidly. In fact, the curve follows quite closely a cube function of frequency cutoff with spacing.

Generally, values of equivalent base resistance of 100 ohms or less will assure transistor stability. In order to obtain these low values and, at the same time, maintain good frequency response, it is necessary to use low-resistivity germanium and smaller contact spacings. As shown in Fig. 1, if germanium having a resistivity of 3.3 ohm-centimeters were used, a point-contact spacing of 20 mils would result in a transistor having an equivalent base resistance of 100 ohms. From Fig. 2, it can be seen that the frequency cutoff of such a transistor would be approximately 3.3 mc. If germanium having a resistivity of 2.0 ohm-centimeters were used, an equivalent base resistance of 100 ohms could be obtained with a spacing of 1.3 mills; the frequency cutoff would be approximately 11.0 mc. If germanium having a resistivity of 1.2 ohm-centimeters were used, an equivalent base resistance of 50 ohms could be obtained with 1-mil spacing; the frequency cutoff would be approximately 20 mc. Thus, by a careful choice of point spacing and ger-

Visual Demonstration of Frequency Response

An interesting visual demonstration of the frequency response of these transistors is given in Fig. 4. The transistor was used as a video amplifier in a flying-spot scanning system. A wedge test pattern was used to indicate the response of the circuit with and without the transistor. Fig. 4(a) shows the test pattern with the transistor by-passed. The narrowest portion of the wedge indicates a frequency of approximately 16 mc. Fig. 4(b) shows the test pattern with a transistor having a cutoff at 30 mc in the circuit; no impairment of the resolution of the pattern can be seen. Fig. 4(c) shows the test pattern with a transistor having a 4-mc cutoff frequency. Considerable resolution is lost when this transistor is used as a video amplifier. Fig. 4(d) shows the test pattern when a wide-spaced transistor having a 100-kc cycle cutoff is used. The resolution of this pattern is very poor. Figs. 5(a) through (d) show oscillograms over an 18-mc bandwidth for the circuit illustrated by the test patterns in Fig. 4. The gain obtained from the transistors in the circuit used for Fig. 4 was only the 2-to-1 current amplification of the transistors because the test was made to give a visual indication of the frequency response of the current-amplification factor. For the transistor illustrated in Figs. 4(b) and 5(b), however, a voltage gain of approximately 25 to 1 could be obtained over the 18-mc bandwidth without any impairment in the quality of the picture.

Fig. 4—Visual demonstration of frequency response of transistors in a video amplifier system. (a) Transistor by-passed. (b) Transistor having 30-mc cutoff. (c) Transistor having 4-mc cutoff. (d) Transistor having 100-kc cutoff.

VHF OSCILLATOR TRANSISTORS

The design methods described in this paper have also been used to obtain transistors capable of operating at frequencies considerably higher than 30 mc. A number of these oscillate at frequencies above 100 mc, with a few oscillating above 200 mc and one achieving 300 mc.

Fig. 5—Oscillograms of frequency response of transistors in video-amplifier system. (a) Transistor by-passed. (b) Transistor having 30-mc cutoff. (c) Transistor having 4-mc cutoff. (d) Transistor having 100-kc cutoff.

TRANSISTORS FOR SWITCHING CIRCUITS

Point-contact transistors which are designed for use in switching circuits require a higher current gain than those used for radio-frequency amplifier devices, and a certain degree of operational instability is, therefore, desirable. In most cases higher values of equivalent base resistance are permissible. Values of resistivity and point spacing may be selected in accordance with the requirements of the applications.

OTHER DESIGN CONSIDERATIONS

Careful selection of point-contact spacings and germanium-crystal resistivity must be made to obtain transistor characteristics which are desirable for either type of application. Other factors, such as the dependence of transistor characteristics upon ambient temperature changes, allowable collector dissipations, and values of r_e and r_c must also be considered in transistor design since these are also functions, to some degree, of the resistivity of the germanium. Factors other than point spacing and germanium resistivity may well have a significant effect upon frequency response and equivalent base resistance. The data presented in this paper are empirical and are not intended to provide an exact design formula. However, the data indicate the significance of the control of point spacing and germanium resistivity in the design of point-contact transistors.

ACKNOWLEDGMENTS

Low-Drain Transistor Audio Oscillator*

D. E. THOMAS†, SENIOR MEMBER, IRE

Summary—A 9-element transistor audio oscillator is described. This oscillator operates with relatively low drain from a single 6-volt battery. The oscillator gives reliable performance with an output uniform to approximately ±1 db with substantially all type 1768 point-contact transistors and without any circuit element adjustment required for variation in transistor parameters from unit to unit or with transistor ambient temperature.

INTRODUCTION

MANY TYPES OF EQUIPMENT must operate at remote locations from power which is provided locally. Often the use of batteries is indicated, and for long service the power-supply drain must be minimized. In very early stages of transistor development, it was recognized that transistors offer new and attractive possibilities in this field since they require no heater power and can operate with quite modest direct electrode power. One such possibility is a low-drain audio oscillator powered by batteries and designed to provide carrier power to a magnetic amplifier. This paper describes a transistor oscillator designed for this application. The transistor used in the oscillator is Western Electric A1768 point-contact transistor, developed by L. B. Valdes of Bell Telephone Laboratories.

The final oscillator circuit developed proved to be extremely simple and very reliable. Since this circuit, or similar circuits, should prove useful for other applications in the communications and electronics field, complete technical information on the oscillator is being presented. The paper gives the requirements on the oscillator, describes the final circuit and circuit element functions, presents complete circuit performance data, and gives an elementary design theory for a "first-order"[1] low-drain oscillator developed to meet the specific requirements given below. It also gives a brief discussion of the broad general reasons for the choice of circuit configuration.

REQUIREMENTS

The following are the specific requirements on the transistor oscillator:

Battery supply voltage: Nominal 6 volts; 12 volts optional. Range, +6 per cent, −15 per cent.

Maximum power drain from battery: Power drain averaged over operating life not to exceed 35 mw.

Frequency of oscillator: Single fixed frequency of 130 cps.

Frequency drift: ±7 per cent over total life.

[1] By a "first-order" circuit is meant one which has been carried to a stage of development where feasibility for the application has been fully demonstrated.

Output voltage: 8.5 volts rms at nominal battery voltage with 20-per cent permissible distortion across an 11,000-ohm resistance. 11,000 ohms is approximately equal to the equivalent shunt resistance of the linear impedance characterizing the magnetic amplifier load at the fundamental frequency.

Output impedance: To be determined by power transfer considerations at the fundamental frequency. To be as low as possible at all harmonics of the fundamental.

Stability: A variation of rms output voltage of the same percentage as the battery voltage over the useful life of the battery is permitted.

In planning the program for the development of a device and circuit to meet the above requirements, additional circuit and device objectives were formulated as follows: (a) to obtain a circuit which would meet requirements for substantially all transistors coded for the application, with no circuit element adjustments required to compensate for variations in transistor electrical parameters from unit to unit, or for variations of a particular unit with operating temperature; (b) to obtain a characterization of the transistor finally designed for application in terms of measurable electrical parameters which would assure that substantially all units chosen on the basis of this characterization would meet the requirements when operated in the oscillator circuit without recourse to any go-no go circuit test.

During the course of development work, several additional requirements arose on the oscillator circuit over and above those stated in the formal requirements already given. These had considerable influence on the design and performance of the circuit, as will be apparent from the discussion to follow.

The first and most important requirement was that the oscillator be self starting at any point in the required operating temperature range. Secondly, the final space allotment for the oscillator proved to be less by almost an order of magnitude than the space assigned to the oscillator in the early stages of the project. Considerable sacrifice in circuit efficiency resulted from the reduction in transformer size which was necessary to accomplish the required space reduction.

The final change in requirements resulted from a change in the oscillator load after the final oscillator circuit had been frozen. This resulted from the fact that the shunt-resistance component of the linear impedance characterizing the magnetic amplifier load at the fundamental frequency proved to be 21,000 ohms for the final magnetic amplifier rather than the 11,000 ohms given in the original requirements. The output voltage requirement was not changed, however. This, of course, was equivalent to lowering the required power output of the oscillator.

Fig. 4—Visual demonstration of frequency response of transistors in a video amplifier system. (a) Transistor by-passed. (b) Transistor having 30-mc cutoff. (c) Transistor having 4-mc cutoff. (d) Transistor having 100-kc cutoff.

VHF Oscillator Transistors

The design methods described in this paper have also been used to obtain transistors capable of operating at frequencies considerably higher than 30 mc. A number of these oscillate at frequencies above 100 mc, with a few oscillating above 200 mc and one achieving 300 mc.

Fig. 5—Oscillograms of frequency response of transistors in video-amplifier system. (a) Transistor by-passed. (b) Transistor having 30-mc cutoff. (c) Transistor having 4-mc cutoff. (d) Transistor having 100-kc cutoff.

Transistors for Switching Circuits

Point-contact transistors which are designed for use in switching circuits require a higher current gain than those used for radio-frequency amplifier devices, and a

certain degree of operational instability is, therefore, desirable. In most cases higher values of equivalent base resistance are permissible. Values of resistivity and point spacing may be selected in accordance with the requirements of the applications.

Other Design Considerations

Careful selection of point-contact spacings and germanium-crystal resistivity must be made to obtain transistor characteristics which are desirable for either type of application. Other factors, such as the dependence of transistor characteristics upon ambient temperature changes, allowable collector dissipations, and values of r_e and r_c must also be considered in transistor design since these are also functions, to some degree, of the resistivity of the germanium. Factors other than point spacing and germanium resistivity may well have a significant effect upon frequency response and equivalent base resistance. The data presented in this paper are empirical and are not intended to provide an exact design formula. However, the data indicate the significance of the control of point spacing and germanium resistivity in the design of point-contact transistors.

Acknowledgments

Low-Drain Transistor Audio Oscillator*

D. E. THOMAS†, SENIOR MEMBER, IRE

Summary—A 9-element transistor audio oscillator is described. This oscillator operates with relatively low drain from a single 6-volt battery. The oscillator gives reliable performance with an output uniform to approximately ±1 db with substantially all type 1768 point-contact transistors and without any circuit element adjustment required for variation in transistor parameters from unit to unit or with transistor ambient temperature.

INTRODUCTION

MANY TYPES OF EQUIPMENT must operate at remote locations from power which is provided locally. Often the use of batteries is indicated, and for long service the power-supply drain must be minimized. In very early stages of transistor development, it was recognized that transistors offer new and attractive possibilities in this field since they require no heater power and can operate with quite modest direct electrode power. One such possibility is a low-drain audio oscillator powered by batteries and designed to provide carrier power to a magnetic amplifier. This paper describes a transistor oscillator designed for this application. The transistor used in the oscillator is Western Electric A1768 point-contact transistor, developed by L. B. Valdes of Bell Telephone Laboratories.

The final oscillator circuit developed proved to be extremely simple and very reliable. Since this circuit, or similar circuits, should prove useful for other applications in the communications and electronics field, complete technical information on the oscillator is being presented. The paper gives the requirements on the oscillator, describes the final circuit and circuit element functions, presents complete circuit performance data, and gives an elementary design theory for a "first-order"[1] low-drain oscillator developed to meet the specific requirements given below. It also gives a brief discussion of the broad general reasons for the choice of circuit configuration.

REQUIREMENTS

The following are the specific requirements on the transistor oscillator:

Battery supply voltage: Nominal 6 volts; 12 volts optional. Range, +6 per cent, −15 per cent.

Maximum power drain from battery: Power drain averaged over operating life not to exceed 35 mw.

Frequency of oscillator: Single fixed frequency of 130 cps.

Frequency drift: ±7 per cent over total life.

Output voltage: 8.5 volts rms at nominal battery voltage with 20-per cent permissible distortion across an 11,000-ohm resistance. 11,000 ohms is approximately equal to the equivalent shunt resistance of the linear impedance characterizing the magnetic amplifier load at the fundamental frequency.

Output impedance: To be determined by power transfer considerations at the fundamental frequency. To be as low as possible at all harmonics of the fundamental.

Stability: A variation of rms output voltage of the same percentage as the battery voltage over the useful life of the battery is permitted.

In planning the program for the development of a device and circuit to meet the above requirements, additional circuit and device objectives were formulated as follows: (a) to obtain a circuit which would meet requirements for substantially all transistors coded for the application, with no circuit element adjustments required to compensate for variations in transistor electrical parameters from unit to unit, or for variations of a particular unit with operating temperature; (b) to obtain a characterization of the transistor finally designed for application in terms of measurable electrical parameters which would assure that substantially all units chosen on the basis of this characterization would meet the requirements when operated in the oscillator circuit without recourse to any go-no go circuit test.

During the course of development work, several additional requirements arose on the oscillator circuit over and above those stated in the formal requirements already given. These had considerable influence on the design and performance of the circuit, as will be apparent from the discussion to follow.

The first and most important requirement was that the oscillator be self starting at any point in the required operating temperature range. Secondly, the final space allotment for the oscillator proved to be less by almost an order of magnitude than the space assigned to the oscillator in the early stages of the project. Considerable sacrifice in circuit efficiency resulted from the reduction in transformer size which was necessary to accomplish the required space reduction.

The final change in requirements resulted from a change in the oscillator load after the final oscillator circuit had been frozen. This resulted from the fact that the shunt-resistance component of the linear impedance characterizing the magnetic amplifier load at the fundamental frequency proved to be 21,000 ohms for the final magnetic amplifier rather than the 11,000 ohms given in the original requirements. The output voltage requirement was not changed, however. This, of course, was equivalent to lowering the required power output of the oscillator.

[1] By a "first-order" circuit is meant one which has been carried to a stage of development where feasibility for the application has been fully demonstrated.

DESCRIPTION OF CIRCUIT AND CIRCUIT-ELEMENT FUNCTIONS

A schematic diagram of the circuit is shown in Fig. 1. Two views of the experimental packaged oscillator, on which all performance data to be presented in this paper
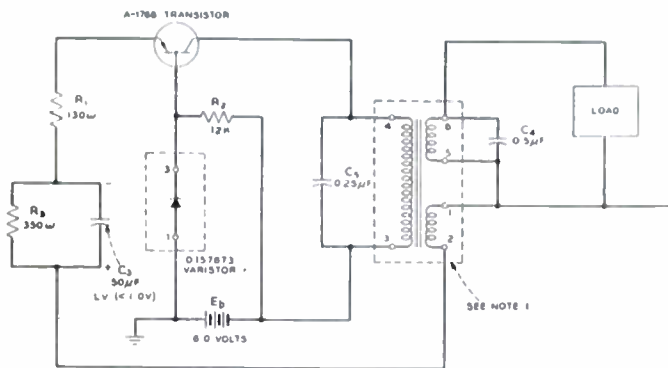
Fig. 1—Schematic diagram of low-drain audio oscillator.

Note 1. Three-winding transformer turns ratio = 1 : 2.8 : 9.6
　　　　　　　　　　　　　　　　　　　　　(1–2) (3–4) (5–6)

$L(5-6) = 3.2$ henrys
$Q(5-6)$ nonloaded $= 13.5$ at 120 cycles

were taken, are given in Fig. 2 and the various circuit elements are identified in terms of their designation on the schematic circuit. A brief discussion of the way in which the final choice of circuit configuration was decided upon follows:

A point-contact transistor was selected because it was known that these could be produced in sufficient quantity to meet any reasonable production requirements of the application for which the oscillator was being designed. The circuit operates on a Class C basis with an operating angle of approximately 150°. Class C was chosen because of its greater efficiency. Six volts was chosen in preference to 12 volts because at the required power output level greater collector conversion efficiency is obtainable at this voltage. For higher power output requirements a higher supply voltage should probably be used.

Transformer feedback was indicated for a number of reasons. In the first place, the load impedance is considerably higher than the optimum collector load impedance for maximum collector conversion efficiency; therefore, a step-up output transformer is called for. Secondly, since it is desirable that the oscillator present as low an impedance as possible to the load at all harmonics of the fundamental frequency, a parallel resonant circuit is indicated, and this is easily obtainable by using the load winding of the output transformer as the inductive element of this circuit. Finally, since a transformer is already called for on the basis of efficiency and impedance considerations, and since the emitter power necessary to maintain oscillations can be most efficiently transferred from the collector by a transformer, the case for a three-winding transformer is practically complete. Feedback winding of the transformer was put in the emitter circuit rather than in the base circuit to reduce possibility of unwanted parasitic oscillations.

Up to this point the reasoning has differed little from that used in the design of a vacuum-tube oscillator and,

in fact, the final circuit does closely suggest a single-tube transformer coupled feedback oscillator. However, here the analogy ends. It is required that the oscillator be operated from a single battery and still be self starting. In
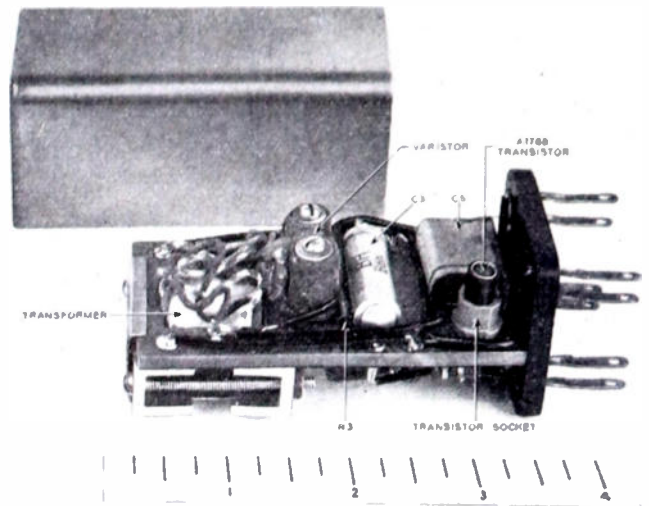
Fig. 2(a)—Experimental packaged low-drain audio oscillator.
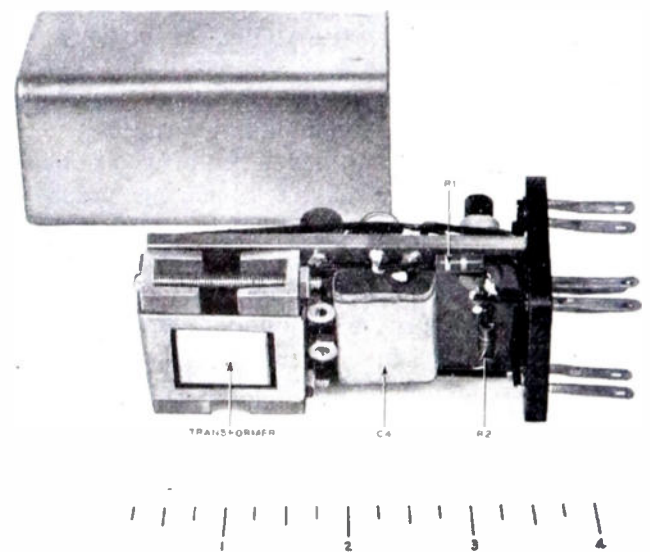
Fig. 2(b)—Experimental packaged low-drain audio oscillator.

a vacuum tube, plate current normally flows with no bias on the control electrode whereupon the tube is biased back negatively to the desired operating point by a cathode resistance. A transistor, on the other hand, requires that its control electrode be biased in a forward direction before substantial collector current flows. In a single-battery circuit some forward bias is provided by the base potential resulting from the collector current at zero emitter current flowing through the internal base resistance of the transistor. However, every effort is made to keep this current low. As a result, oscillations may not start when the collector battery is connected to the circuit unless an initial positive emitter current is provided which is sufficient in magnitude to assure the necessary transistor gain to start oscillation. Once oscillation is started, a self-biasing resistor in the emitter circuit operates to reduce the operating angle and in-

crease efficiency much the same as the tube cathode resistor.

The necessary initial current could have been supplied by inserting a resistance in the base circuit and providing a bleeder resistance from the negative battery to the base terminal. In very low-powered battery circuits this type biasing circuit results in the following difficulties: If the base resistance is made small enough to avoid an undue loss of dc collector potential under steady-state operating conditions, the bleeder current required is so high as to take too large a portion of the maximum permissible battery drain. On the other hand, if the bleeder current is held to a reasonable value from an efficiency standpoint, the base resistance required is so high that the available dc collector potential under steady-state conditions is reduced to a point where conversion efficiency suffers, dc stability problems are introduced, and the variation in transistor collector current with the normal spread of transistor parameters is magnified by the resultant dc positive feedback. As a result, the designer is faced with the three following alternatives, each of which is incompatible with meeting the requirements: 1. to accept poor stability in return for good efficiency; 2. to accept poor efficiency in return for good stability; 3. to abandon the single-battery approach and use two batteries.
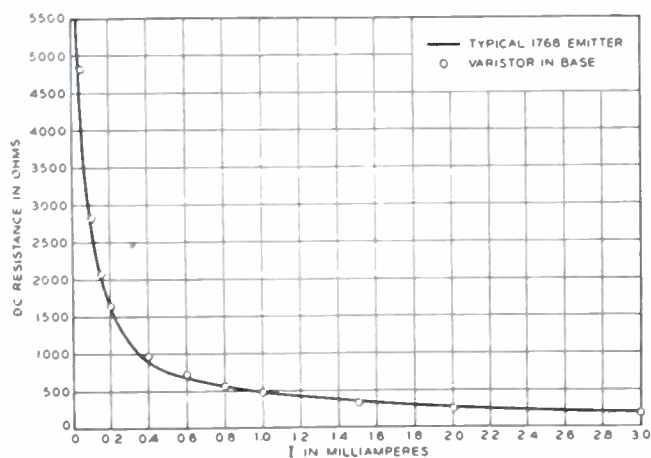
Fig. 3—DC resistance characteristics of transistor emitter and base circuit varistor.

All of the above problems were solved by the use of a varistor rather than a resistor in the base circuit. The varistor, having a negative slope dc resistance-versus-current characteristic, as does the emitter itself, provides the right type of base resistance to permit the maximum emitter current to flow for a given starting current provided by the bleeder. This is illustrated in Fig. 3, where the dc resistance-versus-current characteristic of the emitter of a typical 1768 transistor is shown along with the resistance characteristic of the diode used in the base. It is seen that the two characteristics are practically identical. The starting current provided by the bleeder will therefore divide approximately evenly between the emitter and base circuits. However, as the collector current builds up and the base current increases correspondingly, the varistor resistance drops

from its initially high value to a relatively low value and the dc stability problem is solved. Furthermore, since the voltage drop across the varistor rises only very slowly with increase in current, as illustrated in Fig. 4, the magnification of the normal collector current variation of different transistors due to dc positive base feedback is practically eliminated.
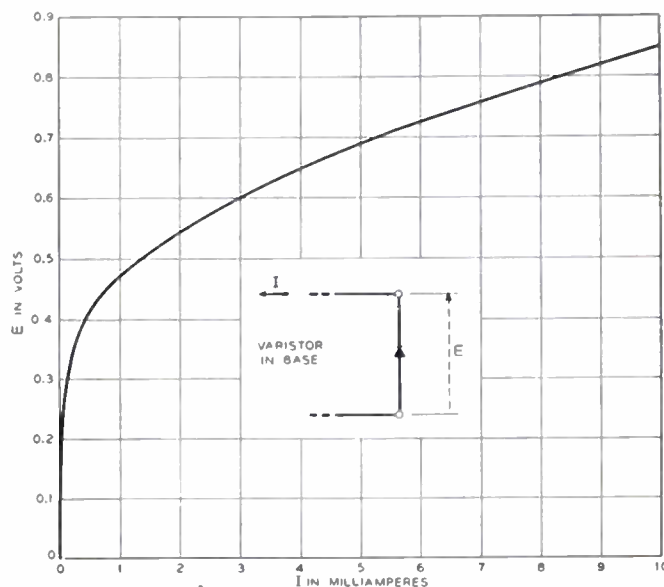
Fig. 4—Base circuit varistor dc-EI characteristic.

Even though the bias resulting from the necessity of introducing resistance in the base circuit to make oscillation self starting is reduced to a minimum by the use of a varistor, as discussed above, it still results in a net positive bias on the emitter under steady-state conditions. In order to get collector current conduction over less than half the operating cycle and thereby obtain Class C operation with its resultant increase in collector conversion efficiency, it is necessary that the net emitter bias be negative in order to bias the transistor beyond cutoff. The required negative bias is provided by the rectifying action of the emitter and the introduction in the emitter circuit of R_3 and C_3 in parallel. The required negative bias results from the rectified emitter current flowing in R_3.

The R_3, C_3 elements perform still another function. Some self-adjustment feature is needed to assure that the variation in output of the oscillator will not be too great with transistors having a reasonable range of operating parameters, and as a result of transistor parameter changes with temperature. The bias across R_3 is proportional to the oscillator output and varies in a way to expand and contract the operating angle, thus providing the required avc action. This avc action uses the battery voltage as its reference potential.

An oscillograph of the collector current versus time for a typical 1768 transistor is shown in Fig. 5(a). It is seen that collector current flows over appreciably less than half the total operating cycle and that the collector is cut off during the remainder of the cycle. The output voltage is shown in Fig. 5(b) and the approximation to

a pure sinusoidal wave is surprisingly good in view of the smallness of the output transformer providing the shunt inductance branch of the tank circuit, the low frequency, and the extreme nonlinearity of the load which is illustrated by the load current versus time shown in Fig. 5(c).
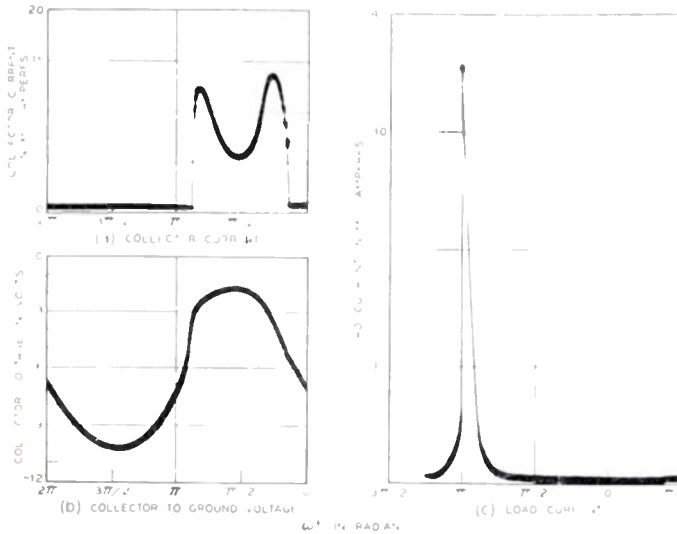
Fig. 5—Oscillator current and voltage wave shapes. (a) Collector current. (b) Collector-to-ground voltage. (c) Load current.

The functions of all elements in the circuit, except the 130 ohms in series with the emitter and the 0.25 mf across the collector winding of the transformer, have now been described. To describe their function it is necessary to return to the varistor in the base. This element
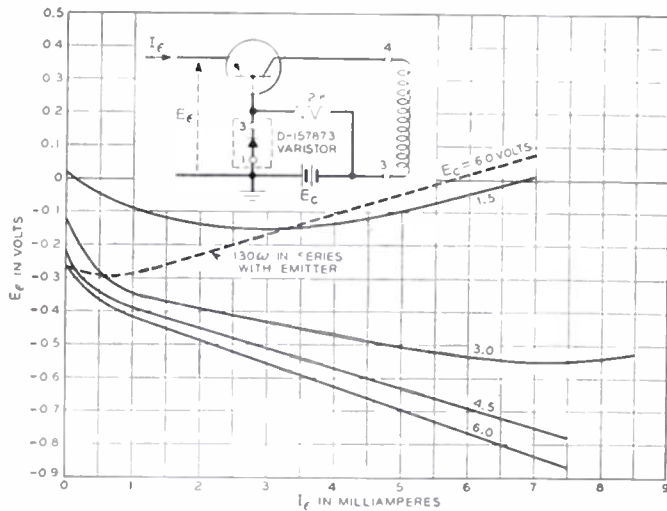
Fig. 6—Emitter input dc-EI characteristic.

not only solved the starting problem, but also made it possible to approximate the square wave of collector current which is desirable from the standpoint of the best Class C operation. To see how this happens, examine the voltage-current characteristic of the emitter-to-ground circuit, as shown in Fig. 6. Note that the slope of the characteristic at collector potentials above 3 volts is negative over a wide range of emitter current. This means that at the point on the positive half cycle of the sine wave of feedback voltage from the trans-

former where the algebraic sum of the instantaneous value of the feedback voltage and the negative bias voltage across R_3 equals the peak emitter voltage on the characteristic shown, the emitter current rises rapidly to a point where the collector current goes into saturation in a way similar to the action in transistor switching circuits. Without slowing down this action, however, the desired sine-wave oscillations may never commence under initial starting conditions since the transistor may either lock up in a conducting state or operate as a blocking oscillator. By placing a resistance, R_1, in the emitter circuit, so chosen that the net resistance is positive but small for all 1768 transistors, we obtain a sufficiently fast rise in emitter current and yet keep the circuit in a dc stable condition under which the desired fundamental ac oscillations can build up to a stable steady state. The dotted line in Fig. 6 shows the positive slope of the input voltage current characteristic at 6.0-volts collector voltage resulting from the introduction of the 130 ohms chosen as the value of R_1 in the emitter circuit. Note that it still takes only a few tenths of a volt change in feedback voltage to produce 6-or-7 mils emitter current, which is adequate to produce collector cur-
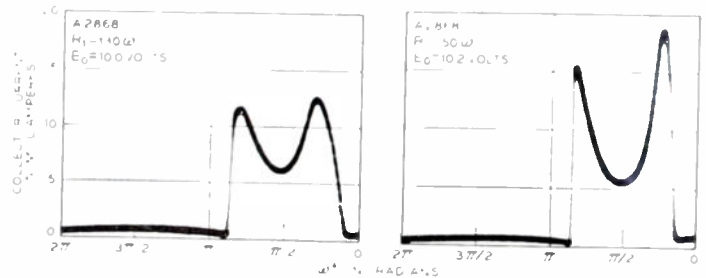
Fig. 7—Oscillographs showing effect of R_1 on collector current.

rent saturation. The resultant fast rise time in collector current, which is necessary for best Class C operation, is a direct consequence of the short-circuit instability inherent in a transistor with a high base resistance. In the low-drain oscillator this base resistance is provided externally by the varistor in the base for the reasons already discussed.

The effect of variation in R_1 on the operation of the circuit is illustrated in Fig. 7, where oscillographs of collector currents with R_1 equal to the normal value of 130 ohms and with R_1 reduced to 50 ohms are shown. The transistor chosen for this illustration was selected from the bottom of the distribution of transistors, considered on the basis of voltage output. This effect of R_1 on collector current demonstrates the desirability of keeping R_1 as small as possible consistent with the maintenance of satisfactory and stable operation over the entire range of variation of parameters of the 1768 transistor.

The function of the 0.25-mf condenser across the collector winding of the transformer is now apparent. This condenser provides the necessary low impedance to harmonics of the collector current required for a high-efficiency Class C oscillator. Without the condenser, the impedance presented to harmonics by the leakage reactance of the transformer would prevent the desired

short collector current rise time. Collector current cut-off occurs on the decreasing side of the positive half of the feedback voltage in a reversal of the action described above for transistor turn on.

From the above description of the operation of the circuit it is now apparent that the transistor is acting substantially as a switch which is turned on and off on the rising and decaying portions, respectively, of the positive half cycle of the voltage fed back to the emitter by the feedback winding of the output transformer. This results in a load curve quite different from that obtained in a Class C oscillator in which the collector is driven into overload only near the peak of the feedback voltage on the conducting half cycle.

This is illustrated in Fig. 8, which gives an oscillograph of the collector current versus the collector voltage over one complete cycle of operation. Load characteristic is,
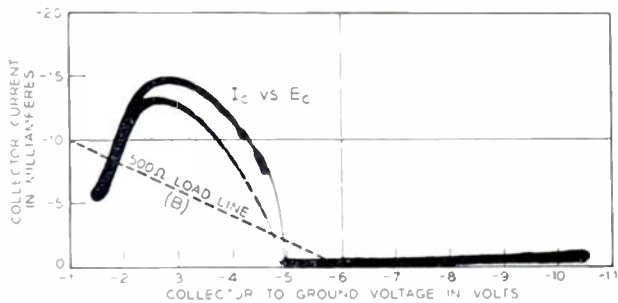
Fig. 8—Oscillograph of collector load curve.

in fact, something of a misnomer since the shape of the characteristic is unchanged by variation of the load resistance over an extremely wide range. The only change resulting from the change in load resistance is the extent to which the transistor is driven into voltage overload as indicated by the minimum value of collector current reached at the maximum positive swing of the collector voltage. The dotted line shows the Class C load characteristic for a 1,000-ohm load (operating angle is assumed close to 180°, in which case the slope is approximately $1/2 R_L$ or 500 ohms) when the emitter and the collector currents are approximately linearly related to the feedback voltage on the conducting half of the cycle.

The load which the oscillator was designed to drive is a magnetic amplifier which draws appreciable current only on the positive half cycle of the load voltage. As a consequence of this characteristic of the load, the poling of the load winding of the output transformer is critical and should be as shown in Fig. 1. If the poling is reversed from that shown, the output voltage of the oscillator is increased. However, the performance of the magnetic amplifier is nevertheless poorer because of the high impedance presented to the fundamental current flowing in the load circuit resulting from the collector current being cut off during the period when the load is conducting. It is therefore necessary, for best operation, that the output be so poled that the load and transistor collector are conducting on the same half cycle. This poling makes it unnecessary for the transformer to store high energy for half a cycle.

CIRCUIT PERFORMANCE

The operation of over five hundred type 1768 transistors was observed in various models of the oscillator. These units were coded as 1768 transistors from laboratory production of the Bell Telephone Laboratories at Murray Hill and from pilot production runs at the Allentown Plant of the Western Electric Company, strictly on the basis of measurements of the electrical parameters of the transistors and without any go-no go circuit tests. Based on a statistical analysis of all results, it can be safely said that at room temperature and 6-volts battery voltage, not more than 1 unit in 1,000 coded 1768 transistors will give an output across the specified magnetic amplifier load of less than 9.0 volts, which is 0.5 volt greater than the minimum requirement, and not more than 1 unit in 1,000 will draw more than 33 mw, which is less than the 35-mw maximum of the requirements.
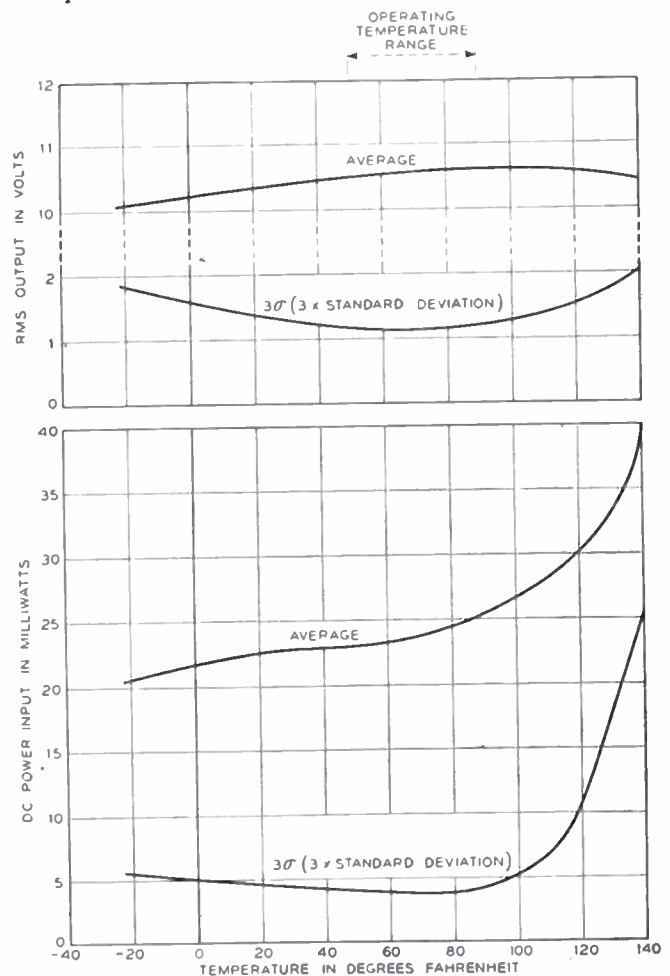
Fig. 9—Circuit-performance curves.

In order to determine the performance to be expected of preproduction Allentown units, 30 units were taken at random from this production and complete operating data taken as a function of temperature over the range from −22°F to +140°F with nominal battery voltage of 6.0 volts. These data were analyzed and the average and 3σ limits of both the output voltage to the magnetic amplifier load and the total dc input power were determined as a function of temperature. The results are given in Fig. 9. An examination of the output voltage

at the lowest temperature in the operating range, which is the most critical temperature from the output voltage standpoint, shows the average output voltage to be 10.5 volts and the 3σ limit 1.2 volts. Not more than 1 unit in 1,000 would therefore be expected to drop below 9.3 volts[2] which is 0.8 volt above the requirement at any temperature in the operating range. The most critical temperature from the power input standpoint is
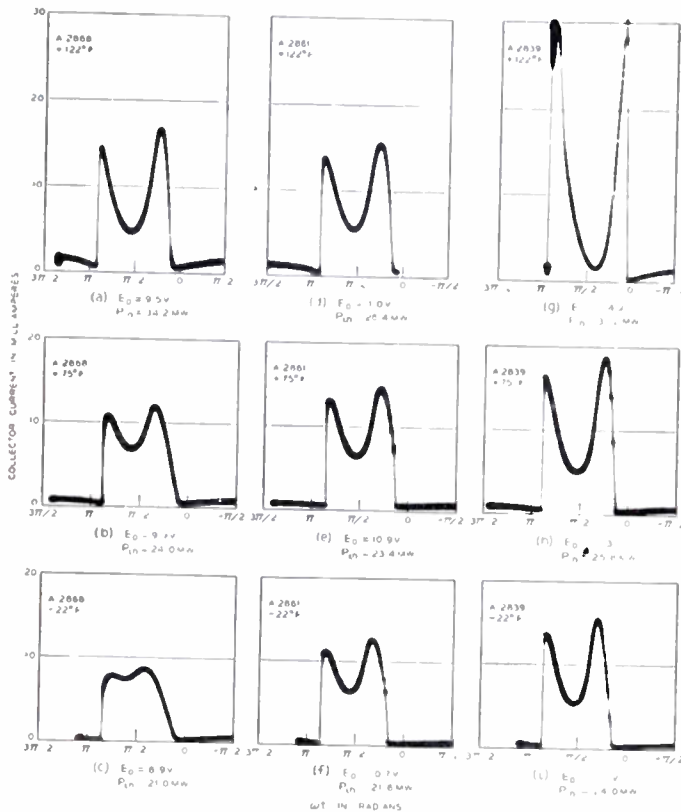
Fig. 10—Oscillographs of collector current for three different transistors at three different ambient temperatures.

at the top end of the operating range. Here the average power input is 25.2 mw and the 3σ limit 4.1 mw, so that not more than 1 unit in 1,000 can be expected to exceed 29.3 mw power drain at any temperature in the operating range, and this is well below the requirement. These results are even better than those quoted above for earlier developmental models of the 1768 transistor at room temperature and current preproduction models of the A1768 transistor are proving to give even slightly better performance than that shown in Fig. 9.

An interesting illustration of the action of the avc circuit in maintaining relatively uniform output was afforded by the temperature tests. Fig. 10 shows oscillographs of the collector current versus time at nominal battery voltage for three A1768 transistors at three different temperatures. One temperature was well below

[2] The statement made here and later on the number of units to be expected to fall outside an upper or lower 3σ limit is based on the assumption of a normal distribution. Experience with a large quantity of performance data on the oscillator has led to the conclusion that this assumption is a fairly good one. However, it would be more conservative to consider the numbers quoted as falling outside certain limits as order of magnitude rather than as absolute values, and that the true numbers falling outside these limits might therefore be two or even three times as large as a result of failure of the distribution to be a truly normal one.

the operating range, one was room temperature, which is within the operating range, and one at a temperature well above the operating range. The numbers in the upper left of each oscillograph are serial numbers of A1768 units. The three transistor units were selected to be representative of the bottom, the middle, and the top of the normal distribution of A1768 transistors from an output voltage standpoint. Below each oscillograph of collector current, the oscillator rms output voltage, E_o, and the total dc power input P_{in}, are given for the condition indicated. It is seen that, although the range of peak collector current is greater than 3:1, the output voltage varied only over a range of ± 12 per cent. The power input, of course, varied more, but even here the highest power input is only 2 mw above the maximum requirement, and this at a temperature well in excess of the maximum required. Within the required operating range the power input stayed well below the maximum permissible.

The oscillator frequency, which is 130 cycles, is substantially constant for different transistor units, and with change in transistor ambient temperature over the operating range, variations of less than 1 per cent due to these causes were observed. The chief cause of frequency variation will be inductance and capacitance changes of the tank circuit with time and temperature and inductance changes caused by dc current variation, and these changes should easily be made sufficiently small with proper component design, to hold the oscillator frequency variation within the required ± 7 per cent. Fig. 11 shows curves of the oscillator frequency as a
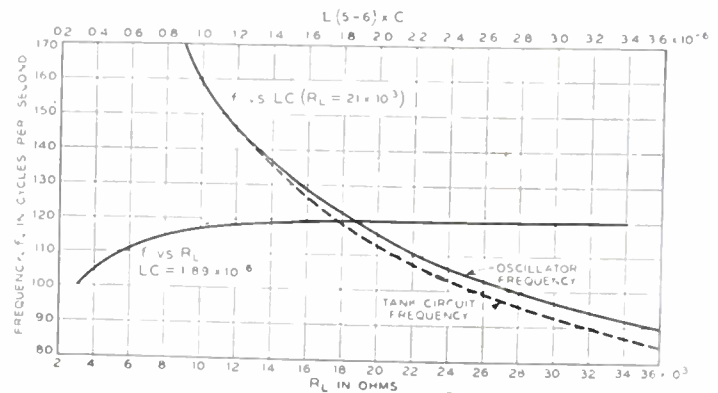
Fig. 11—Oscillator frequency.

function of load resistance and of the LC product of the load winding tank circuit. The frequency is seen to be independent of load resistance above 17,000 ohms and to vary only 3 per cent down to a load resistance of 9,000 ohms. The frequency as a function of tank circuit LC product is interesting largely from the standpoint of the difference between the computed tank frequency and the actual oscillating frequency. This is seen to be approximately 3 per cent for $LC = 1.89 \times 10^{-6}$, which is the LC value for the present circuit. This difference is accounted for by the phase shift in the emitter circuit because of the fact that C_3 is not sufficiently large to make its reactance negligible at the oscillating frequency and this phase shift must be compensated for by the circuit

oscillating off the resonant frequency of the tank circuit. It is apparent that the frequency is controlled principally by the tank circuit, and, with proper design of the tank circuit components should be held within the required range with little difficulty, as pointed out above.

Insofar as harmonics are concerned, the ratio of harmonic voltages to fundamental voltage for the magnetic amplifier load and a 21,000-ohm resistance load are shown in Table I. The harmonic components of the col-

TABLE I

Harmonic Components of Load Voltage and Collector Current in Percentage of their Respective Fundamental Magnitudes

	Load	
	Magnetic Amplifier	21K Resistance
Load Voltage		
Fundamental	100.0	100.0
2nd harmonic	15.4	5.6
3rd "	5.4	3.9
4th "	5.1	2.9
5th "	4.0	0.6
6th "	2.3	0.6
7th "	2.1	0.5
8th "	1.7	0.2
9th "	1.3	0.1
10th "	1.2	0.1
Collector current		
Fundamental	100.0	100.0
2nd harmonic	28.5	26.9
3rd "	40.0	36.8
4th "	55.2	45.3
5th "	24.0	16.3
6th "	13.3	13.9
7th "	25.2	15.8
8th "	9.7	2.1
9th "	9.4	8.4
10th "	10.9	4.7

lector current for the same load conditions are also given. It is seen that the maximum harmonic voltage with the magnetic amplifier load is only 15 per cent of the fundamental voltage, which is well within the requirement and which, when we consider the shape of the collector current wave and the extreme nonlinearity of the load, coupled with the small transformer size and the low fundamental frequency, is surprisingly good. The improvement in the shape of the output voltage wave form for a linear load is apparent from the table.

Finally, regarding stability of output voltage with change in battery voltage, the results are shown in Fig. 12. It is seen that the rms output voltage is practically a linear function of the battery voltage. The percentage drop in output voltage is slightly greater than the percentage drop in battery voltage, but this is more than compensated by the better-than-asked-for performance at the nominal voltage. The function of the battery as a reference voltage in the avc circuit is illustrated by the variation in peak collector voltage and the difference between peak collector voltage and battery voltage, both of which are plotted as a function of battery voltage in the same figure. The difference in peak collector voltage and battery voltage is seen to be almost constant with battery voltage change over a wide range. This difference is equal to the sum of the magnitude of the minimum collector voltage reached when the collec-

tor is driven to voltage saturation and the voltage drop in the base circuit at the same instant.

The circuit operation has been discussed with chief emphasis placed on transistor variations from unit to unit and with temperature. However, because of the reliability and stability of operation obtained with the diode biasing and avc circuits, the lack of criticality of the circuit to transistor parameter changes also extends to the other elements of the circuit. Note, for instance,

Fig. 12—Output voltage versus battery voltage.

the range of variation of load resistance in Fig. 11 over which the circuit continues to oscillate. The one exception is the series emitter resistance R_1. If the design value of this element is set at the absolute minimum value for which stable operation is obtained for all 1768 units, then obviously a decrease in this resistance would result in circuit instability for at least some units, the number increasing with the amount by which R_1 falls below the design value set. In selecting the final value of R_1 for a production design oscillator, the value should be chosen sufficiently higher than the critical value to allow a margin equal to the tolerance on R_1.

Circuit Design Theory

A certain amount of qualitative design theory has already been presented in connection with the general description of the circuit and discussion of the circuit element functions. However, nothing was said about the transformer design. This will now be discussed.

The choice of transformer turns ratio will first be considered. Fig. 13 gives a small signal linear analysis of the oscillator circuit. Equation (2) of this figure gives an expression for the feedback loop gain, $\mu\beta$, of the circuit. The requirement for oscillation of a linear Class A oscillator is that the feedback loop gain shall be unity and

have zero phase angle. In the Class C circuit under consideration, it was assumed that: (a) the requirement for oscillation is that $\mu\beta$ averaged over a complete operating cycle should equal unity; (b) $\mu\beta$ is equal to zero over the nonconducting portion of the operating cycle; and (c) the average value of $\mu\beta$ over the operating portion of the cycle will be equal to $\mu\beta$ determined from the small signal Class A values of the transistor parameters and the base varistor resistance. With these assumptions, (2) of Fig. 13 was solved for a for an operating angle of 150° and a collector load of 1,000 ohms.[3] Two real values of a satisfy the requirement, but the smaller value, namely
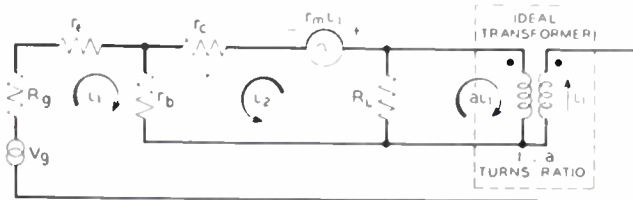
$$i_1(R_g + r_e + r_b + a^2R_L) + i_2(r_b + aR_L) = V_g$$
$$i_1(r_b + aR_L + r_m) + i_2(r_e + r_b + R_L) = 0 \tag{1}$$

$$\Delta = \begin{vmatrix} R_g + r_e + r_b + a^2R_L & r_b + aR_L \\ r_b + aR_L + r_m & r_e + r_b + R_L \end{vmatrix}$$

$$\mu\beta = 1 - \frac{\Delta}{\Delta_0} = \frac{\Delta_0 - \Delta}{\Delta_0} \quad \text{where} \quad \Delta_0 = \Delta \quad \text{when} \quad r_m = 0$$

$$\mu\beta = \frac{r_m(r_b + aR_L)}{(r_e + r_b + R_L)(R_g + r_e + r_b + a^2R_L) - (r_b + aR_L)^2}, \tag{2}$$

where r's are given in terms of the measured R's of the transistor as follows:

$$\left. \begin{array}{lll} r_b = R_{12} & = & 75 \text{ ohms} \\ r_e = R_{11} - r_b = & 80 \text{ ohms} \\ r_m = R_{21} - r_b \doteq & 16,000 \text{ ohms} \\ r_c = R_{22} - r_b \doteq & 6,800 \text{ ohms} \end{array} \right\} \text{ for average M1768 unit.}$$

Note: Any external resistance added in base must be added to r_b.
For feedback winding in series with base rather than emitter expression for $\mu\beta$ is

$$\mu\beta = \frac{r_m(r_b + a^2R_L + aR_L)}{(r_e + r_b + R_L + a^2R_L + 2aR_L)(R_g + r_e + r_b + a^2R_L) - (r_b + a^2R_L + aR_L)^2} \tag{2a}$$

See above note on external base resistance.

Fig. 13—Small-signal linear equivalent circuit of oscillator.

0.325, was selected for more efficient loading. This corresponds to a collector-to-feedback winding turns ratio of 2.9. An experimental check resulted in a final value of 2.8, which shows how close the linear analysis approximation proves to be.

The optimum collector load was initially found to be approximately 1,000 ohms. With an 11,000 ohm load, this would call for a load winding-to-collector winding turns ratio of 3.3, 3.4 was finally determined experimentally, so here again the experimental and computed results check very well.

Once the transformer turns ratio has been determined, the inductance of one of the windings must be selected. The load winding, which is also the tank circuit inductance, was selected. Two conflicting factors govern this choice. It is desirable to keep L as low as possible to provide as small an L/C ratio as possible and

[3] The referred to a is a transformer turns ratio and not the $a = r_m/r_c$ associated with the equivalent circuit of a transistor.

thereby present the minimum possible impedance to the load at harmonics of the fundamental frequency. On the other hand, with a given transformer core size and frequency the nonloaded Q of the transformer is essentially fixed, and the smaller the L is made the greater becomes the ratio of transformer core losses to load power. This latter factor forced a larger value of load winding inductance in the circuit than would have been desirable from the standpoint of the impedance presented by the
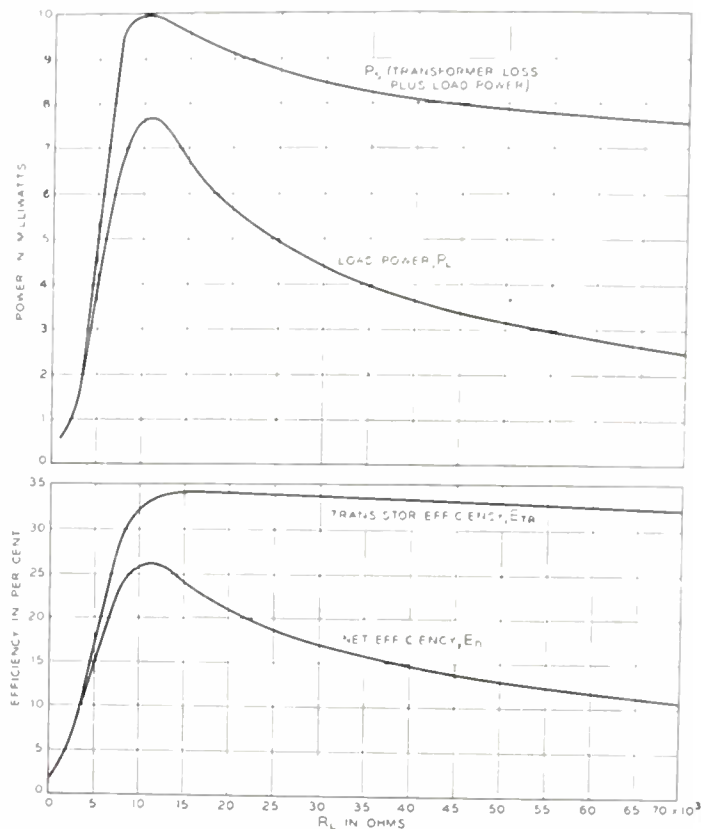
Fig. 14—Power output and efficiency of oscillator.

oscillator output circuit to harmonics of the fundamental frequency. This was a consequence of the low Q resulting from the small transformer selected with prime consideration to space reduction. This will be considered further in connection with the discussion of the power output of the oscillator.

In presenting circuit-performance data, the output performance of the oscillator circuit was discussed only in terms of the voltage delivered to the actual load it was designed to drive. This is a satisfactory load performance criterion only when the load is predetermined and fixed. If general oscillator design theory is to be discussed, and the means for obtaining maximum operating efficiency determined, the oscillator output must be considered in terms of useful power delivered to the load. It was pointed out earlier that in the interest of size reduction, a smaller transformer was used than would have been chosen if maximum circuit conversion efficiency were the prime consideration. The cost of obtaining this reduction in transformer size in terms of loss of circuit conversion efficiency is illustrated in Fig. 14, where curves of power output and conversion efficiency versus load resistance are given. Two curves of power output

are shown, one for the power P_L delivered to the load resistance R_L and the other for the power delivered to the load plus the power dissipated in transformer core losses. The sum of these two powers is indicated as P_S. The corresponding efficiencies, E_N and E_{TR}, are also shown. E_N is the net circuit efficiency given by the ratio of the power delivered to the load to the total dc power input and E_{TR} is the transistor efficiency given by the ratio of the sum of the power delivered to the load and the transformer core losses, to the dc power input.

These curves show that at 11,000 ohms, the load for which the oscillator was originally designed, the load power P_L, and the over-all circuit efficiency E_N are maximized. Transformer losses are, however, absorbing 23 per cent of the total power output. When the load was changed and the equivalent shunt resistance of the linear impedance characterizing the new load proved to be 21,000 ohms, the output power, as shown on the curves, dropped from 7.7 mw at 11,000 ohms to 5.5 mw at 21,000 ohms and the net efficiency dropped from 26 to 20 per cent. Because of a lower power input requirement on the new load, the oscillator still supplied sufficient drive. However, the transformer core losses consume two-thirds as much power as the load itself under the new load conditions.

It is interesting at this point to note an important difference between a low-power vacuum-tube oscillator and a low-power transistor oscillator. Because of the necessity of supplying a minimum amount of heater power, the potential vacuum-tube efficiency at low-power outputs is sufficiently low so that a relatively inefficient load coupling circuit may not appreciably reduce the over-all circuit efficiency. In contrast, the potential transistor efficiency is sufficiently high at low-power outputs so that there is much to be gained in over-all circuit efficiency by maximizing the efficiency of the coupling circuit to the load. An examination of the efficiency curves of Fig. 14 might lead to the conclusion that an increase in load-to-collector transformer winding turns ratio would result in more efficient loading; but without an increase in nonloaded Q of the load winding of the transformer, this change results in no improvement. An examination of the reason for this leads to a solution of the optimum turns ratio for any given set of load conditions, transformer Q, and tank circuit L/C ratio for this circuit.

In obtaining a general solution to the optimum turns ratio of the load to the collector winding for the oscillator circuit being described, it is convenient to combine not only the power delivered to the load and the transformer losses, but also the resistances which account for these losses. R_C, the shunt resistance at the load winding impedance level which accounts for the transformer core losses in the equivalent circuit of the transformer, is therefore combined in parallel with the load resistance, R_L, to get a net load resistance R_S. Thus R_S is the load which would be effective if the transformer were ideal and the losses of the actual transformer were lumped with the load. R_C is closely approximated by $Q_C \omega L_C$ at the fundamental frequency (in this case 120 cycles) since

a resistance load is being considered. Q_C is the nonloaded Q of the load winding of the transformer and L_C is the load-winding inductance. The total power output P_S and the transistor efficiency E_{TR}, given in the curves of Fig. 14, are then replotted as a function of the net load R_S, and the resultant curves are shown in Fig. 15.
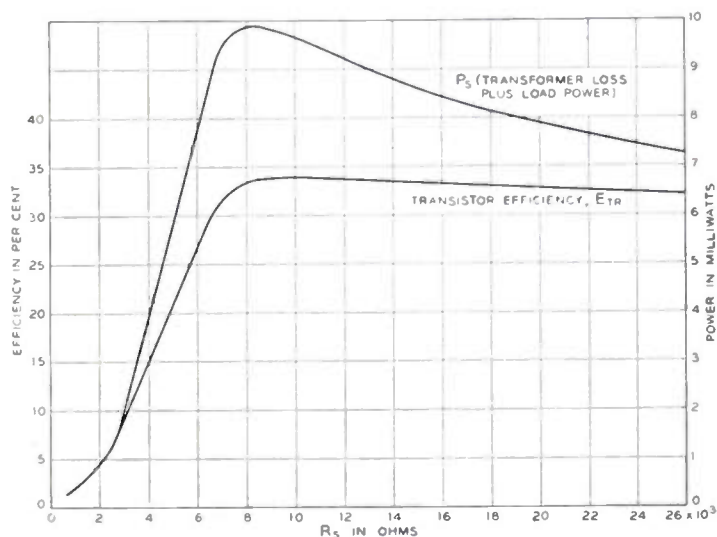
Fig. 15—Transistor power output and efficiency curves.

The design approach to the turns ratio of the load winding to the collector winding of the transformer is given in Fig. 16. The analysis given is a single-frequency analysis which assumes a sufficiently low tank circuit L/C ratio that the harmonics are suppressed to a point where they have only a second-order effect on the efficiency of the circuit. If the L/C ratio is allowed to rise to a point where appreciable power is dissipated at harmonic frequencies, then the wave form and efficiency will both suffer and the analysis given will no longer apply. However, the L/C ratio requirement for best magnetic amplifier performance should preclude an L/C selection which would be high enough to be outside the limits of the analysis.

The design approach of Fig. 16 starts by segregating the load and collector windings, and the load, from the rest of the transistor circuit. An ideal transformer of turns ratio n is then placed between the load resistance and the transformer load winding. The load resistance is then moved through the ideal transformer and appears across the load winding of the transformer with the appropriate impedance transformation, as shown in the figure. Finally, the ideal transformer is eliminated by combining it with the actual transformer in the approximation given in the figure, and it is this final approximation in which the load winding is now changed by a factor of n in turns ratio that is to be considered.

It is convenient to take the ratio of the loaded Q of the transformer load winding Q_S to the nonloaded Q, Q_C, both as defined in Fig. 16 as a fixed parameter in the discussion. For any given value of this ratio it is possible to determine the value of R_S for any given load resistance R_L and any given ratio in load winding turns to the original circuit load winding turns as given by the factor

n, from (3). For this discussion, R_L will be taken as the 21,000 ohms of the present oscillator load. With R_S determined as a function of n, R_L, and Q_C/Q_S, it is now possible, from the curves of Fig. 15, to determine P_S and

way for any other load resistance, however.) The present oscillator transformer load winding has a nonloaded Q of 13.5 at 120 cycles. The maximum value of load-winding inductance consistent with a sufficiently low L/C ratio to satisfy circuit output impedance requirements was selected; for this inductance and a nonloaded Q of 13.5, the loaded Q is 5.2, giving a Q_C/Q_S ratio of 2.5. The "A" curves of Fig. 17 therefore apply, and

$$\omega L_C G_S = \frac{1}{Q_S} = \omega L_C (G_C + G) = \frac{1}{Q_C} + \frac{n^2 \omega L_C}{R_L}$$

$$\frac{1}{Q_S} - \frac{1}{Q_C} = \frac{n^2 \omega L_C}{R_L} = \frac{n^2 R_S}{R_L Q_S},$$

$$R_S = \frac{R_L}{n^2} \left[1 - \frac{Q_S}{Q_C} \right], \tag{3}$$

$$\frac{G}{G_S} = \left[1 - \frac{Q_S}{Q_C} \right], \tag{4}$$

$$P_L \text{ (power into load)} = \frac{G}{G_S} P_S = P_S \left(1 - \frac{Q_S}{Q_C} \right). \tag{5}$$

Where P_S is sum of load power & transformer loss,

$$E_n \text{ (net efficiency)} = \frac{G}{G_S} E_{TR} = E_{TR} \left(1 - \frac{Q_S}{Q_C} \right). \tag{6}$$

Where

$$E_{TR} = \frac{\text{Load power} + \text{transformer loss}}{\text{Total battery power input}} \times 100,$$

$$L_{C'} = n^2 L_C = \frac{n^2 R_S}{\omega Q_S} = \frac{R_L}{\omega Q_S} \left[1 - \frac{Q_S}{Q_C} \right],$$

$$Q_S L_{C'} = \frac{R_L}{\omega} \left[1 - \frac{Q_S}{Q_C} \right]. \tag{7}$$

Fig. 16—Optimization of output transformer turns ratio.
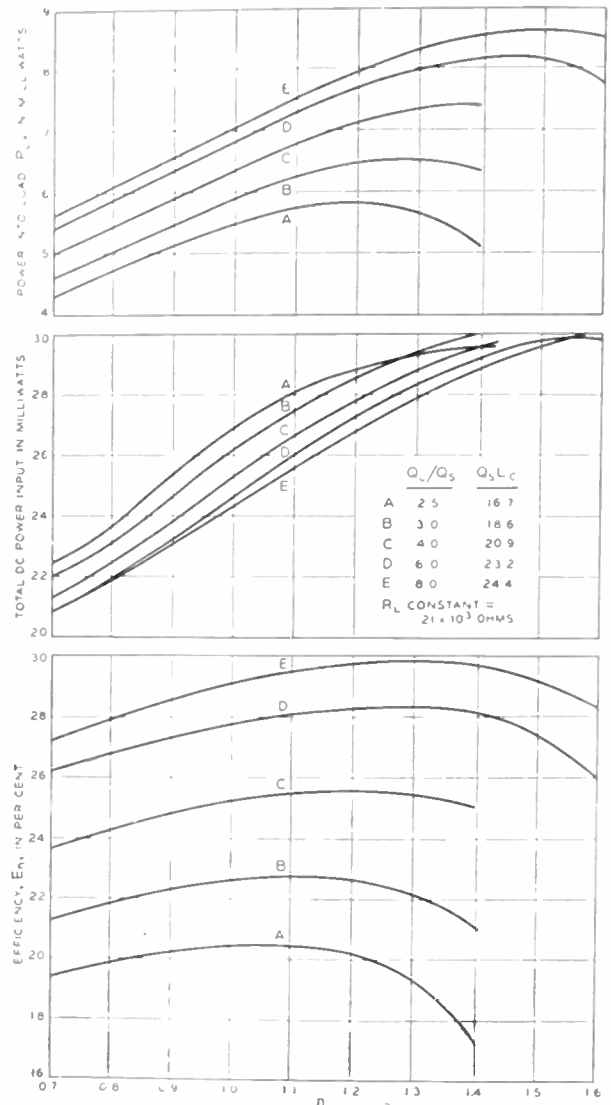
Fig. 17—Power output, power input, and efficiency curves.

E_{TR}. Given P_S and E_{TR}, the load power P_L and the net efficiency E_N can be determined from (5) and (6), respectively. Given P_L and E_N the, total dc power input can then be computed directly. These three operating parameters, P_L, E_N, and total dc power input have been determined as a function of n with Q_C/Q_S as a parameter, and the results are plotted in Fig. 17.

Now consider how the curves of Fig. 17 may be used to arrive at an optimum load-to-collector winding turns ratio. (Note that these curves apply only to a 21,000-ohm load. Similar curves may be determined in the same

an examination of the efficiency curve at $n = 1.0$ shows that the present circuit is already operating at close to maximum potential efficiency. An examination of the "A" power-output curve shows that a slight increase in power output could be obtained by increasing the turns ratio of the load-to-collector windings by a factor of $n = 1.2$, but the increase was not sufficient to merit a change in the final experimental model of the oscillator.

In order to obtain a worth while improvement in either power output or circuit efficiency, a larger transformer with considerably higher nonloaded Q must be used. Consider, for example, operating on curve "C" with $n = 1.33$. This would give 7.4-mw output for

29-mw input,[4] as compared with 5.5-mw output for 27-mw input with the present circuit. Therefore, 1.9 mw of the 2.0 mw added input power, or practically the entire amount, are obtained in added load power. To do this and retain the same L/C ratio as at present, Q_s must be increased to 20.9/16.7 times the present Q_s, or 6.1. (This ratio is determined from Fig. 16.) Since Q_c/Q_s for the "C" curves is 4.0, the required transformer load winding Q is 24.4, or nearly double that of the present transformer.

Any increase in transformer Q obtained by increasing the core size could just as well be used to lower the L/C ratio of the load tank circuit, thereby lowering the impedance to harmonics of the fundamental with a resultant improvement in output voltage wave shape, or the increase in Q could be used to lower the power input retaining the same power output and L/C ratio. Regardless of the way in which increased Q is used to improve performance, an approximate determination of the necessary transformer change may be obtained from the curves of Fig. 17.

If the transistor is not charged with the high core losses in the present transformer, which are a direct result of the emphasis on space reduction in the oscillator model covered in this paper, it is seen from Fig. 17 that the transistor is converting dc power into ac power with a net efficiency of 34 per cent. The theoretical maximum efficiency for a Class C square-wave current, sine-wave voltage oscillator for a 150° operating angle is 74 per cent. An analysis of the collector conversion efficiency of the transistor in the oscillator under discussion shows it to be close to 60 per cent, which is quite close to the ideal. The difference between the 60-per cent collector conversion efficiency and 34-per cent net efficiency is accounted for by power dissipated in the bias circuit and emitter circuit series resistance, bleeder power necessary for self starting, power dissipated in the base circuit, and emitter power necessary to maintain oscillations. Since most of the power dissipated in these elements contributes to the constancy and reliability of circuit performance and should not be considered lost power, the net efficiency of 34 per cent is highly satisfactory.

CONCLUSIONS

The above results demonstrate the suitability of a transistor as the active element in an oscillator to meet the requirements given. The required circuit is simple, reliable, and compact, and the same efficiency could not be achieved using a currently available vacuum tube since the lowest powered tube which could supply the required output power would require a filament power about equal to the total power input to the present transistor oscillator. Furthermore, two batteries would be required, one of which would be of higher potential than the single battery used in transistor oscillator circuit.

[4] Comparable power inputs for the magnetic amplifier load are as much as 3 mw lower than for "equivalent" resistance load since equivalence is not exact.

A Junction Transistor Tetrode for High-Frequency Use*

R. L. WALLACE, Jr.†, L. G. SCHIMPF†, SENIOR MEMBER, IRE, AND E. DICKTEN†

Summary—If a fourth electrode is added to a conventional junction transistor and biased in a suitable way, the base resistance of the transistor is reduced by a very substantial factor. This reduction in r_b permits the transistor to be used at frequencies ten times or more higher than would otherwise be possible. Tetrodes of this sort have been used in sine-wave oscillators up to a frequency of 130 mc and have produced substantial gain as tuned amplifiers at frequencies of 50 mc and higher.

INTRODUCTION

A FEW EXPERIMENTAL MODELS of the junction transistor tetrode to be described in this paper have been built in the laboratory and have been subjected to a preliminary evaluation of their circuit performance.

The measurements which have been made to date indicate that these tetrodes are useful at very much higher frequencies than are conventional junction transistors. Sine-wave oscillations have been obtained at frequencies as high as 130 mc and tuned amplifiers have shown gains as high as 11.8 db at 50 mc. Further-more, these transistors are capable of rather wide-band operation. For example, a grounded-base stage operated between resistive terminations can produce 22.3-db gain with the cutoff frequency at 5 mc. This amounts to a gain-band product of a little more than 800 mc. As a band-pass amplifier for frequencies centered around 32 mc, this same transistor gives 15-db gain over a bandwidth of 9 mc.

Time has not permitted a thorough study of the high-frequency properties of these devices. For this reason, we are forced to set rather limited objectives for this paper. First, we will attempt to show that the device does have useful high-frequency properties by presenting the measured performance of various amplifier and oscillator circuits in which it has been tried. Second, we will present some rough approximations to indicate that three factors may be principally responsible for the improvement in high-frequency performance as compared to that of conventional junction transistors. These are the following:

1. The base resistance of the transistor has been reduced through the use of the tetrode structure.

2. Cutoff frequency of α has been increased by using somewhat thinner p layers than often used.
3. The collector capacitance has been reduced by decreasing the area of the collector junction to approximately 10^{-4} square inches.

The Structure

Fig. 1 shows the geometrical configuration of the junction tetrode and the symbol which will be used in this paper to represent it. This is identically the same structure as previously reported[1,2] for the junction triode, with the exception that a fourth electrode, $b2$, has been added. This connection is made to the p layer in just the same way as the base connection is made but on the opposite side of the bar.
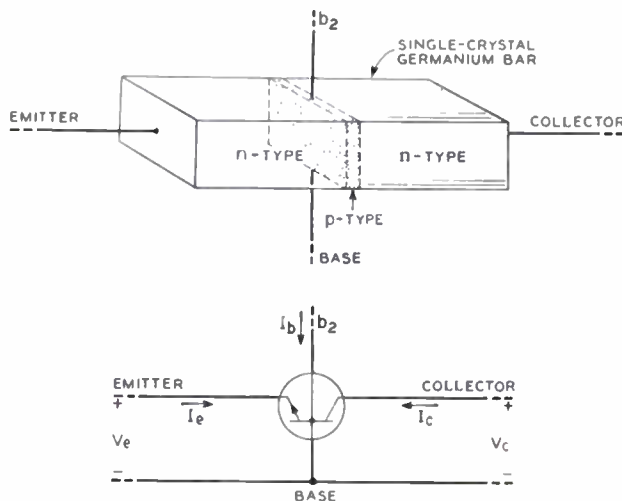
Fig. 1—The tetrode structure is shown above and the symbol used to represent it is shown below.

Throughout this paper we will consider the performance which is obtained by applying a fixed current bias, I_{b2}, between the base and the added electrode. The magnitude and direction of this bias are so chosen as to make the potential of $b2$ something like -6 volts with respect to base. The emitter and collector electrodes are biased approximately the same as they are in conventional junction transistors. In particular, the emitter is normally at something like -0.1 volts with respect to base.

Under these conditions, that part of the emitter junction which is near $b2$ is biased in the reverse direction by almost 6 volts, and hence does not emit electrons into the p layer. In fact, the only part of the emitter junction which is biased in the proper direction to serve as an emitter is in the immediate vicinity of the base contact. This makes all the transistor action take place very near the base contact, and hence might be expected to reduce the base resistance, r_b, of the transistor. In the following sections it will be shown that r_b is reduced and that this results in improved high-frequency performance.

[1] W. Shockley, M. Sparks, and G. K. Teal, "p-n transistors," *Phys. Rev.*, vol. 83, p. 151; 1951.
[2] R. L. Wallace, Jr. and W. J. Pietenpol, "Some circuit properties and applications of n-p-n transistors," *Bell Sys. Tech. Jour.*, vol. 30, p. 530; 1951.

On the basis of the above discussion one might suppose that the same results could be obtained without adding the fourth electrode by simply reducing the dimensions of a conventional junction transistor. It turns out, however, that the required dimensions are considerably smaller than can be handled conveniently. Rough calculations indicate that with the biases mentioned above, the whole active part of the emitter must lie within something like $\frac{1}{6}$ of a mil of the base contact.

Low-Frequency Properties

With a constant current bias applied to $b2$, the remaining three electrodes constitute a triode with properties which are qualitatively the same as those of a conventional junction transistor. The static characteristics are of the same general shape, similar biasing conditions are suitable, and the usual triode equivalent circuit shown in Fig. 2 is appropriate at low frequencies.
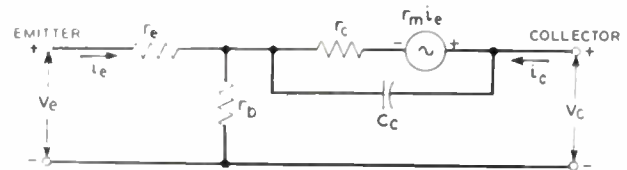
Fig. 2—The low-frequency equivalent circuit of a junction tetrode when operated with fixed bias on the added electrode, $b2$.

The effect of the bias applied to $b2$ is to modify the values of the resistances in this equivalent circuit as shown in Figs. 3 through 6. From these figures it can be seen that none of the parameters varies rapidly with $-I_{b2}$ so long as this current is greater than about 1 ma.
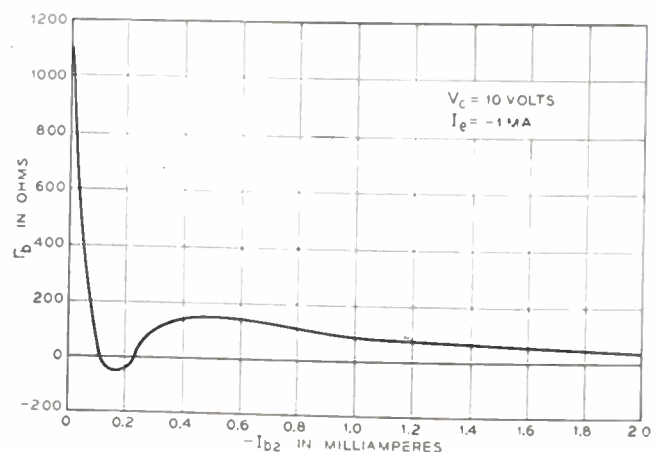
Fig. 3—An important result of bias on the added electrode is the reduction of base resistance.

This is important because it means that the good high-frequency properties of the device are not dependent on a critical setting of this bias. All values between 1 and 2 ma have been found about equally satisfactory.

Fig. 3 shows that with no bias on $b2$, the base resistance is 1,100 ohms for this transistor. As $-I_{b2}$ is increased to 2 ma, r_b decreases to about 40 ohms.

The interesting behavior of r_b at small values of $-I_{b2}$ has not been investigated thoroughly. There is some reason, however, to believe that in this range r_b is

quite frequency dependent and that it has the low values shown only at low frequencies.

Fig. 4 shows that α decreases more or less uniformly from 0.99 to about 0.75 as $-I_{b2}$ is increased from zero to 2 ma. This increase in alpha has the effect of increasing bandwidth at the expense of gain.
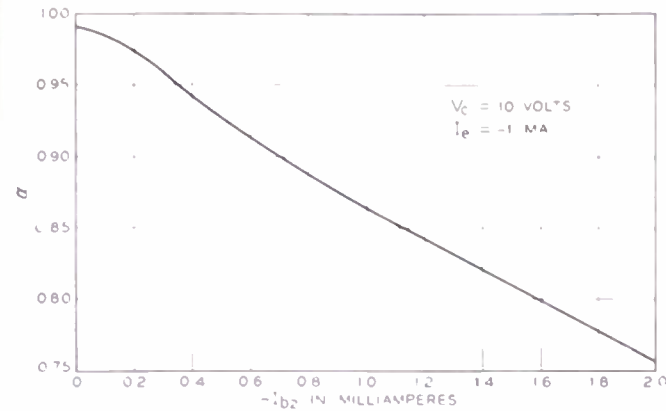
Fig. 4—The bias applied to b2 reduces the current amplification factor, α. Note the expanded scale to which α is plotted.

Fig. 5 shows that r_c is appreciably reduced by the bias applied to b2. This, too, is in the direction of decreasing gain.
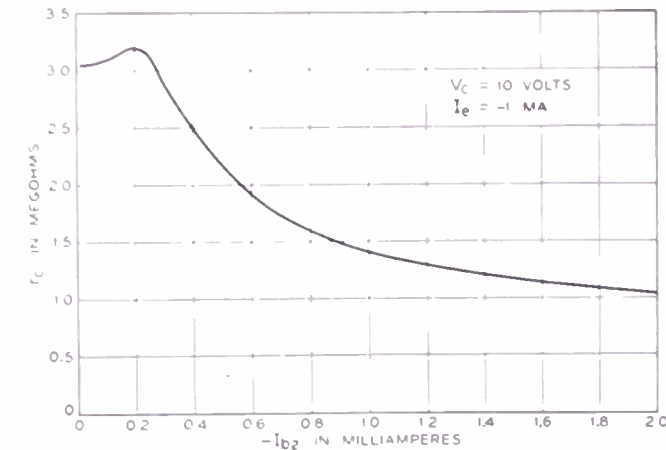
Fig. 5—The collector resistance is appreciably reduced by bias on the fourth electrode.

Fig. 6 shows that in the range 1 to 2 ma r_e is increasing as $-I_{b2}$ is increased, but the total change in r_e is not very great.
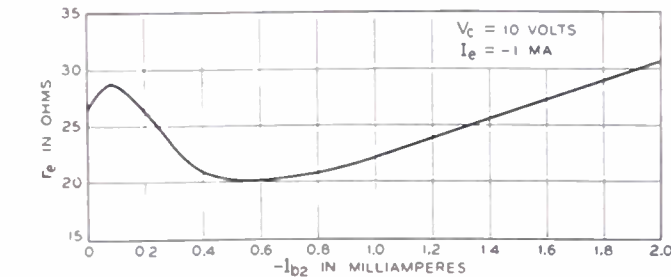
Fig. 6—The emitter resistance is not greatly changed as the bias on b2 is changed.

There is some variation in the properties discussed above from one transistor to another, but the general trends, particularly with respect to r_b and α, are the

same. Table I shows the values of these parameters for five other tetrode transistors.

HIGH-FREQUENCY PROPERTIES

At high frequencies the simple equivalent circuit of Fig. 2 is not adequate to account for the performance of transistors. Among other things it neglects the effect of emitter junction capacitance and fails to account for the fact that α eventually begins to decrease with increasing frequency and to have associated with it a phase angle.

In discussing the dependence of α on frequency it is convenient to define an "α cutoff frequency," $f_{c\alpha}$, as the frequency at which the magnitude of α has been reduced from its low-frequency value by a factor of $1/\sqrt{2}$. This frequency is tabulated for several junction tetrodes in Table I and is of the order of 15 to 20 mc.

Shockley[1] has shown that $f_{c\alpha}$ should be inversely proportional to the square of the thickness of the p layer and should be about 20 mc for the p layers of roughly 0.0005 inches used in these transistors. This is in agreement with the measured values of $f_{c\alpha}$.

TABLE I

MEASURED PARAMETERS FOR SEVERAL JUNCTION TETRODES
$I_e = -2$ ma, $I_{b2} = -1.5$ ma, $V_c = 24$ v

Transistor no	I	II	III	IV	V
r_e (ohms)	6.9	7.2	7.2	16.8	9.6
r_b (ohms)	92.5	68.1	93.0	165	48
r_c (meg)	0.825	0.655	0.720	1.50	0.335
α_0	0.82	0.795	0.819	0.831	0.857
$f_{c\alpha}$ (mc)	18.5	18.2	16.5	16.0	16.5
C_c ($\mu\mu f$)	1.5	3.2	1.8	2.8	2.6

Fig. 7 shows the magnitude of α plotted as a function of frequency. Here again it can be seen that the effect of the bias applied to b2 is to reduce the low-frequency value of α, in this case to about 0.82. Cutoff frequency
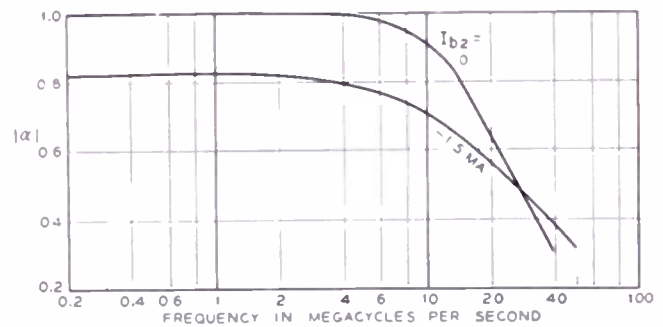
Fig. 7—The cutoff frequency of α is not much changed when the tetrode bias is applied, but the shape of the curve is altered in a favorable way.

is not appreciably changed by bias on b2, but above cutoff frequency α appears to decrease less rapidly with increasing frequency when the bias is applied.

In order to describe completely and accurately high-frequency properties of these transistors it would be necessary to measure the magnitude and phase of all its internal impedances as functions of frequency. Such measurements have not as yet been made, but it has been found that rough approximate calculations can be based on two simplifying assumptions. These are:

1. Assume that the equivalent circuit of Fig. 2 is correct and complete except for the variation of α with frequency.
2. Assume that the variation of α with frequency is given by

$$\alpha = \frac{\alpha_0}{1 + jf//f_{ca}}, \qquad (1)$$

where α_0 is the low-frequency value and f_{ca} is the cutoff frequency for α.
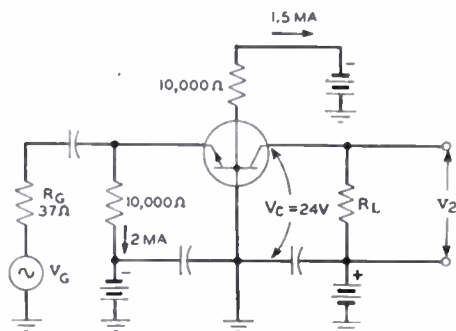
Fig. 8—The circuit used in obtaining the response curves shown in Fig. 9.

Although these assumptions are not sufficiently good to give accurate agreement with high-frequency measurements, they are close enough to permit one to find out which parameters of the transistor are most important at high frequencies. We will use such calculations to show that r_b is very important and that one should expect substantial improvement in high-frequency performance when r_b is reduced as it is in junction tetrodes.

Consider the circuit of Fig. 8 in which the transistor is used as a grounded-base amplifier between a generator of internal resistance R_g and a load of resistance R_L. Such a circuit was treated in an earlier paper,[2] in which it was shown (17) that

$$v_2/v_g = \frac{\alpha R_L}{(r_e + r_b + R_g)(1 + R_L/r_c) - \alpha r_b}. \qquad (2)$$

In this same paper it was shown that the effect of collector capacitance is properly taken into account if r_c is replaced by $r_c/(1+j2\pi r_c C_c f)$. If this substitution is made in (2) and if α is replaced by the value given in (1), the result can be reduced to

$$(v_2/v_g)^2$$

$$= \frac{(\alpha_0 R_L/r)^2}{\left[\begin{array}{l} (1-n+R_L/r_c)^2+(1+R_L/r_c)^2 x_a^2 \\ +(R_L/r_c)^2 x_c^2+(2nR_L/r_c)x_a x_c+(R_L/r_c)^2 x_a^2 x_c^2 \end{array} \right]}, \qquad (3)$$

where

$$r = r_e + r_b + R_g$$
$$n = \alpha_0 r_b/r \qquad (4)$$
$$x_a = f/f_{ca}$$

and

$$x_c = 2\pi r_c C_c f.$$

The power gain can be computed by multiplying this value of $(v_2/v_g)^2$ by $4R_g/R_L$.

In order to show the drastic effect which r_b can have on high-frequency performance, suppose $C_c = 0$ so that $x_c = 0$ and suppose further that R_L/r_c is much less than one. In this case (3) reduces to

$$(v_2/v_g)^2 = \frac{(\alpha_0 R_L/r)^2}{(1 - n)^2 + x_a^2}. \qquad (5)$$

From this it can be seen that the response will be down 3 db when $x_a = (1 - n)$, that is, when

$$f//f_{ca} = 1 - \frac{\alpha_0 r_b}{r_e + r_b + R_g}. \qquad (6)$$

If $r_b = 0$, then the gain of the stage is down 3 db when $f = f_{ca}$. Suppose, however, that $R_g = 25$ ohms and that the constants of the transistor are those shown in Figs. 3 through 5 for $-I_{b2} = 0$ (i.e., for the transistor used as a triode). In this case the response is down 3 db at $f = 0.055 f_{ca}$. If $-I_{b2}$ is raised to 1.5 ma, the same sort of calculation shows that the cutoff frequency is raised to $f = 0.603 f_{ca}$. The effect of the tetrode bias, therefore, is to raise the cutoff frequency of the stage by a factor of eleven. This increase is partly due to the reduction of r_b and partly to the reduction of α_0.
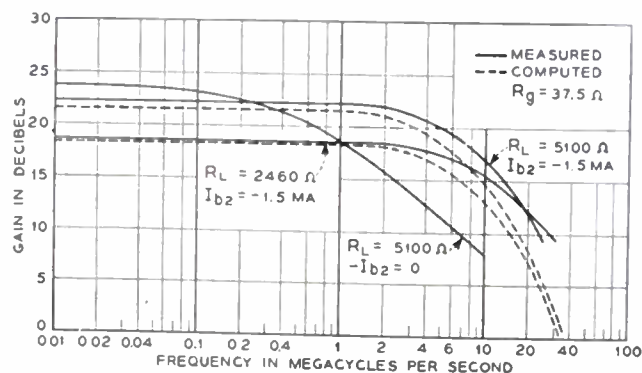
Fig. 9—Measured and computed gains for the stage shown in Fig. 8.

Fig. 9 shows the measured response of transistor I (Table I) in the circuit of Fig. 8. Note that with $R_L = 5,100$ ohms, the gain is 22.3 db and the 3-db point is at 5 mc. This is comparable with the performance of good vacuum tubes. Note that if this transistor is used as a triode ($-I_{b2} = 0$) in the same circuit, the cutoff frequency is reduced to approximately 0.5 mc.

The lower solid curve shows that reducing R_L to 2,460 ohms raises the cutoff frequency to 10 mc and reduces the low-frequency gain to 18.4 db.

The dotted curves in Fig. 9 are computed from (3). They agree reasonably well with the measured response at low frequencies, but the computed gains are appreciably smaller than those measured at high frequencies. This is surprising in view of the fact that emitter capacitance was neglected in the computations. It may be

due to the fact that (1) is not a sufficiently good approximation for the frequency dependence of α. On the other hand, it may be due to a decrease of r_b with increasing frequency.

Assuming that the latter might be the case, (3) was used to compute the values of r_b required to bring the computed curve into agreement with the measured response. For this purpose, a new response curve was measured with $R_g = 0$ and $R_L = 600$. Fig. 10 shows
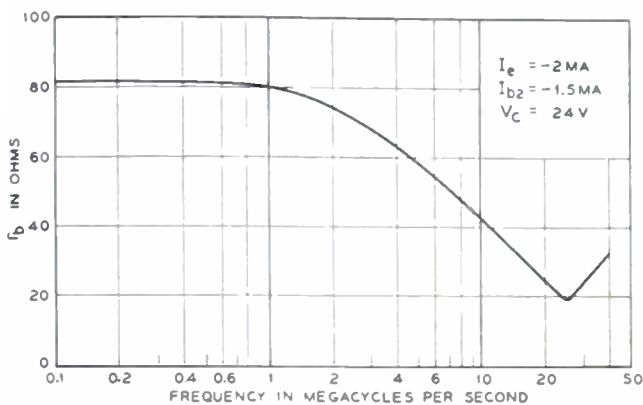
Fig. 10—Errors in the computed curves of Fig. 9 may be due to a variation of r_b, with frequency. The values of r_b, shown here would bring measured and computed curves into agreement.

the values of r_b computed in this way. Obviously, this is a completely arbitrary way of accounting for the failure of (3) to give the right answers at high frequencies. Further measurements are needed to show whether it has any basis in fact. These computed values of r_b will be used in the next section to compute the response of tuned amplifiers.

AMPLIFIERS WITH RESISTIVE GENERATOR AND TUNED LOAD

The output impedance of a grounded-base amplifier driven by a resistive generator looks like a resistance and capacitance in parallel. If the amplifier is loaded by a shunt inductance and resistance as shown in Fig. 11, the inductance will resonate with the output ca-
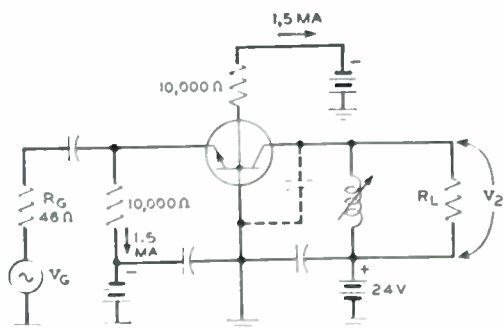
Fig. 11—Simple tuned amplifier circuit used in obtaining the response curves of Fig. 12.

pacitance to give a peaked response. At the frequency of resonance, maximum gain will be obtained when R_L is adjusted for match.

The curves of Fig. 12 were obtained by varying the load inductance to obtain various resonance frequencies

and then adjusting R_L at each frequency to give maximum gain. Note that the measured gain at 50 mc is 11.8 db.
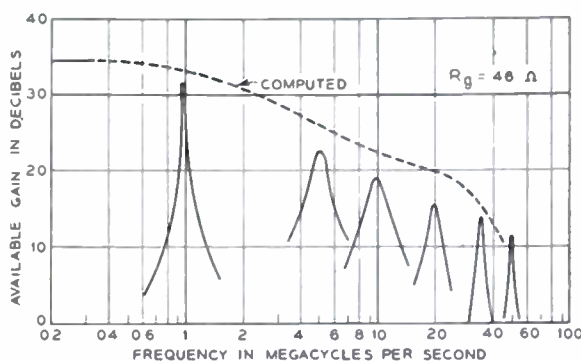
Fig. 12—The peaks of these response curves show the available gain in the circuit of Fig. 11. The dashed curve shows the computed locus of these peaks.

Fig. 13 shows the measured values of the output impedance of the stage in terms of an equivalent parallel resistance and capacitance.
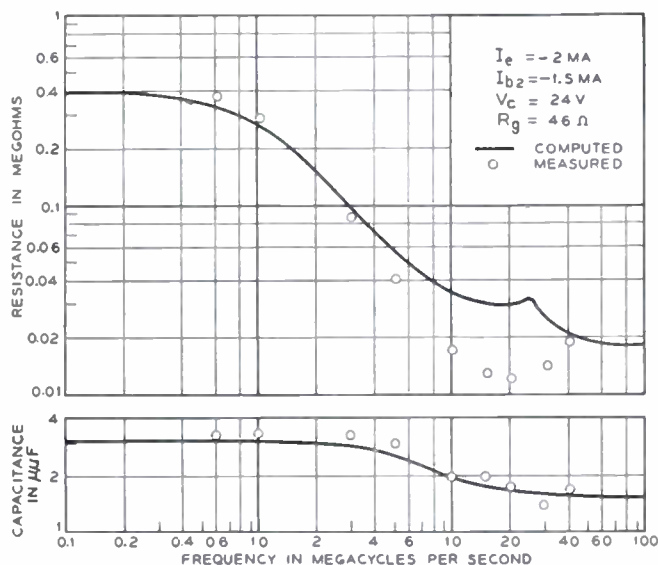
Fig. 13—The output impedance of a grounded-base stage in terms of a shunt resistance and capacitance.

Under the same assumptions made in deriving (3) we can compute these components of the output impedance and the gain which should be measured at the peak of the resonance curve. The results are

$$R_0 = r_c \frac{(1 - n)^2 + x_a^2}{(1 - n) + n x_a x_c + x_a^2}, \tag{7}$$

$$C_0 = C_c \frac{(1 - n) - n x_a / x_c + x_a^2}{(1 - n)^2 + x_a^2}, \tag{8}$$

$$\text{gain} = \frac{(R_g r_c \alpha_0^2 / r^2)}{(1 - n) + n x_a x_c + x_a^2}. \tag{9}$$

Again, if it is assumed that r_b is independent of frequency, the computed gains are much too low at the higher frequencies. If, on the other hand, it is assumed that the values of r_b from Fig. 10 are correct, then (7),

(8), and (9) give the computed values shown on Figs. 12 and 13. The agreement between these measured and computed values is far from excellent, but it is close enough to suggest that r_b may decrease with increasing frequency in some such way as indicated in Fig. 10.

BAND-PASS AMPLIFIERS

The low values of output capacitance and resistance shown in Fig. 13 suggest that it should be possible to build reasonably wide band-pass amplifiers without much sacrifice in gain as compared to the values shown for narrow-band amplifiers in Fig. 12. To illustrate this, the stage shown in Fig. 14 was designed to pass a 9-mc
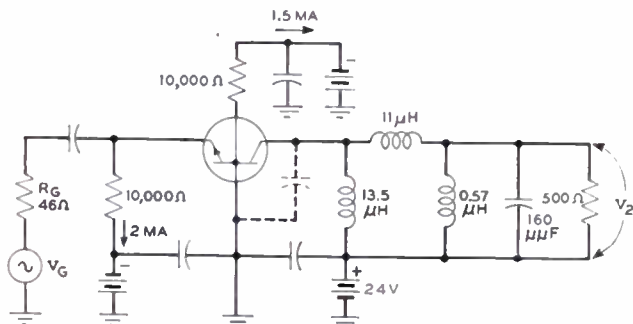
Fig. 14—Band-pass amplifier stage.

band of frequencies centered around 32 mc. The measured response curve is shown in Fig. 15, from which it can be seen that the gain is 15 db, giving a gain-band product of 280 mc. The other response curves shown on Fig. 15 were measured with coupling networks similar to that shown in Fig. 14, but deliberately designed for smaller bandwidths at lower frequencies.
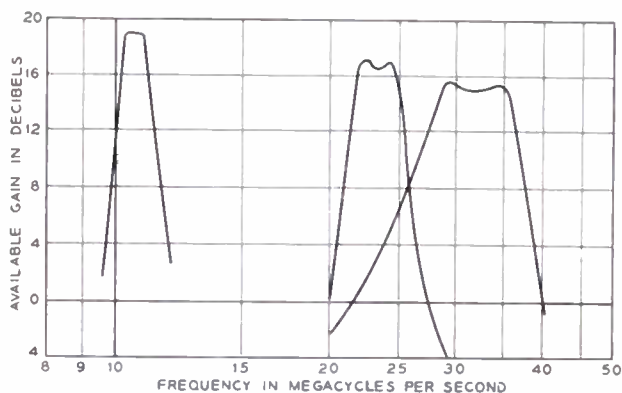
Fig. 15—Measured response for various band-pass amplifiers like that shown in Fig. 14.

OSCILLATORS

The circuit of Fig. 16 has been used to observe the performance of these transistors as oscillators at high frequencies. It resembles a conventional Hartley circuit, with the exception that provision is made for adjusting the value of the capacitance in the feedback path to the emitter. At the higher frequencies the input impedance at the emitter tends to become inductive, and it is necessary to adjust this condenser. Used in this circuit, most of the tetrodes which have been made will produce sinusoidal oscillations up to frequencies as high as 80 to 100 mc. Four or five have been capable of oscillating at frequencies above 100 mc and two will oscillate at 130 mc.
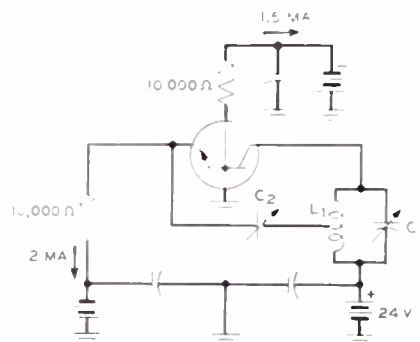
Fig. 16—An oscillator circuit suitable for use with junction tetrodes.

Power output of this oscillator has been measured as a function of frequency for one transistor. Between 40 and 75 mc the measured output was approximately 1 mw. At 100 mc the output is 0.25 mw and at 115 mc, 0.06 mw. The collector dissipation was held to about 30 mw during these measurements.

ACKNOWLEDGMENTS

Effects of Space-Charge Layer Widening in Junction Transistors*

J. M. EARLY†, ASSOCIATE MEMBER, IRE

Summary—Some effects of the dependence of collector barrier (space-charge layer) thickness on collector voltage are analyzed. Transistor base thickness is shown to decrease as collector voltage is increased, resulting in an increase of the current-gain factor (α) and a decrease in the emitter potential required to maintain any fixed emitter current. These effects are shown to lead to two new elements in the theoretical small-signal equivalent circuit. One, the collector conductance (g_c'), is proportional to emitter current and varies inversely with collector voltage. This term is the dominant component of collector conductance in high-quality junction transistors. The other element, the voltage feedback factor (μ_{ec}), is independent of emitter current, but varies inversely with collector voltage. The latter element is shown to modify the elements of the conventional equivalent tee network.

INTRODUCTION

THIS PAPER DESCRIBES an extension of the theory of junction transistors which explains some recent experimental results. In particular, dependence of collector-barrier thickness on collector potential is shown to account for observed collector resistances and to explain high base resistances recently observed, together with low emitter resistances.

Principal attention is given to small-signal theory at low frequencies, since such measurements are most easily made. Emphasis is placed primarily on physical behavior and equivalent circuits, with the mathematical work appearing in the appendix.

First, the existing theory is reviewed and interpreted. The problem is then restated and the change in the assumptions described. A discussion of the physical behavior is followed by a presentation of formal results which are interpreted in terms of physical behavior and then compared with experimental results.

PRESENT THEORY

Junction transistors were first described and analyzed by Shockley in 1949.[1] In 1951, Shockley, Sparks, and Teal described and analyzed some high-quality junction units.[2] The later paper and a companion article by Wallace and Pietenpol[3] showed the characteristics of the units to be almost identical to those predicted by Shockley in his first paper.

In that analysis,[1] Shockley assumed a structure similar to that of Fig. 1, except that it was *p-n-p* rather than

n-p-n. He assumed uniform, but different, resistivities and minority-carrier diffusion lengths in the emitter, collector, and base regions. He related the minority-
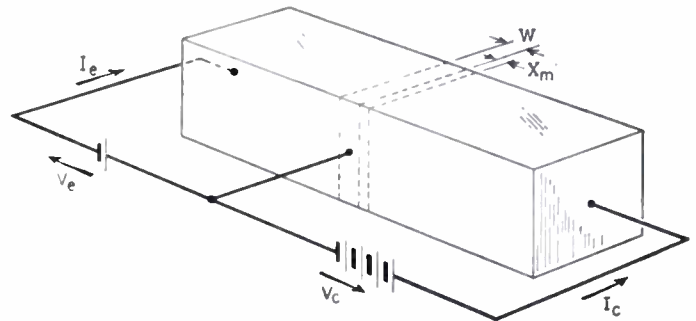
Fig. 1 *n-p-n* transistor.

carrier densities on both sides of each of the *p-n* barriers to the potentials across these barriers and proceeded to solve the diffusion equation for minority carriers in each of the three regions. The relation of the currents across the barriers to the barrier potentials was then easily computed from the relation of minority-
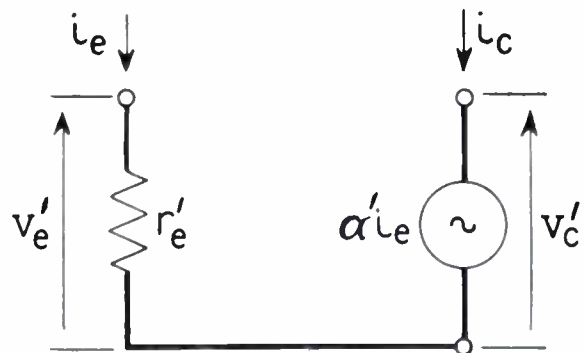
$$r_e' \simeq kT/qI_e$$
$$\alpha' = \beta\gamma$$
$$\beta = \operatorname{sech}\left(\frac{w}{L_{nb}}\right)$$
$$\gamma = \frac{1}{1 + \frac{\sigma_b L_{nb}}{\sigma_e I_{pe}}} \tanh\left(\frac{w}{L_{nb}}\right)$$

Fig. 2—Approximate theoretical ac equivalent circuit for a junction transistor.

carrier densities to these potentials. In the same paper, it was shown that increase in the potential across a barrier increased the barrier thickness. However, it was assumed that changes in collector- and emitter-barrier thicknesses did not affect base-layer thickness.

Under these assumptions, the collector current was shown to be independent of the collector reverse potential whenever the latter exceeded a few times kT/q.

[1] W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, p. 435; July, 1949.
[2] W. Shockley, M. Sparks, G. K. Teal, "The *p-n* junction transistors," *Phys. Rev.*, vol. 83, p. 151; July, 1951.
[3] R. L. Wallace, Jr. and W. J. Pietenpol, "Some Properties and Applications of *n-p-n* Transistors," *Bell Sys. Tech. Jour.*, vol. 30, p. 530; July, 1951. Also, PROC. I.R.E., vol. 39, p. 753; July, 1951.

Shockley, Sparks, and Teal showed that a very simple equivalent circuit results, as shown in Fig. 2.[4] At high frequencies, collector capacitance shunted across the $\alpha' i_e$ current generator must be included.

In general, for active four-terminal devices, the small-signal equivalent parameters depend on bias conditions. From this standpoint, the circuit of Fig. 2 is unusual since both of the parameters are independent of collector potential. In addition, this circuit has no feedback or output elements. A practical feedback element of considerable importance, the majority-carrier resistance[1] in the base layer (r'_b), which will be introduced later in the analysis, has been neglected here for simplicity.

EXTENSION OF THE THEORY

The fundamental problem in determining equivalent-circuit parameters of a junction transistor is that of relating the currents flowing at emitter and collector barriers to the potentials across these barriers. If this problem is solved using Shockley's assumptions, modified only by the additional assumption that the base width is dependent on the collector potential, an equivalent circuit whose parameters are dependent on bias conditions and which has feedback and output elements is obtained.

The physical implications of this assumption may be visualized using Fig. 1. The base-layer thickness is indicated by w and the collector-barrier thickness by x_m. Since the emitter barrier is usually forward biased, it is very thin and may be neglected. If reverse collector potential V_c is increased, the barrier thickness x_m is likewise increased, as Shockley has shown. This reduces the base-layer thickness w, since the barrier layer always spreads in both directions, although not necessarily equally. It is therefore clear that base thickness depends on collector-barrier potential.

The decrease in base-layer thickness has two principal effects: It decreases recombination of injected minority carriers in the base layer since the average carrier diffuses across the narrower base in a shorter time. This increases the transport factor β, which is the fraction of emitted minority carriers which reach the collector.

The second principal effect of decrease in base-layer thickness is a decrease in the impedance presented to minority-carrier current injected by the emitter. The impedance seen by this injected current depends on base-layer resistivity and base-layer thickness. A reduction in thickness results in a reduction in the impedance. Dependence of this impedance on the base-layer thickness is indicated in Fig. 3, in which are shown emitter conductances for electrons injected into the base by the emitter and for the holes which are simultaneously extracted from the base by the emitter. Now since narrowing of the base layer decreases the

resistance to injection of electrons into the base, it must increase the proportion of emitter current carried by electrons. This is an increase in the emitter γ, which is, for an n-p-n unit, the ratio of electron emitter current to total emitter current.

Since both β and γ increase as the base layer is narrowed, it is obvious that an increase of collector voltage produces an increase of the current-gain factor, α', since for most simple junction transistors $\alpha' \simeq \beta\gamma$. Furthermore, what might be called the direct-current emitter resistance, $[R_e = V_e/I_e]$, was shown to become smaller as the collector voltage increased since the decreased base-layer thickness reduced the resistance to injection of electrons by the emitter.
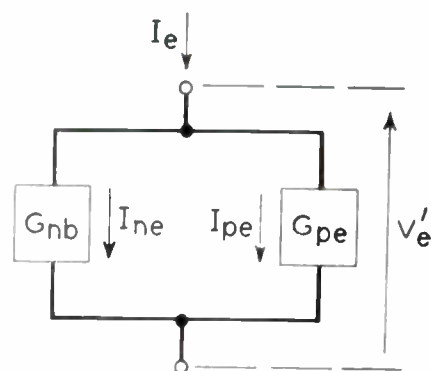
$$G_{nb} = f(\sigma_b, w)$$
$$G_{pe} = f(\sigma_e, I_{pe})$$

Fig. 3—Emitter equivalent circuit.

Dependence of α' and R_e on collector potential is indicated schematically in Fig. 4. The small-signal effects of this dependence are shown next.
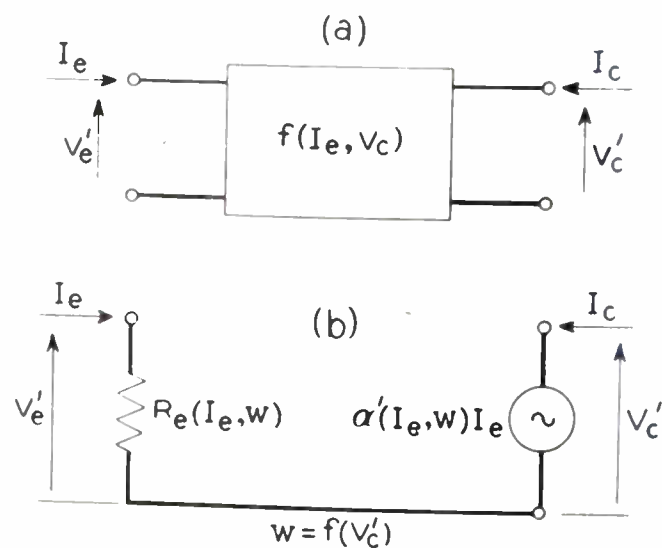
Fig. 4—Generalized dc equivalent circuits for a junction transistor.

SMALL-SIGNAL EFFECTS

Fig. 5 shows the equivalent circuit which is obtained by including the effects of collector-barrier widening. Since $I_c \simeq \alpha' I_e$, the output element g_c' is given in its simplest form as

---

[4] In this and succeeding equivalent circuits, primes are used to distinguish the theoretical circuit parameters from those defined by measurements [see, for example, Fig. 6 of footnote ref. 3]. Primes on potentials indicate barrier potentials rather than terminal potentials.

$$g_c' = \frac{\partial I_c'}{\partial V_c'}\bigg]_{I_e} \simeq I_e \frac{\partial \alpha'}{\partial V_c'} \cdot \qquad (1)$$

This says that small-signal collector conductance is simply a representation of the dependence of alpha on collector potential. The conductance increases linearly

$$r_e' = \frac{kT}{qI_e} \qquad \mu_{ec} \simeq \frac{kT}{qw}\frac{\partial w}{\partial V_c'} \qquad \alpha = \beta\gamma$$

$$g_c' = \frac{1}{r_c'} = I_e\frac{\partial \alpha}{\partial V_c'} \simeq \frac{I_e}{w}\left[(1-\alpha) + (1-\beta)\right]\frac{\partial w}{\partial V_c'}$$
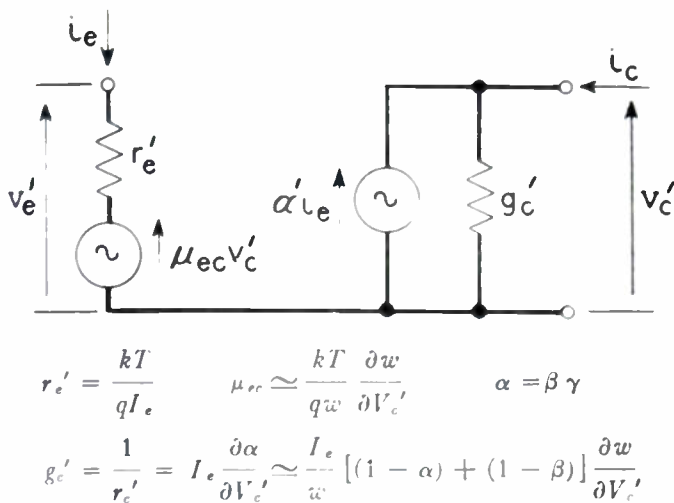
Fig. 5—Detailed theoretical ac equivalent circuit for a junction transistor.

with emitter current since the current change corresponding to any change in alpha depends on the current which is flowing.

For junction transistors having $\alpha' \simeq \beta\gamma$ in the range 0.95–1.0, (1) reduces to the forms derived in the appendix [(11) and (12)].

$$g_c' = \frac{I_e}{w}\left[2(1-\beta) + (1-\gamma)\right]\frac{\partial w}{\partial V_c'}, \qquad (2)$$

or

$$g_c' = I_e\left[\frac{w}{L_{nb}^2} + \frac{\sigma_b}{\sigma_e L_{pe}}\right]\frac{\partial w}{\partial V_c'} \cdot \qquad (3)$$

The factor of two multiplying the $(1-\beta)$ term in (2) results from the fact that $(1-\beta)$ depends on the square of the base-layer thickness w, while $(1-\gamma)$ is directly proportional to base-layer thickness.

Static collector characteristics for an n-p-n junction transistor are shown in Fig. 6. The slope of the constant emitter-current lines decreases with increasing emitter current, but the effect is not obvious when plotted to the scale shown. Fig. 7 shows collector conductance measured at $V_c = 3$, 6, and 12 volts as indicated by the dashed lines in Fig. 6. The theoretical curves were computed using (2); w was determined from alpha cutoff-frequency measurement and the dependence of w on V_c' was estimated from the dependence of this cutoff frequency on collector potential. As a check, it was also estimated from the dependence of collector capacitance on collector potential. $(1-\beta)$ was estimated from diffusion-length measurement for minority carriers and $(1-\gamma)$ was obtained by difference between

measured $(1-\alpha)$ and calculated $(1-\beta)$. The agreement of measured and theoretical curves is as good as can be expected. The decrease of collector conductance with
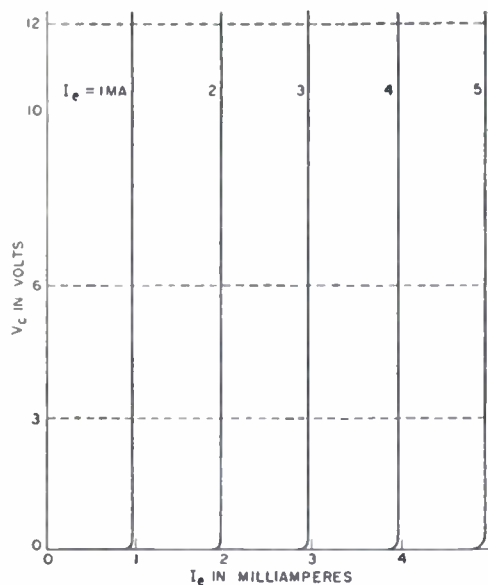
Fig. 6—Collector characteristics of an n-p-n transistor.

increase of collector bias voltage is the result of decrease in the w/V_c'. As collector voltage is increased, the rate of increase in space-charge layer thickness becomes smaller, thus decreasing the sensitivity of base-layer thickness to collector voltage. Fig. 7 shows clearly
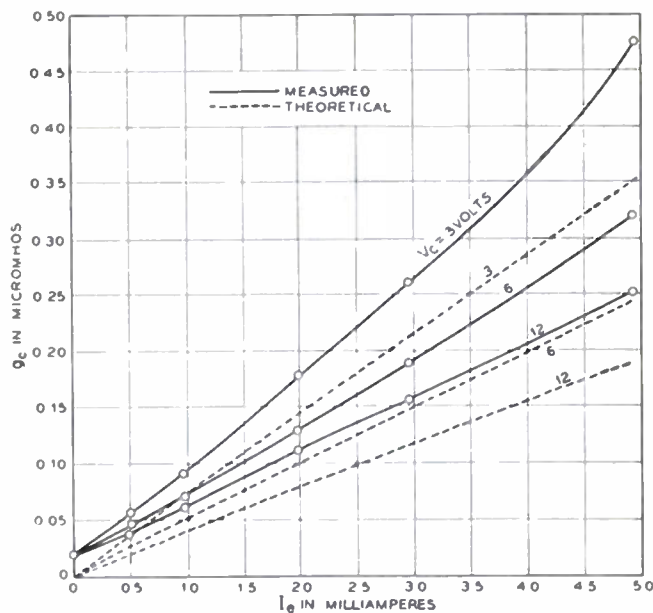
Fig. 7—Collector conductance of an n-p-n transistor.

that the conductance described by (1) is the major component of collector conductance in this unit.

The feedback element $(\mu_{ec}v_c')$ in Fig. 5 is a small-signal parameter representing the effect of collector-voltage increase in decreasing the emitter resistance.[6] At con-

[6] The existence of a back voltage on the emitter resulting from collector bias was first pointed out by Shockley (footnote ref. 1, eq. 5.12). Investigation of the ac effect of this feedback was suggested by Ryder.

stant emitter current, the required forward emitter potential is decreased as the collector reverse bias is increased since less emitter voltage is needed as the base layer becomes thinner. Mathematically, the relation of emitter current, emitter potential, and base-layer thickness is approximately[6]

$$I_e \simeq \frac{n_p q D_n}{w} e^{qV_e/kT}.$$

It is easily shown that

$$\mu_{ec} \simeq \left.\frac{\partial V_e}{\partial V_c}\right|_{I_e} \simeq \frac{kT_1}{qw}\frac{\partial w}{\partial V_c}. \qquad (4)$$

The feedback factor μ_{ec} is independent of emitter current because of the logarithmic relation between emitter current and emitter potential.

In addition to μ_{ec}, two additional feedback effects should be represented in the equivalent circuit. Small-signal current through the base electrode produces a small-signal feedback voltage in the base-layer spreading resistance r_b'. Another feedback voltage results from variation in the value of r_b' caused by space-charge layer width changes.[7] The variation of r_b' produces a small-signal variation in the voltage drop $(I_b r_b')$ in the base layer. A feedback voltage proportional to the average current flowing through the base electrode results. It can be represented by a feedback voltage generator $\mu_{bc} v_c'$ in series with the base-layer resistance as shown in Fig. 9(a). Analytically,

$$\mu_{bc} = I_b \frac{\partial r_b'}{\partial V_c'} = I_b \frac{\partial r_b'}{\partial w}\frac{\partial w}{\partial V_c'}.$$

Since r_b'/w is strongly dependent on the variation of resistivity across the base layer from emitter to collector and since this variation is not known, μ_{bc} has not been calculated; r_b' is a lumped representation of a distributed effect and its calculation is not included here. For an *n-p-n* unit it should be of the order of ρ_b/w, where ρ_b and w are base-layer resistivity and thickness.

Examination of the circuit of Fig. 9(a) will show that at low frequencies, the total open-circuit feedback voltage
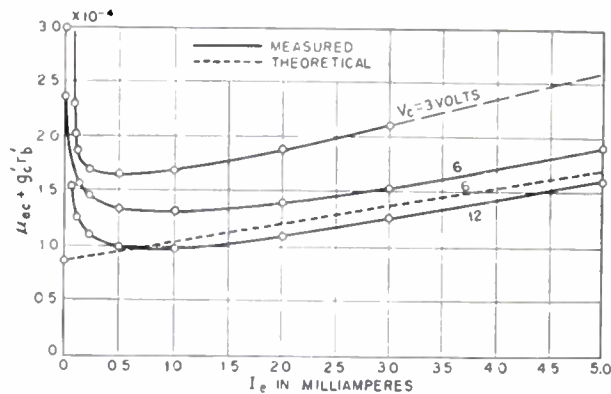
Fig. 8—Voltage-feedback ratio in an *n-p-n* transistor.

[6] An approximate form of (6.2), footnote reference 1.
[7] Existence of this effect was suggested by J. N. Shive.

is $(\mu_{ec}+\mu_{bc}+g_c'r_b')v_c'$. Qualitatively, μ_{bc} should be [for $|I_c| < |I_e|$] less than $g_c'r_b'$ and of opposite sign. The measured open-circuit feedback voltage ratio is plotted in Fig. 8 for the *n-p-n* transistor whose collector conductance is given by Fig. 7. In Fig. 8, the term μ_{bc} is omitted. The rise of measured feedback ratio is attributed to the term $g_c'r_b'$ which increases almost linearly with collector current. Comparison of Figs. 7 and 8 leads to a value for r_b' of 330 ohms for the bias condition $V_c = 6.0$ volts. This value of r_b' is required to produce the calculated curve in Fig. 8 (dashed line) using g_c values at $V_c = 6$ from Fig. 7; μ_{ec} is given by the intercept of the extrapolated experimental curve for $V_c = 6$ in Fig. 8. It is approximately 10^{-4}. The rise in μ_{ec} with decrease of collector potential is the result of the more rapid variation in space-charge layer thickness at lower voltages. The sudden increase in feedback at very low emitter currents is attributed to leakage from collector to emitter across the base layer.
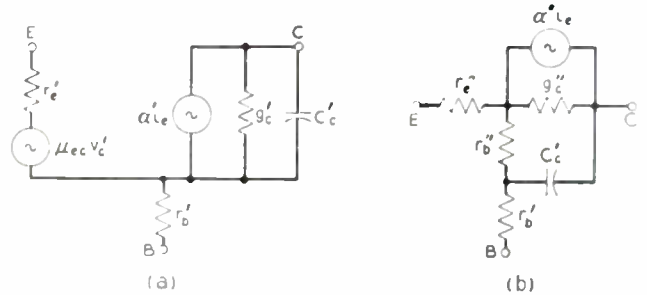
Fig. 9—AC equivalent circuit including collector capacitance and base resistance.

In Fig. 9(b), the parameters r_e', μ_{ec}, α', and g_c' of Fig. 9(a) have been converted to an equivalent tee by use of the relations shown in Fig. 10. Fig. 9(b) can be used together with theoretical expressions for the parameters of a conventional transistor tee network[3] to point out significant frequency variation of the effective base impedance. If the circuit of Fig. 9(a), omitting $\mu_{bc}\pm v_c'$ is reduced to a conventional four-element network,[3] there results

$$z_e = r_e' - \frac{\mu_{ec}(1-\alpha')}{g_c'+j\omega C_c'},$$

$$z_b = r_b' + \frac{\mu_{ec}}{g_c'+j\omega C_c'},$$

$$z_c = \frac{1-\mu_{ec}}{g_c'+j\omega C_c'},$$

$$a = \frac{\alpha'-\mu_{ec}}{1-\mu_{ec}}.$$

Inspection of the expression for z_b shows that at low frequencies it is $r_b'+\mu_{ec}/g_c$, while at frequencies such that $\omega C_c' \gg g_c'$, it is r_b'. This can be understood from inspection of Fig. 9(b), which shows the capacitance C_c' grounded through r_b' rather than through $r_b'+\mu_{ec}/g_c' = r_b'+r_b''$. In high-quality junction transistors r_b'' is ordinarily much greater than r_b'. At $V_c = 6.0$ volts, $I_e = 1.0$ ma for the

unit of Figs. 7 and 8, $r_b'' = 1800$ ohms while $r_b' = 330$ ohms.

The more complete analysis in the appendix shows deviations from the simple theory described above at emitter currents of the order of the saturation currents.

$$r_b'' = \mu_{ec} r_c' = \frac{r_e'}{(1 - \alpha') + (1 - \beta)}.$$

$$r_c'' = r_c'(1 - \mu_{ec})$$

$$r_e'' = r_e' - (1 - \alpha')\mu_{ec} r_c' = \frac{r_e'(1 - \beta)}{(1 - \alpha') + (1 - \beta)}$$

$$\alpha'' = \frac{\alpha' - \mu_{ec}}{1 - \mu_{ec}}$$

Fig. 10—Relations of equivalent-circuit parameters in Fig. 9.

If β and γ are less than 0.95, the more complete analysis is likewise preferable.

CONCLUSIONS

Introduction of the effects of the dependence of collector-barrier thickness on collector voltage has led to explanation of output resistance (r_c') and feedback voltage ($\mu_{ec} v_c'$) in junction transistors. Agreement of measurements and theory is excellent qualitatively and fair quantitatively. It was shown that the low-frequency conventional base resistance ($r_b' + r_b''$ of Fig. 9(b)) does not properly characterize the high-frequency base resistance of a junction transistor.

ACKNOWLEDGMENTS

The writer is particularly indebted to R. M. Ryder and J. N. Shive for suggestions and encouragement, likewise to W. J. Pietenpol and J. A. Morton. Thanks are also due R. L. Johnston for the measurements.

APPENDIX

This appendix indicates one method by which expressions for the collector conductance[8] g_c and the voltage feedback factor μ_{ec} of junction transistors may be derived. The assumptions are, in general, those used by Shockley. Each of the three transistor regions indicated in Fig. 1 may be characterized by a resistivity and a diffusion length or life-path for minority carriers. Injected carriers flow through the base layer principally by diffusion. The additional assumption is made that base thickness w depends on collector potential V_c, through variation of the barrier thickness x_m.

The currents and potentials used are those at the emitter and collector barriers.[8] Majority-carrier resistance effects in the base layer are neglected and frequency effects are not considered. It is assumed that collector reverse bias potential is many times kT/q (e.g., > 0.5 volts) so that the classical p-n diode reverse conductance may be neglected. At 0.5 volt, it is of the order of 10^{-5} micromhos or less. Its magnitude decreases one decade per 60 mv.

[8] For simplicity, no primes are employed in the appendix.

Collector Conductance

Collector conductance is defined as

$$g_c = \frac{\partial I_c}{\partial V_c}\bigg|_{I_e}. \tag{5}$$

It may be shown that for an *n-p-n* junction transistor, emitter and collector currents are given by[9]

$$I_e = \coth\left(\frac{w}{L_n}\right) I_{n0}(V_e)$$
$$\qquad - \operatorname{csch}\left(\frac{w}{L_n}\right) I_{n0}(V_c) + I_{p0}(V_e), \tag{6}$$

$$I_c = - \operatorname{csch}\left(\frac{w}{L_n}\right) I_{n0}(V_e)$$
$$\qquad + \coth\left(\frac{w}{L_n}\right) I_{n0}(V_c) + I_{p0}(V_c), \tag{7}$$

in which the I_0's have the form

$$I = \frac{qD}{L}\left(e^{-qV/kT} - 1\right).$$

Equation (7) may be combined with (6) to yield

$$I_c = -\beta\gamma I_e + I_{n0}(V_c)\left[\coth\left(\frac{w}{L_n}\right)\right.$$
$$\qquad\left. - \beta\gamma \operatorname{csch}\left(\frac{w}{L_n}\right)\right] + I_{p0}(V_c), \tag{8}$$

in which[1] $\beta = \operatorname{sech}(w/L_n)$ and

$$\gamma = \frac{1}{1 + \dfrac{I_{p0}(V_e)}{I_{n0}(V_e)}\tanh\left(\dfrac{w}{L_n}\right)}.$$

The collector conductance may be found by substitution of (8) into (5)

$$g_c = \frac{\partial I_c}{\partial V_c}\bigg|_{I_e} = \frac{\partial I_c}{\partial w}\bigg|_{I_e}\frac{\partial w}{\partial V_c}. \tag{9}$$

In general, the second and third terms of (8) may be neglected since they are very much less than the first term. There then results, after some manipulation,

$$g_c = \frac{I_e}{L_n}\left[\gamma\beta^2 \sinh\left(\frac{w}{L_n}\right) + \gamma^2\beta^3 \frac{I_{p0}(V_e)}{I_{n0}(V_e)}\right]\frac{\partial w}{\partial V_c}, \tag{10}$$

which is a substantially complete expression for the collector conductance. If β and γ are assumed to be almost unity

$$\left(I_{p0}(V_e) \ll I_{n0}(V_e) \text{ and } \frac{w}{L_n} \ll 1\right),$$

and the hole and electron components of emitter current are expressed in terms of conductivities and diffusion lengths, (10) simplifies to

[9] Footnote reference 1, (6.2) and (6.1) modified to this notation.

$$g_c = I_e \left[\frac{w}{L_n{}^2} + \frac{\sigma_b}{\sigma_e L_p} \right] \frac{\partial w}{\partial V_c}, \tag{11}$$

or to

$$g_c = \frac{I_e}{w} \left[2(1 - \beta) + (1 - \gamma) \right] \frac{\partial w}{\partial V_c}. \tag{12}$$

The term $\partial w / \partial V_c$ may be written as

$$\frac{\partial w}{\partial V_c} = \frac{\partial w}{\partial x_m} \frac{\partial x_m}{\partial V_c}, \tag{13}$$

in which collector-barrier thickness x_m is given for a junction of uniform impurity-concentration gradient by[10]

$$x_m{}^3 = \frac{3\kappa\epsilon_0 V_c}{qa}, \tag{14}$$

and for a step junction having high p-conductivity to one side and uniform but much lower n-conductivity on the other side by[10]

$$x_m{}^2 = \frac{2\kappa\epsilon_0 V_c}{qn_n}. \tag{15}$$

The $\partial w/\partial x_m$ is $-1/2$ for the junction of uniform concentration gradient since the space-charge layer spreads equally in the two directions, and is -1 for the second case.

The $\partial w/\partial V_c$ then has the value for the graded junction,

$$\frac{\partial w}{\partial V_c} = - \frac{x_m}{6V_c}. \tag{16}$$

and for the step junction

$$\frac{\partial w}{\partial V_c} = - \frac{x_m}{2V_c}. \tag{17}$$

Voltage feedback factor

The feedback factor is defined by

$$\mu_{ec} = \frac{\partial V_e}{\partial V_c}\bigg|_{I_e}.$$

It may be obtained from (6) by setting $dI_e/dV_c = 0$ and solving for $\partial V_e/\partial V_c$.

$$\frac{dI_e}{dV_c} = 0 = \left[I_{n0}(V_e) \frac{\partial}{\partial w} \coth\left(\frac{w}{L_n}\right) \right.$$
$$\left. - I_{n0}(V_e) \frac{\partial}{\partial w} \operatorname{csch}\left(\frac{w}{L_n}\right) \right] \frac{\partial w}{\partial V_c} \tag{18}$$
$$+ \coth\left(\frac{w}{L_n}\right) \frac{\partial I_{n0}(V_e)}{\partial V_c} + \frac{\partial I_{p0}(V_e)}{\partial V_c}$$
$$- \operatorname{csch}\left(\frac{w}{L_n}\right) \frac{\partial I_{n0}(V_c)}{\partial V_c}.$$

The terms containing $\partial I_{n0}(V_c)/\partial V_c$ may be neglected since $I_{n0}(V_c)$ is substantially independent of V_c for collector reverse bias voltage more than a few times kT/q.

If the operations indicated in (18) are carried through, there is obtained

$$0 = - \frac{qD}{L_n{}^2} \left[\frac{e^{-qV_c/kT} - 1}{\sinh^2\left(\frac{w}{L_n}\right)} + \frac{\cosh\left(\frac{w}{L_n}\right)}{\sinh^2\left(\frac{w}{L_n}\right)} \right] \frac{\partial w}{\partial V_c}$$
$$- \frac{q}{kT} \left[\coth\left(\frac{w}{L_n}\right) \frac{qD}{L_n} \frac{e^{-qV_c/kT}}{\gamma} \right] \frac{\partial V_e}{\partial V_c}, \tag{19}$$

which yields

$$\frac{\partial V_e}{\partial V_c} = \frac{-kT}{qL_n} \frac{\beta\gamma}{\sinh\left(\frac{w}{L_n}\right)} \frac{\partial w}{\partial V_c}$$
$$\cdot \frac{e^{-qV_c/kT} - 1 + \cosh\left(\frac{w}{L_n}\right)}{e^{-qV_c/kT}}. \tag{20}$$

If β and γ are close to unity, this becomes

$$\mu_{ec} = \frac{\partial V_e}{\partial V_c} \simeq + \frac{kT}{qw} \frac{\partial w}{\partial V_c}. \tag{21}$$

The $\partial w/\partial V_c$ was discussed in the section on collector conductance.

LIST OF SYMBOLS

a—impurity concentration gradient in atoms/cm^4

D—diffusion constant for minority carriers

$g_c = \partial I_c/\partial V_c|_{I_e}$—collector conductance

G_{pe}, G_{nb}—dc emitter conductances for electrons and holes

I_e, I_c—dc emitter and collector currents

i_e, i_c—small-signal emitter and collector currents

k—Planck's constant

I_{n0}, I_{p0}—electron and hole saturation currents of junctions considered as isolated diodes

L_n—diffusion length for electrons in base layer

L_p—diffusion length for holes in emitter body

n_n—electron density in n-semiconductor

q—electronic charge 1.6×10^{-19} coulombs

r_c, r_e, r_b—small-signal equivalent collector, emitter and base resistances

R_e—dc emitter resistance

T—temperature in degrees K

V_e, V_c—dc emitter and collector voltages

v_e, v_c—small-signal emitter and collector voltages

w—base-layer thickness

x_m—collector-barrier thickness

α—current-gain factor

β—transport factor

γ—emitter efficiency

ϵ_0—permittivity of space

κ—relative dielectric constant

σ_b, σ_e—base and emitter conductivities

μ_{ec}—voltage feedback factor

---

[10] Footnote reference 1, (2.20) and (2.54) modified to this notation. Here only space-charge layer-widening effects are considered. As mentioned previously, classical reverse conductance is neglected. It is many orders of magnitude smaller than the term analyzed here.

New Photoelectric Devices Utilizing Carrier Injection*

KURT LEHOVEC†

Summary—The detection of injected carriers by their absorption and by emission due to their recombination with majority carriers is discussed. Two new components are proposed: (1) the "photo-modulator," which permits modulation of a light beam by the change in absorption due to injected carriers (the photomodulator in connection with a phototransistor, or some other type of photocell, may be useful as an amplifier); and (2) the "graded seal junction," which permits the light to be transmitted to and from a *p-n*-junction without substantial absorption in the bulk semiconductor.

I. INTRODUCTION

I**N ALKALI HALIDE CRYSTALS** optical investigation of the injection of electrons is possible since visible coloration results from trapping injected electrons in halogen vacancies (*F*-centers). The migration of coloration into the crystal accompanies the initial phase of the injection process.[1,2,3]

The question arises whether injected carriers in germanium and silicon could be observed optically. The word "optically" refers to absorption or emission.

II. ABSORPTION OF INJECTED CARRIERS

The "Photomodulator"

There are in principle three types of absorption which may arise from the injected carriers: (1) the absorption of the injected carriers proper; (2) the absorption arising from trapping of the injected carriers (comparable to *F*-centers in alkali halides); (3) the increase in absorption due to the increase in concentration of majority carriers, resulting from neutralization of the space charge of the injected carriers.

We shall be concerned in what follows with the use of absorption due to injected carriers [case (1) or (3)] for a photoelectric device. The absorption of free carriers is responsible for the absorption in the region on the long wave side of the lattice absorption edge. Briggs[4] has correlated the absorption coefficient of silicon in this region with the concentration of carriers. From Briggs' data one obtains the following relation between hole concentration and absorption coefficient at 2 microns: $n_h \sim 8.9 \times 10^{15} \Gamma$ (*n* in cm⁻³, Γ in cm⁻¹). One should expect therefore that the absorption in the wavelength region on the long wave side of the absorption edge can

be modulated by carrier injection. This effect could be used for an electronic means to modulate a light beam ("photomodulator").

This device works as follows: A light beam of wavelength longer than that of the absorption edge of the semiconductor is passed through the region in which carriers are injected (left part of Fig. 1). The intensity of the light beam emerging from the semiconductor is modulated by the modulated absorption of the semiconductor due to the injected carriers. A quantitative estimate of the order of magnitude of the effect is given in the appendix.
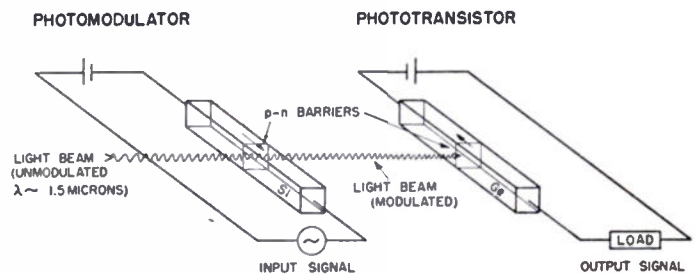
Fig. 1—Principle of a "photoamplifier," consisting of a "photomodulator" and a phototransistor.

The modulated light beam can be transformed back into an electric signal using, for example, a phototransistor (right part of Fig. 1). The material of the phototransistor must absorb the light in order to convert it into free carriers; on the other hand, the material of the photomodulator must transmit (most of) the light. Thus the phototransistor must be made from a material differing from that of the photomodulator (e.g., Si and Ge; see Fig. 1).

Amplification may be achieved with the combination of a photomodulator and a phototransistor for two reasons: (1) The intensity of the light beam to be modulated may be chosen in principle to be very high (independent of the input signal); it seems therefore possible that after transforming the light quanta into current, current-amplification may be achieved. The calculations of the appendix indicate that rather high light intensities, of the order of 20 watts/cm², will have to be used in order to obtain current amplification. In practice the intensity of the light beam will be restricted in order to avoid heating effects of the semiconductor. (2) The ratio of the impedances of carrier injector (*p-n*-junction biased in the forward direction) and phototransistor (*p-n*-junction biased in the blocking direction) is very low. Thus a low impedance signal can be transformed into a high impedance signal.

By passing the same light beam through several photomodulators in series, mixing of several signals can

[1] O. Stasiw, "Zur elektrischen Wanderungsgeschwindigkeit der Farbzentren in Alkalihalogenidkristallen," *Gött. Nachr.*, Heft 4, p. 387; 1933.

[2] R. W. Pohl, "Electron conductivity and photochemical processes in alkali halide crystals," *Proc. Phys. Soc.* (London), vol. 49, p. 3; 1937.

[3] R. Hilsch and R. W. Pohl, "Steuerung von Elektronenströmen mit einem Dreielektrodenkristall und ein Modell einer Sperrschicht," *Z. Phys.*, vol. 111, p. 399; 1938.

[4] H. B. Briggs, "Infra-red absorption in silicon," *Phys. Rev.*, vol. 77, p. 727; 1950.

be achieved. Frequency response of the photomodulator is the same as that of carrier injection over a p-n-junction.

III. Light Emission Due to Recombination of Injected Carriers

Graded Seal Junctions

Recently,[5] a new type of semiconductor light source was described. The light source consists of a combination of a p-n-barrier and a phosphor. The excitation process consists of injection of carriers over the p-n-barrier by passing a current in the forward direction. The light emission results from the recombination of these carriers. We have studied the effect on silicon carbide.[5] The effect was recently observed also on germanium and silicon.[6] In the case of germanium and silicon the emitted light has wavelengths close to the absorption edge. Therefore, part of the emitted light is absorbed in the bulk material adjacent to the junction over which injection took place. In what follows, a junction is described which eliminates the loss of light in the bulk semiconductor.

In semiconductor devices p-n-junctions have found many useful applications as rectifiers, elements in transistors, and photocells, and will be used probably as light emitters. The junctions used so far consisted of two adjacent regions of the same material, containing acceptors- and donor-type impurities, respectively. It is desirable (if not necessary) that the two regions mentioned above be part of a single crystal in order to avoid undesirable effects of grain boundaries (surface states).

In what follows, a new type of p-n-junctions will be discussed. The p- and n-regions differ not only in the type of conduction but also in the crystal structure. A combination of materials which form solid solutions, for example, germanium and silicon,[7] or selenium and tellurium, is of particular interest. In the case of the existence of mixed crystals in any compositions, a transition from material A to material B within a single crystal can be achieved.

In the following we shall mention briefly three of the advantages which the "graded-seal junctions" have as compared to ordinary p-n-junctions: (1) Graded seal junctions can be prepared conveniently by fusing material A to material B at temperatures above the melting point of A but below the melting point of B, and subsequent slow cooling (e.g., using a seed of B in the melt of A).[8] (2) It is known that the spectral region of the strongest barrier photoeffect coincides with the long wavelength absorption edge of a semiconductor. Since the position of the absorption edge of materials A and B differs, one material (say B) will be transparent in a spectral region (photon energy $h\nu$), where material A has its maximum photosensitivity. Thus a graded seal junction can be used to construct a large area junction-type photocell with a negligible loss of light by absorption outside of the junction. (See Fig. 2; the proportions of the energy level diagram correspond to a p-silicon versus n-germanium junction.) (3) Current passing a p-n-junction in the forward direction may lead to light emission. In a graded seal junction of the type shown in Fig. 2, light emitted from material A will not be absorbed when passing through material B.
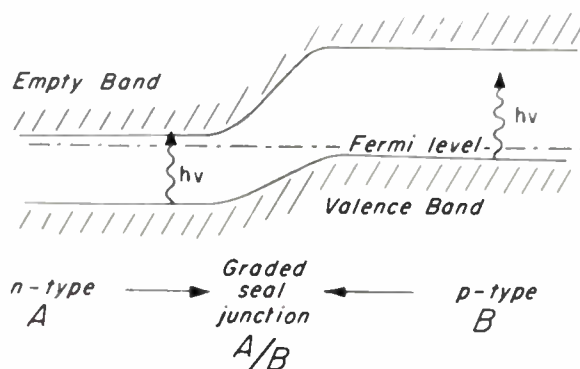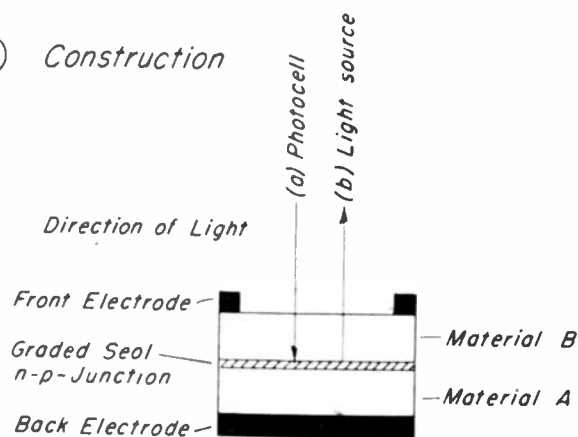
Fig. 2—Energy-level diagram (1) and construction (2) of a graded seal junction for use as (a) photocell and (b) as a light source.

IV. Note on "Electroluminescence"

Light emission from phosphor powders, for example, ZnS or ZnO, at the application of an ac field has been known for some time under the name "electroluminescence."[9] This effect increases with the field intensity and the frequency of the applied field.

[5] K. Lehovec, C. A. Accardo and E. Jamgochian, "Injected light emission of silicon carbide crystals," *Phys. Rev.*, vol. 83, p. 603; 1951; "Light emission produced by current injected into a green silicon carbide crystal," *Phys. Rev.*, in print.

[6] J. R. Haynes and H. B. Briggs, "Radiation produced in germanium and silicon by electron-hole recombination," *Bull. Am. Phys. Soc.*, vol. 27, p. 14; 1952.

[7] H. Stöhr and W. Klemm, "Über Zweistoffsysteme mit Germanium. I. Germanium/Aluminium, Germanium/Zinn, und Germanium/Silicium," *Z. Anorg. allg. Chem.*, vol. 4, p. 305; 1939.

[8] Another method suitable for preparation of a graded junction consists in the deposition of material A from the vapor phase on a heated substrate of material B.

[9] G. Destriau, "Researches on electrophotoluminescence," *Trans. Faraday Soc.*, vol. 35, p. 227; 1939; "The new phenomenon of electroluminescence and its possibilities for the investigation of crystal lattice," *Phil. Mag.*, vol. 38, pp. 700, 774, 880; 1947; and others.

We propose the following explanation for the electroluminescence: Near the boundaries of the crystallites there is a region of distortion, causing an extremely short life time of minority carriers. Accordingly, also the rate of thermal creation of minority carriers is very high in this region. An applied field is able to sweep the minority carriers into the bulk of the semiconductor, where they recombine under light emission. Thus we consider the "electroluminescence" as arising from injection of minority carriers from a region of high rate of thermal creation and of low relative probability of optical recombination into a region of low rate of thermal creation and of high relative probability of optical recombination.

V. APPENDIX

A. Theory of the Photomodulator

The current in the forward direction over a *p-n*-junction is due to injection of minority carriers (concentration $n(x)$). The minority carriers migrate into the bulk semiconductor by (a) diffusion and (b) under the influence of an electric field. This migration is terminated by recombination with majority carriers.

For a quantitative estimate of $n(x)$ we shall assume that diffusion current exceeds largely the field current of injected carriers. The condition of continuity then takes the simple form,

$$D \frac{d^2 n}{dx^2} = \frac{n}{\tau}. \qquad (1)$$

($D = vkT/e$ diffusion constant of injected carriers, $v =$ mobility, $k =$ Boltzman constant, $T =$ absolute temperature, $e =$ elementary charge, $\tau =$ life time of injected carriers.)

Integration gives

$$n = n_0 \exp\left(-\frac{x}{\sqrt{D\tau}}\right),$$

where n_0 is the concentration of minority carriers immediately adjacent to the junction. The concentration n_0 can be expressed by the current I, using the equation

$$I = -HWDe \frac{dn}{dx}\Big|_{x=0} = HWn_0 e \sqrt{\frac{D}{\tau}}. \qquad (3)$$

($H =$ height of the rectangular junction and $W =$ width.) Hence,

$$n_0 = \frac{I}{HWe} \sqrt{\frac{\tau}{D}}. \qquad (4)$$

According to (2), n_0 decays by a factor $1/e = 1/2.72$ at the distance

$$L = \sqrt{D\tau}. \qquad (5)$$

With $\tau \approx 100$ μsec and $D \approx 50$ cm²/sec, which appear to be values of reasonable order of magnitude for germanium and silicon of proper preparation, we obtain $L \approx 0.07$ cm.

Assume a light beam of intensity per unit area J (expressed in quanta per cm² per second) and of rectangular cross section (height H, width X) incident on the semiconductor parallel to the junction and to the side of width W (Fig. 1). The fraction R of the light beam is reflected and the intensity

$$Q = JXH(1 - R)\exp(-W\Gamma) \qquad (6)$$

emerges from the semiconductor; Γ can be modulated by the carrier injection.

The change of light intensity with current is linear for sufficiently small values of current. Then we may set

$$\Gamma \approx \Gamma_0 + \Delta\Gamma \qquad (7)$$

and

$$Q = Q_0 + \Delta Q; \qquad (8)$$

and we obtain

$$\Delta Q \approx JXH(1 - R)\exp(-W\Gamma_0)W\Delta\Gamma. \qquad (9)$$

Maximum value $(\Delta Q)_W$ is obtained by choosing the width of the bar

$$W = \frac{1}{\Gamma_0}. \qquad (10)$$

$$(\Delta Q)_W = \frac{JXH(1 - R)}{2.72} \frac{\Delta\Gamma}{\Gamma_0}. \qquad (11)$$

Since $\Delta\Gamma$ is proportional to Δn_0,² we have

$$\Delta\Gamma = \Delta n_0/p, \qquad (12)$$

where p is a constant $\sim 10^{16}$cm⁻² for silicon (see text). Replace n_0 in (12) by using (4) and insert the value of $\Delta\Gamma$ so obtained into (10) to obtain

$$(\Delta Q)_W = \frac{TX(1 - R)\Delta I/e}{2.72\ p} \sqrt{\frac{\tau}{D}}. \qquad (13)$$

Since only light passing within a distance $\sim L$ from the junction will be modulated, one shall choose $X \approx L$. Then from (13) and (5), the efficiency of light modulation may be calculated

$$\frac{(\Delta Q)_W}{\Delta I/e} = \frac{T(1 - R)\tau}{2.72 \cdot p}. \qquad (14)$$

The numerical factor $1 - R/2.72$ is of the order of 0.3. For $\tau \approx 10^{-4}$ sec, and $p \sim 10^{16}$cm⁻², efficiencies larger than unity may be obtained for light intensities larger than 3×10^{20} quanta/cm²sec (for light of near infrared about 20 watts/cm²).

Properties of the M-1740 *P-N* Junction Photocell*

JOHN N. SHIVE†

Summary—The *p-n* junction photocell has a sensitivity of 30 ma per lumen for light of 2,400 degrees K color temperature, corresponding to a quantum yield approximately unity in the spectral range from visible to the long wave cutoff at 1.8 microns. Dark currents of a few microamperes are observed at room temperature, with a temperature coefficient of about +10 per cent per degree C. Both dark and light currents exhibit saturation in the range from 1 to 90 volts applied. The frequency response is flat into the 100-kc region. Short-circuit noise currents are observed around 20 $\mu\mu$a in a 1-cps band at 1,000 cps. The photocell element is encapsulated in a plastic housing $\frac{1}{4} \times \frac{3}{16} \times \frac{3}{8}$ inch in dimensions.

1. Introduction

THE PHOTOCONDUCTIVE and photovoltaic properties of rectifying barriers between semiconductors and metals are well known and have been studied by many investigators. Only recently, however, has there been much study of the photoelectric properties of rectifying *p-n* junctions in semiconductors. In 1941 Ohl[1] described the photovoltaic behavior of a *p-n* junction in a silicon ingot. In 1947 Benzer[2] observed a photoconductive effect resulting from the illumination of a natural *p-n* junction in series with a rectifying point contact. He called this arrangement a "photodiode." In 1951 Goucher, Pearson, Sparks, Teal, and Shockley[3] published the results of some observations of the collection of photoelectrically liberated charges at *p-n* junctions in grown crystals of germanium. Shortly thereafter, Pietenpol[4] described a practical *p-n* junction photocell element. With this background the M-1740 photocell has evolved as a small, sensitive, rugged photoelectric device. It is the purpose of this paper to describe this photocell and to give as comprehensive a discussion as present knowledge permits of its electrical and optical characteristics, temperature behavior, noise, and reliability.

2. Description of the Cell

The photoelectric element consists of a bar of single-crystal germanium, approximately 0.030 by 0.030 by 0.125 inch in dimensions, having a *p-n* conductivity type transition midway between its two ends. Terminal leads are attached to the ends of the bar, and the element is encapsulated in a small rectangular block of plastic $\frac{1}{4} \times \frac{3}{16} \times \frac{3}{32}$ inch in dimensions. Light is incident upon the element through one of the $\frac{1}{4} \times \frac{3}{32}$ faces of the block, while the leads come out the opposite face. Lateral surfaces of the block are painted black, and the lead wire which is to be made positive in the biasing circuit is indicated by a red dot on the side of the block. Fig. 1 shows a sectional drawing of the photoelement and its encapsulation, while Fig. 2 shows a photograph of several completed M-1740 photocells.
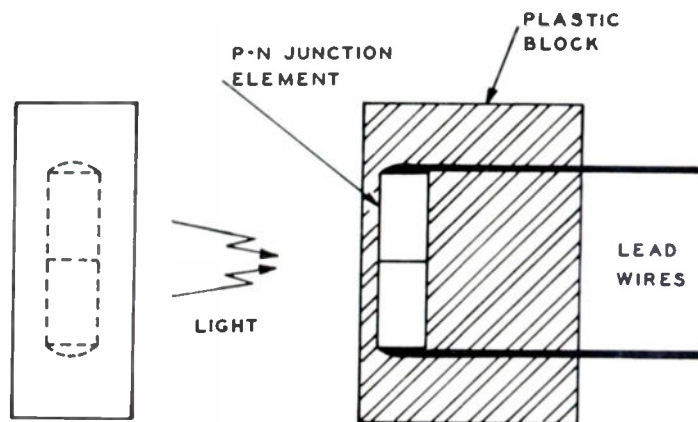
Fig. 1—Left, front-view drawing of photocell; right, side section drawing of element and capsule.
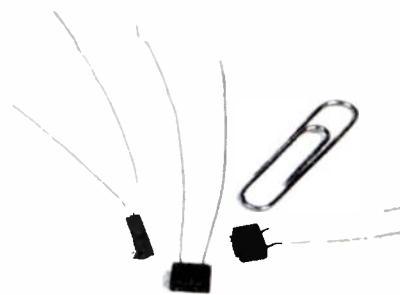
Fig. 2—Photograph of M-1740 photocells.

In normal use the cell is placed in a circuit in series with a dc biasing voltage source and a load across which the output signal is to be developed. The direction of the bias is such that the *p-n* barrier is operated in the blocking direction (*p*-type end negative). Under these conditions a small reverse leakage current flows even with the cell in the dark. When the cell is illuminated, the reverse current increases.

3. Spatial Distribution of Sensitivity

The response of a *p-n* junction element to light depends on how well the light is focused upon the *p-n* junction region. To be effective the light must fall within a few tenths millimeter of the junction line. Fig. 3 shows the response of a typical M-1740 photocell as a function of the position of a small light spot when the latter is traversed across the germanium along a line perpendicular to the junction. The width of this response curve at half maximum is about $\frac{3}{4}$ mm.

[1] R. S. Ohl, U. S. Patent No. 2,402,662.
[2] S. Benzer, "Excess-defect germanium contacts," *Phys. Rev.*, vol. 72, p. 1267; December, 1947.
[3] F. S. Goucher, G. L. Pearson, M. Sparks, G. K. Teal, W. Shockley; "Theory and experiment for a germanium *p-n* junction," *Phys. Rev.*, vol. 81, p. 637; April, 1951.
[4] W. J. Pietenpol, "*p-n* junction rectifier and photocell," *Phys. Rev.*, vol. 82, p. 120; January, 1951.

The shape of this curve is capable of simple physical explanation. The photocurrent is produced by the collection at the *p-n* junction of minority carriers (electrons in the *p*-type region and positive holes in the *n*-type region) liberated by the photoelectric process at and near the junction. Except in the junction layer itself there is negligible electric field, and most of the carriers reach the junction by diffusion. The farther away from the junction they are liberated the farther they must diffuse to be collected and the greater is the probability that they will disappear by recombination on the way. This accounts for the decay of sensitivity on both sides of the maximum, which is coincident with the junction.
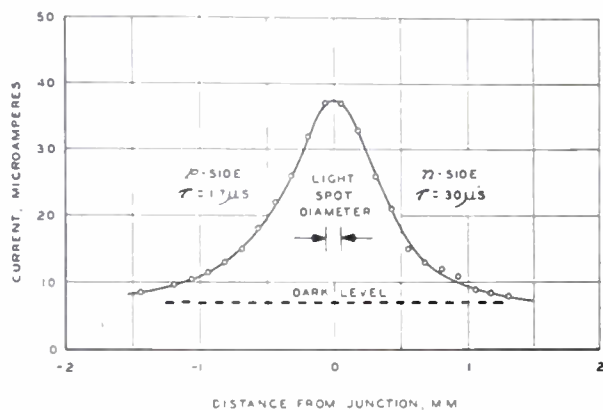
Fig. 3—Sensitivity profile of exposed face of
p-n junction element.

The curve of Fig. 3 exhibits a noticeable asymmetry. This is due in part to the fact that electrons and holes have different diffusion mobilities and in part to the difference in the lifetimes of electrons in the *p*-type material and holes in the *n*-type material. Goucher[2,5] has described an ingenious method for determining these lifetime values for any particular junction specimen by measuring the decrement of the response curve on the two sides of the maximum. Analysis of the curve of Fig. 3 gives a lifetime of 30 μsec for holes on the *n*-type side of the junction and 17 μsec for electrons on the *p*-type side of the photoelement depicted.

There is considerable expectation of increasing the responsive area of *p-n* junction photocells in the direction normal to the junction line. From the *p-n* junction theory[6] the breadth of the response curve is proportional to the square roots of the minority carrier lifetimes. Considerable progress has been made in the direction of increasing these lifetimes, and junction specimens have been produced[7] in which the lifetimes are an order of magnitude greater than those indicated above.

[5] F. S. Goucher, "Measurement of hole diffusion in *n*-type germanium," *Phys. Rev.*, vol. 81, p. 475; March, 1951.
[6] W. Shockley, "Electrons and Holes in Semiconductors," chap. 12, D. Van Nostrand Co., Inc., New York, N. Y.; 1950.
[7] W. Shockley, M. Sparks, and G. K. Teal, "*P-n* junction transistors," *Phys. Rev.*, vol. 83, p. 151; January, 1951.
[8] F. S. Goucher, "Photon yield of electron-hole pairs in germanium," *Phys. Rev.*, vol. 78, p. 816; June, 1950.

4. ELECTRICAL CHARACTERISTICS

The performance of the M-1740 photocell is described by the family of current-voltage characteristics given in Fig. 4. These characteristics are the result of averaging individual data for ten typical cells. They give the performance for the dark condition and for several different values of incident light flux. In the experiment from which these data were obtained the light was focused into a small spot about ½ mm in diameter, centered on the *p-n* junction, as determined by manipulating the position of the cell with respect to the light spot so as to maximize the output. The light source was a tungsten filament operated at about 2,400 degrees K color temperature.
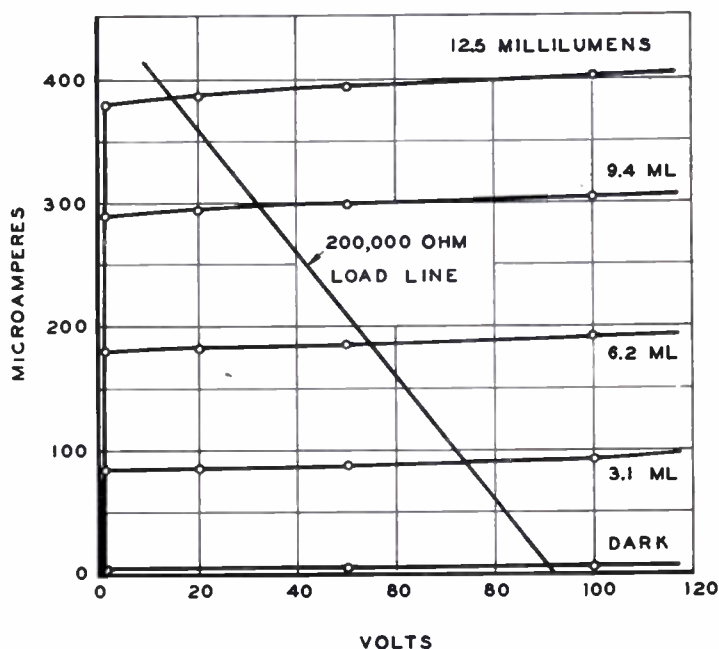
Fig. 4—Current-voltage characteristics,
average for ten M-1740 photocells.

In order for a photocell to be acceptable as an M-1740 unit its dark current at 90 volts must not exceed 20 μa. Most M-1740 photocells now being made have dark currents in the range from 1 to 5 μa at room temperature. These reverse leakage currents across the junction result from the constant thermal generation of new minority charges in the germanium within a few diffusion lengths of the junction and from their diffusion into and collection by the junction.

From a knowledge of the photon equivalent of the lumen, the shape of the spectral energy distribution of the source lamp, and the long wavelength photoelectric threshold of germanium it can be shown that the photocurrents given in Fig. 4 correspond to quantum efficiencies of about 1 hole-electron pair liberated per quantum absorbed by the photoelement at the junction. This agrees with Goucher's findings[8] that light absorption in the wavelength range from 1 to 2 microns in germanium results in electron-hole pair production with a quantum yield of approximately unity. There is thus no hope that the current response of a single *p-n* junc-

tion photocell can be increased by making the photoelectric activation process more efficient. Since these cells have no such mechanisms as those exhibited by point-contact phototransitors[9] or by multiple-junction devices[7] for multiplying the primary photocurrent, present sensitivities can be but little further improved.

5. AC Equivalent Circuit

A noteworthy feature of the characteristics of Fig. 4 is that they exhibit a current saturation in the range from 1 to 100 volts. This is a consequence of the fact that the diffusion of photocurrent carriers from the point of their generation to the junction occurs in essentially field-free space regardless of the voltage drop across the terminals of the bar, nearly the whole of which appears as a drop across the junction. The slopes of the curves of Fig. 4 in this saturation region correspond to ac impedances of the order of 10 megohms.

These characteristics are describable in terms of an equivalent circuit (Fig. 5, left) consisting of a junction resistance r_c in series with a germanium body resistance r_b, with a current generator shunting the r_c component. The current delivered by this generator is given by kl,

Fig. 5—Left, low-frequency ac equivalent circuit; right, high-frequency equivalent circuit.

where k is a constant of proportionality and l is the effective light flux on the photocell. The flux is usually given in millilumens at some specified color temperature, while k is generally expressed as ma per millilumen. The short-circuit output current is then given by

$$i_c = kl \frac{r_c}{r_c + r_b} = \kappa l,$$

where κ is a new constant of proportionality easily accessible to external measurement.[10] Since r_c is usually several megohms and r_b is about a hundred ohms, k and κ are nearly equal numerically, and can be so regarded for all practical purposes.

Any p-n junction photocell is specified when r_c, r_b and k (or κ) have been determined. The barrier resistance r_c is too high to measure conveniently by dynamic means. It is best determined from the slopes of the static characteristics. Numerical values of r_b can be computed from the resistivities and geometry of the ger-

[9] J. N. Shive, "A new germanium photoresistance cell," *Phys. Rev.*, vol. 76, p. 575; April, 1949; also "The Phototransistor," *Bell Labs. Rec.*, vol. 28, p. 289; June, 1950.

[10] This equivalent-circuit terminology was chosen to correspond with that used in describing transistors. In the latter case the externally measured current gain is denoted by $\alpha = i_c/i_e$ for ac short-circuited collector, while the internal equivalent-circuit generator is denoted by ai_e. The constants a and α are then related through the expression $a = \alpha(r_b + r_c/r_c)$.
In most cases a and α are only slightly different numerically.

manium element or measured by biasing the junction in the *forward* direction and determining the asymptotic resistance approached with increasing voltage in this direction.

From a practical point of view, however, there is only occasional need for knowing r_c and r_b. The former is numerically much greater than any load which one is likely to use with this device, while r_b is much smaller than any such load. Hence one may set up an approximate equivalent circuit in which r_c is considered infinite and r_b is considered zero. One is then left with the ac current generator kl operating directly into the output load. If this load is anywhere in the range from two thousand to five hundred thousand ohms, the approximation is valid to within 5 per cent. The measurement problem then becomes merely that of determining k for the photocell. (For the approximate equivalent circuit the distinction between k and κ disappears.) This is done simply by measuring the dc current into a short-circuiting load, first with the unit in the dark and then with a known light flux incident upon the junction, the difference being equal to kl. In current practice this is done in a circuit with a 90-volt bias and a flux of 6.25 millilumens of 2,400 degrees K color temperature light focused into a spot of $\frac{1}{2}$-mm diameter centered on the junction line. Values for k in the range from 0.03 to 0.04 ma per millilumen are obtained.

When the photocell is to be used with step-function light inputs or ac light signals having frequencies of the order of 100 kc or higher, the effect of the barrier capacitance must be taken into account. This capacitance is represented in the high-frequency equivalent circuit of Fig. 5 (right) as a capacitance in shunt with r_c. This circuit can be solved for the load current by simple analysis. If, as before, one introduces the practical assumptions that R_L is very much smaller than r_c and very much larger than r_b, then the load current reduces to $i \approx kl/(1 + jR_L C\omega)$. This expression is equal to kl at low frequency, remains substantially flat with increasing frequency until the second term in parenthesis becomes comparable with unity, then decreases with further increase in frequency according to the $1/f$ relationship, in quadrature with the light input.

Barrier layer capacitances of M-1740 photocells have not yet been measured directly. Calculations based on the p-n junction theory[11] give capacitances of a few micromicrofarads for such junctions back-biased at 50 volts. For a load resistance of 100,000 ohms the term $R_L C \omega$ becomes comparable with unity at frequencies of the order of a megacycle.

The simple barrier capacitance effects are usually dominated in practice by the effects of diffusion transit time of the carriers. Electrons or holes liberated some distance away from the junction may require microseconds or tens of microseconds to diffuse into the barrier and be collected. This process superposes an addi-

[11] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, p. 435; May, 1949.

tional lag upon that attributable to barrier capacitance. Transit time effects have been studied quantitatively by Anderson (unpublished), who measured the delay in response to short, steep light pulses incident upon a photocell at various distances away from the junction. His results show that as the illuminated spot is moved farther away from the junction the current collection pulses are more delayed, more attenuated, and more spread out. These experiments lacked resolution-in-time below about 2 μsec because of the shape of the light flashes and other experimental limitations. However, they do indicate that by masking down the photocell window so that light can fall no farther away from the junction than about 0.010 inch, the delay in response can be reduced to at least this value.

6. Temperature Dependence of the Characteristics

The dark current of a germanium *p-n* junction photocell increases by a factor of approximately 10 in going from a temperature of 20 degrees C to a temperature of 55 degrees C. This appears true regardless of what the room temperature dark current of the cell may be. Table I gives measurements of dark and light currents for a number of cells at these two temperatures. These tests were made at 24 volts dc with a flux of 6.25 millilumens for the light condition.

TABLE I

Unit No.	Dark current, μa		Light current, μa	
	20°C	55°C	20°C	55°C
1	3.7	35	190	270
2	8.6	91	160	240
3	1.8	19	180	252
4	3.1	26	206	292
5	12.4	108	196	268
6	4.7	52	164	242

Examination of these data shows that the increase in light current caused by temperature increase is larger than the dark-current increase for the same cell. Thus the sensitivity increases with increasing temperature. This is accounted for by temperature variations of mobility, diffusion constant, minority carrier lifetime.

7. Noise in *p-n* Junction Photocells

Measurements of noise in *p-n* junction photocells have been made by Montgomery. The quantity determined was the ac short-circuit noise current in a 1-cps band at 1,000 cps. Values for a typical photocell are given in Table II for different light-flux conditions and two different dc bias-voltage conditions.

TABLE II

DC supply voltage	Dark		Light		More light	
	dc μa	noise $\mu\mu a$	dc μa	noise $\mu\mu a$	dc μa	noise $\mu\mu a$
45	6.7	30	151	25	620	50
90	7.6	45	144	55	600	60

An observation of unusual interest is that, for a given bias voltage, while the dc current may increase by a factor of nearly 100 from the dark to the highest light

condition, the noise current changes by an amount hardly distinguishable from the uncertainties of measurement. Junction photocells have been examined for the appearance of their noise on an oscilloscope, and no irregularities have been observed over periods which, added together, aggregate several hours. The noise is well behaved in the sense that it does not exhibit bursts or spikes of amplitude greater than what is expected on the basis of smooth distribution.

It is interesting to compare quantitatively the noise power available from a *p-n* junction photocell with that from a point-contact phototransistor. Calculation from Table II for the *p-n* photocell gives around 2×10^{-15} watt per cycle at 1,000 cycles. For a point-contact photocell the corresponding figure is around 10^{-13} watt per cycle, giving a 17-db advantage for the junction device.

8. Reliability

Comprehensive life tests for observing the reliability of junction photocells over large periods of time under conditions of environmental abuse have been initiated but not yet completed. Prolonged exposure of the M-1740 photocell to a high-humidity environment results in an upward drift of dark current. The rate of increase is greater the higher the temperature. If the increase is not carried too far, the effect is reversible and the initial characteristics are restored when the unit is dried out. Present experience indicates that operation should be avoided above 60 degrees C and that shipping and storage exposure to temperatures of 70 degrees C or higher for more than a few days may result in a permanent upward shift in dark current.

An operational life test under conditions of ordinary room temperature and humidity was set up by Pietenpol more than two years ago. Except for the failure of one cell out of the set of 9, for assignable reasons of faulty encapsulation, the units have substantially the same characteristics as initially. Other tests under the same environmental conditions, but under different electrical and optical conditions, have been in operation for shorter times, with same favorable results to date.

Because the device is small, its heat dissipation is limited. In applications where large supply voltages are used and photocurrents of a milliampere or more are expected, the load resistance should be chosen large enough to keep the dissipation in the photocell itself below 50-mw continuous. Instantaneous dissipations such as those encountered in switching transitions may exceed one watt for a few milliseconds.

9. Conclusions

The characteristics of the M-1740 *p-n* junction photocell have been described. Its small size and weight, its low dark current, rapid response, and mechanical ruggedness recommend it for applications where these qualities are particularly desired in a semiconductor photocell. It is further characterized by low noise and high sensitivity. In its present form it is unsuited for use in extremes of temperature or humidity, but under ordinary conditions its operation is reliable.

tion photocell can be increased by making the photo-electric activation process more efficient. Since these cells have no such mechanisms as those exhibited by point-contact phototransistors[9] or by multiple-junction devices[7] for multiplying the primary photocurrent, present sensitivities can be but little further improved.

5. AC Equivalent Circuit

A noteworthy feature of the characteristics of Fig. 4 is that they exhibit a current saturation in the range from 1 to 100 volts. This is a consequence of the fact that the diffusion of photocurrent carriers from the point of their generation to the junction occurs in essentially field-free space regardless of the voltage drop across the terminals of the bar, nearly the whole of which appears as a drop across the junction. The slopes of the curves of Fig. 4 in this saturation region correspond to ac impedances of the order of 10 megohms.

These characteristics are describable in terms of an equivalent circuit (Fig. 5, left) consisting of a junction resistance r_c in series with a germanium body resistance r_b, with a current generator shunting the r_c component. The current delivered by this generator is given by kl,

Fig. 5—Left, low-frequency ac equivalent circuit; right, high-frequency equivalent circuit.

where k is a constant of proportionality and l is the effective light flux on the photocell. The flux is usually given in millilumens at some specified color temperature, while k is generally expressed as ma per millilumen. The short-circuit output current is then given by

$$i_c = kl \frac{r_c}{r_c + r_b} = \kappa l.$$

where κ is a new constant of proportionality easily accessible to external measurement.[10] Since r_c is usually several megohms and r_b is about a hundred ohms, k and κ are nearly equal numerically, and can be so regarded for all practical purposes.

Any p-n junction photocell is specified when r_c, r_b and k (or κ) have been determined. The barrier resistance r_c is too high to measure conveniently by dynamic means. It is best determined from the slopes of the static characteristics. Numerical values of r_b can be computed from the resistivities and geometry of the ger-

manium element or measured by biasing the junction in the forward direction and determining the asymptotic resistance approached with increasing voltage in this direction.

From a practical point of view, however, there is only occasional need for knowing r_c and r_b. The former is numerically much greater than any load which one is likely to use with this device, while r_b is much smaller than any such load. Hence one may set up an approximate equivalent circuit in which r_c is considered infinite and r_b is considered zero. One is then left with the ac current generator kl operating directly into the output load. If this load is anywhere in the range from two thousand to five hundred thousand ohms, the approximation is valid to within 5 per cent. The measurement problem then becomes merely that of determining k for the photocell. (For the approximate equivalent circuit the distinction between k and κ disappears.) This is done simply by measuring the dc current into a short-circuiting load, first with the unit in the dark and then with a known light flux incident upon the junction, the difference being equal to kl. In current practice this is done in a circuit with a 90-volt bias and a flux of 6.25 millilumens of 2,400 degrees K color temperature light focused into a spot of $\frac{1}{2}$-mm diameter centered on the junction line. Values for k in the range from 0.03 to 0.04 ma per millilumen are obtained.

When the photocell is to be used with step-function light inputs or ac light signals having frequencies of the order of 100 kc or higher, the effect of the barrier capacitance must be taken into account. This capacitance is represented in the high-frequency equivalent circuit of Fig. 5 (right) as a capacitance in shunt with r_c. This circuit can be solved for the load current by simple analysis. If, as before, one introduces the practical assumptions that R_L is very much smaller than r_c and very much larger than r_b, then the load current reduces to $i \approx kl/(1 + jR_L C\omega)$. This expression is equal to kl at low frequency, remains substantially flat with increasing frequency until the second term in parenthesis becomes comparable with unity, then decreases with further increase in frequency according to the $1/f$ relationship, in quadrature with the light input.

Barrier layer capacitances of M-1740 photocells have not yet been measured directly. Calculations based on the p-n junction theory[11] give capacitances of a few micro-microfarads for such junctions back-biased at 50 volts. For a load resistance of 100,000 ohms the term $R_L C\omega$ becomes comparable with unity at frequencies of the order of a megacycle.

The simple barrier capacitance effects are usually dominated in practice by the effects of diffusion transit time of the carriers. Electrons or holes liberated some distance away from the junction may require microseconds or tens of microseconds to diffuse into the barrier and be collected. This process superposes an addi-

[9] J. N. Shive, "A new germanium photoresistance cell," Phys. Rev., vol. 76, p. 575; April, 1949; also "The Phototransistor," Bell Labs. Rec., vol. 28, p. 289; June, 1950.

[10] This equivalent-circuit terminology was chosen to correspond with that used in describing transistors. In the latter case the externally measured current gain is denoted by $\alpha = i_c/i_e$ for ac short-circuited collector, while the internal equivalent-circuit generator is denoted by ai_e. The constants a and α are then related through the expression $a = \alpha(r_b + r_c/r_c)$.
In most cases a and α are only slightly different numerically.

[11] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," Bell Sys. Tech. Jour., vol. 28, p. 435; May, 1949.

tional lag upon that attributable to barrier capacitance. Transit time effects have been studied quantitatively by Anderson (unpublished), who measured the delay in response to short, steep light pulses incident upon a photocell at various distances away from the junction. His results show that as the illuminated spot is moved farther away from the junction the current collection pulses are more delayed, more attenuated, and more spread out. These experiments lacked resolution-in-time below about 2 μsec because of the shape of the light flashes and other experimental limitations. However, they do indicate that by masking down the photocell window so that light can fall no farther away from the junction than about 0.010 inch, the delay in response can be reduced to at least this value.

6. TEMPERATURE DEPENDENCE OF THE CHARACTERISTICS

The dark current of a germanium *p-n* junction photocell increases by a factor of approximately 10 in going from a temperature of 20 degrees C to a temperature of 55 degrees C. This appears true regardless of what the room temperature dark current of the cell may be. Table I gives measurements of dark and light currents for a number of cells at these two temperatures. These tests were made at 24 volts dc with a flux of 6.25 millilumens for the light condition.

TABLE I

Unit No.	Dark current, μa		Light current, μa	
	20°C	55°C	20°C	55°C
1	3.7	35	190	270
2	8.6	91	160	240
3	1.8	19	180	252
4	3.1	26	206	292
5	12.4	108	196	268
6	4.7	52	164	242

Examination of these data shows that the increase in light current caused by temperature increase is larger than the dark-current increase for the same cell. Thus the sensitivity increases with increasing temperature. This is accounted for by temperature variations of mobility, diffusion constant, minority carrier lifetime.

7. NOISE IN *P-N* JUNCTION PHOTOCELLS

Measurements of noise in *p-n* junction photocells have been made by Montgomery. The quantity determined was the ac short-circuit noise current in a 1-cps band at 1,000 cps. Values for a typical photocell are given in Table II for different light-flux conditions and two different dc bias-voltage conditions.

TABLE II

DC supply voltage	Dark		Light		More light	
	dc μa	noise μμa	dc μa	noise μμa	dc μa	noise μμa
45	6.7	30	154	25	620	50
90	7.6	45	144	55	600	60

An observation of unusual interest is that, for a given bias voltage, while the dc current may increase by a factor of nearly 100 from the dark to the highest light

condition, the noise current changes by an amount hardly distinguishable from the uncertainties of measurement. Junction photocells have been examined for the appearance of their noise on an oscilloscope, and no irregularities have been observed over periods which, added together, aggregate several hours. The noise is well behaved in the sense that it does not exhibit bursts or spikes of amplitude greater than what is expected on the basis of smooth distribution.

It is interesting to compare quantitatively the noise power available from a *p-n* junction photocell with that from a point-contact phototransistor. Calculation from Table II for the *p-n* photocell gives around 2×10^{-16} watt per cycle at 1,000 cycles. For a point-contact photocell the corresponding figure is around 10^{-13} watt per cycle, giving a 17-db advantage for the junction device.

8. RELIABILITY

Comprehensive life tests for observing the reliability of junction photocells over large periods of time under conditions of environmental abuse have been initiated but not yet completed. Prolonged exposure of the M-1740 photocell to a high-humidity environment results in an upward drift of dark current. The rate of increase is greater the higher the temperature. If the increase is not carried too far, the effect is reversible and the initial characteristics are restored when the unit is dried out. Present experience indicates that operation should be avoided above 60 degrees C and that shipping and storage exposure to temperatures of 70 degrees C or higher for more than a few days may result in a permanent upward shift in dark current.

An operational life test under conditions of ordinary room temperature and humidity was set up by Pietenpol more than two years ago. Except for the failure of one cell out of the set of 9, for assignable reasons of faulty encapsulation, the units have substantially the same characteristics as initially. Other tests under the same environmental conditions, but under different electrical and optical conditions, have been in operation for shorter times, with same favorable results to date.

Because the device is small, its heat dissipation is limited. In applications where large supply voltages are used and photocurrents of a milliampere or more are expected, the load resistance should be chosen large enough to keep the dissipation in the photocell itself below 50-mw continuous. Instantaneous dissipations such as those encountered in switching transitions may exceed one watt for a few milliseconds.

9. CONCLUSIONS

The characteristics of the M-1740 *p-n* junction photocell have been described. Its small size and weight, its low dark current, rapid response, and mechanical ruggedness recommend it for applications where these qualities are particularly desired in a semiconductor photocell. It is further characterized by low noise and high sensitivity. In its present form it is unsuited for use in extremes of temperature or humidity, but under ordinary conditions its operation is reliable.

Hall Effect*

OLOF LINDBERG†

Summary—The Hall effect is one of the rich sources of information about the conduction properties of semiconductors. The mobility and carrier concentration can be obtained from the Hall constant in conjunction with the resistivity; this cannot be done with the resistivity alone. The mobility is pertinent to the understanding of transistors since such things as high-frequency cut-off and the intrinsic current gain of the transistor are related to this property of germanium. The Hall effect and associated thermomagnetic and galvanomagnetic (Ettingshausen, Nernst, Righi-Leduc, and Ohmic) effects are discussed. The elimination of the effect of associated phenomena from the Hall measurement can be achieved in several ways. Some of the methods which are used today in the study of germanium are discussed, and typical apparatus is described.

INTRODUCTION

THE TRANSISTOR makes use of the special conduction properties of semiconductors to gain the advantages it has over the vacuum tube. To understand the advantages and limitations of the transistor one must understand these special conduction properties. For example, the high-frequency cut-off of the type A transistor can be predicted from the injected carrier's transit time which is related to the mobility. The intrinsic current gain in the type A transistor is related to the ratio of the mobility of electrons to the mobility of holes. In dealing with current flow in semiconductors, one must take account of the fact that other carriers than electrons may be present. The nature of these carriers (whether they are electrons or holes), the number of carriers per unit volume, and the ease with which they respond to an applied electric field (mobility) are all important quantities to know. The Hall effect provides a direct measurement of the carrier type and concentration and, in conjunction with the resistivity, yields the mobility. The density of carriers in semiconductors is determined in part by the density of foreign atoms in the material (for example, aluminum in germanium). Thus the Hall measurement is used to determine the density of impurity atoms. This measure of impurity density in high-purity samples is several orders of magnitude more sensitive than the best chemical procedures. Because of this, the measurement of the Hall constant is one of the basic procedures in experimental studies of semiconductors.

The Hall effect occurs when a substance carrying a current is subjected to a magnetic field perpendicular to the direction of the current. If the current is flowing in the x-direction and the magnetic field is applied in the z-direction, a potential gradient will appear across the sample in the y-direction. This transverse potential gradient is found to be proportional to the product of the current density in the sample and the applied magnetic field; the constant of proportionality is called the "Hall constant." Mathematically this can be expressed as follows:

$$\text{Grad } V_H = - RiH. \tag{1}$$

Grad V_H—(the transverse potential gradient) $= - E_H$, where E_H is the Hall field.

i—the current density.

H—the applied magnetic field.

R—the Hall constant.

If the sample is a rectangular solid of width a and thickness b and if the distribution of current is assumed uniform, (1) can be rewritten in terms of the total current I, the Hall voltage V_H, and the dimensions of the sample.

$$\text{Grad } V_H = \frac{V_H}{a} = \frac{- RIH}{ab}$$

$$V_H = \frac{- RIH}{b}. \tag{2}$$

The Hall effect can be explained on the basis of the particle nature of conduction. The current consists of streams of charged particles drifting under the influence of the electric field. When not under the influence of the magnetic field, the current flows longitudinally in the sample. On application of the magnetic field, the current carriers experience a force $e/c(\bar{v} \times \bar{H})$ and are swept to the edges of the sample. The charge continues to build up on the edges of the sample until the field due to the nonuniform charge distribution exerts a force equal to the deflecting force of the magnetic field.
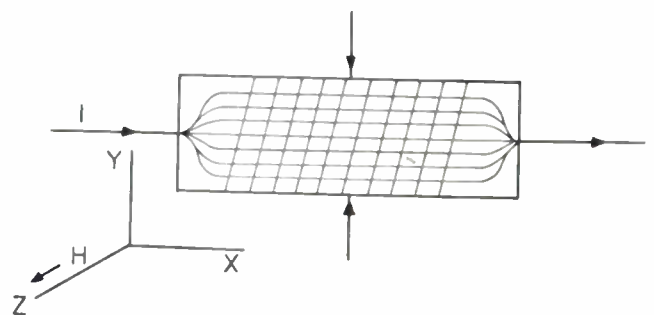
Fig. 1—Rotation of equipotentials by Hall effect. The Hall effect rotates the equipotentials so that they are no longer normal to the current flow.

Fig. 1 shows that the equipotentials are no longer normal to the current flow, but have been rotated through an angle θ, called the "Hall angle." Examination of the vector diagram of the fields in Fig. 2 (opposite page) shows that the Hall angle θ is determined by the following relation:

$$\tan \theta = \frac{E_H}{E_x} \simeq \theta \quad \text{for small angles.} \tag{3}$$

Now $E_H = Ri II$ and $E_x = \dfrac{i}{\sigma}$, thus $\theta = \dfrac{Ri II \sigma}{i} = R II \sigma$,

where σ is the conductivity.
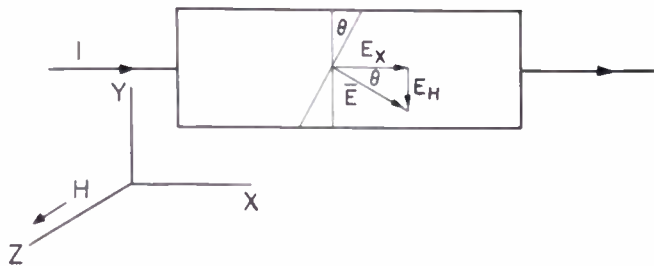
Fig. 2—Vector diagram for the Hall effect. The Hall angle θ is the angle of rotation of equipotentials. E_H is the Hall field if the carriers are electrons.

It is possible to calculate an expression for the Hall constant on the basis of the particle theory. In the derivation we assume that only electrons are present; but if only holes are present, the procedure is analogous. From the discussion above, the condition for the steady state is that the deflecting force of the magnetic field on a current carrier just equals the force exerted by the transverse electric field due to the charge build-up at the edges of the sample. This condition can be met mathematically by setting the y-component of the electric-field force equal in magnitude but opposite in sign to the force experienced by a charge moving in the magnetic field.

$$eE_H = e \frac{V_z II_z}{c}. \tag{4}$$

$$E_H = \frac{i_z II_z}{nec}, \text{ the current density } i_z = nev_z \tag{5}$$

where n the carrier concentration; i_z and II_z can be replaced by i and II since $i_y = i_z = II_z = II_y = 0$.

$$E_H = \frac{i II}{nec} = Ri II$$
$$R = \frac{1}{nec}. \tag{6}$$

The Hall constant derived by this method is valid only insofar as the particle picture of conduction is valid; that is, it applies to simple metals and impure semiconductors. To obtain an expression of more general applicability the mechanism of conduction must be examined more closely on the basis of a Boltzmann distribution of velocity of carriers. The result of such an investigation gives a relation which is generally used in semiconductor work when only one type of carrier is involved.[1]

[1] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y., p. 277; 1950.

$$R = \frac{3\pi}{8} \frac{1}{nec}. \tag{7}$$

From the measurement of the Hall constant, the carrier concentration can be determined. In an impurity-type semiconductor the carrier concentration is determined by the density of the dominant impurity. For example, in germanium at room temperature the intrinsic carrier concentration is approximately 5×10^{13} carriers per cubic centimeter, and impurity densities ten times as great are commonly found; thus it introduces a small error to attribute the entire carrier concentration to impurity atoms. From (6), the carrier concentration is proportional to $1/R$. This gives a quantitative measure of the impurity density. The value of this procedure can be seen from a calculation of the per cent impurity concentration for an impurity density of 4×10^{15}. There are approximately 4×10^{22} atoms of germanium per cubic centimeter; thus concentrations of one part in 10^7 are commonly met. To use chemical analysis to determine such impurity concentrations is a hopeless task. The Hall voltage, on the other hand, becomes larger for lower impurity concentration, and therefore more easily measured.

The carrier concentration does not give the complete picture of the conduction properties. The conductivity σ is related to the carrier concentration, the charge of the electron, and the ease with which the carriers move in an electric field (mobility). The latter is defined as the steady-state average velocity of the conducting particle (cm/sec) in unit electric field (1 volt/cm). In germanium the mobility will vary from 1,000 to 3,600 cm/sec per volt/cm. In a semiconductor with both negative and positive (holes) carriers the conductivity must be a function of the concentration and mobility of both holes and electrons.

$$\sigma = n|e|\mu_n + p|e|\mu_p \tag{8}$$

n—electron concentration (no/cc).
p—hole concentration (no/cc).
μ_n—electron mobility (up to 3,600 cm²/volt-sec in germanium).
μ_p—hole mobility (up to 1,700 cm²/volt-sec in germanium).

In the range where the concentrations of holes and electrons are about the same this expression for conductivity must be used. At the operating temperature of a semiconducting device, the concentration of one type carrier is much greater than the concentration of the other so that the conductivity will be dependent on only one term or the other in (8).

For p-type sample: $\sigma = p|e|\mu_p$.
For n-type sample: $\sigma = n|e|\mu_n$.

Carrier concentration can be determined from the conductivity if the mobility is known or a value is assumed. The Hall effect gives a method for determining the mobility and the carrier concentration for both n-type and p-type semiconductors without assumptions.

For an *n*-type sample the Hall constant is negative since *e* is negative.

$$R = \frac{-3\pi}{8} \frac{1}{n|e|c}. \tag{9}$$

For a *p*-type sample the Hall constant is positive because the charge of a hole is positive.

$$R = \frac{3\pi}{8} \frac{1}{pec}. \tag{10}$$

The Hall-angle formula differs for holes and electrons since mobility is different for holes and electrons.

$$n\text{-type: } \theta = \frac{H\mu_n}{c} \tag{11}$$

$$p\text{-type: } \theta = \frac{H\mu_p}{c} \tag{12}$$

Solving (9) and (10) for the carrier concentrations and (11) and (12) for the mobilities, the following expressions are obtained:

$$n = \frac{3\pi}{8} \frac{-1}{Rec} \qquad p = \frac{3\pi}{8} \frac{1}{Rec}$$

$$\mu_n = \frac{\theta c}{H} = |R|\sigma c \qquad \mu_p = \frac{\theta c}{H} = R\sigma c.$$

In measurements of the Hall voltage, certain associated effects give rise to potentials which must be corrected in order to avoid error in the measured value. The largest of these effects is the potential which appears because of the experimental difficulty in aligning the measuring probes on the same equipotential plane.
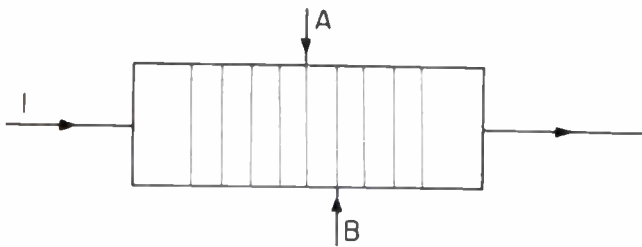
Fig. 3—Source of *IR* drop error. It is difficult to align the probes *A* and *B* so that no voltage will be measured in the absence of a magnetic field.

If Fig. 3, *A* and *B* are probes for measuring the Hall potential. With no field applied, the equipotentials are ideally planes perpendicular to the lines of current flow. If probes *A* and *B* are not exactly on the same equipotential, a potential will be measured between them, giving a constant error to the Hall voltage measured. This voltage can easily be of the order of magnitude of the Hall voltage itself. This effect is sometimes referred to as the "*IR* drop." The *IR* drop is dependent only on the current and the conductivity of the sample, not being affected by a reversal of magnetic field.

In addition to the *IR* drop, there are three thermomagnetic or galvanomagnetic effects, the Ettingshausen effect, the Nernst effect, and the Righi-Leduc effect. These effects will give rise to a temperature gradient or a potential gradient when either an electric current or a thermal current is subjected to a magnetic field perpendicular to the direction of current flow. A temperature gradient as well as a potential gradient will cause an error in the Hall-effect measurement. In Fig. 4, junction *A* is at temperature T_2 and junction *B* is at temperature T_1, which is less than T_2. Since the probe material is in general not the same as the sample material, the probes and the sample form a thermocouple and produce a voltage dependent in sign and magnitude on the materials of the probes and the sample.
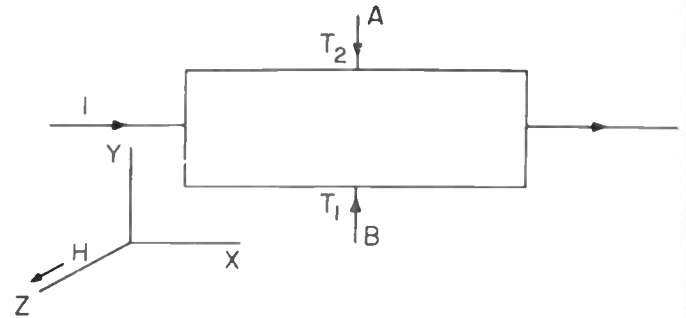
Fig. 4—Ettingshausen effect. The Ettingshausen effect causes the edge at probe *A* to be at temperature T_2 greater than T_1 the temperature of the other edge.

In the Ettingshausen effect, a permanently maintained temperature gradient will appear if an electric current is subjected to a magnetic field perpendicular to its direction of flow. The temperature gradient is found to be proportional to the product of the current density and the magnetic field.[2]

$$\frac{\Delta T}{a} = PiH \qquad i = \frac{I}{ab}.$$
$$\Delta T = \frac{PIH}{b} \tag{13}$$

ΔT—the difference of temperature between the edges of the sample.

i—current density.

I—total current.

H—magnetic field perpendicular to the direction of current flow.

P—Ettingshausen coefficient.

a—width of sample.

b—thickness of sample.

As in the case of the Hall effect, uniform current distribution and a rectangular solid sample have been assumed.

The Nernst effect and the Righi-Leduc effect are similar to the Hall effect and the Ettingshausen effect except that they are produced by a thermal current and a perpendicular magnetic field rather than an elec-

[2] P. W. Bridgman, "The Thermodynamics of Electrical Phenomena in Metals," Macmillan Co., New York, N. Y., pp. 135–138; 1934.

ric current and a perpendicular magnetic field.[3] In the Nernst effect, a potential gradient appears in the *y*-direction if a thermal current flows in the *x*-direction and a magnetic field is applied in the *z*-direction.[2]

$$E_N = Q \frac{wH}{K}. \tag{14}$$

E_N—Nernst transverse potential gradient.
w—thermal current density.
K—thermal conductivity of the sample.
Q—Nernst coefficient.

The Righi-Leduc effect produces a temperature gradient in the *y*-direction when a thermal current flows in the *x*-direction, and a magnetic field is applied in the *z*-direction.[2]

$$\frac{\Delta T}{a} = S \frac{wH}{K}. \tag{15}$$

ΔT—difference in the temperature between the edges of the sample.
S—Righi-Leduc coefficient.
a, w, H, K—as previously defined.

Associated with a thermal current, there exists an electron current. When two ends of a sample are at different temperatures, the high-velocity electrons at the hot end will diffuse toward the cold end; but for each high-velocity electron that diffuses away, a low-velocity electron must take its place so that a uniform charge distribution will be maintained.

A satisfactory description of the mechanism of the Ettingshausen effect on the basis of the classical theory of conduction by particles has not been given. However, following a method given by Bridgman,[3] some justification of the Ettingshausen effect can be made on a thermodynamic basis. Where an electric current flows longitudinally in a material and is subjected to a magnetic field perpendicular to the current flow, a transverse temperature gradient builds up because of a "thermomotive force." This temperature gradient is maintained as long as the current and field are maintained. When the temperature gradient builds up, a transverse thermal current must flow from the cold edge to the hot edge. The energy required for the thermomotive force to drive the thermal current against the temperature gradient must be supplied by the electric current in the sample. Because of the Nernst effect, the transverse thermal current produces a longitudinal potential; the electrical current flowing against this potential provides the energy to build up and maintain the Ettingshausen temperature gradient.

[3] In this discussion the assumption has been made that the thermal current and the electric current are independent of each other. In reality a part of the total thermal current arises from the electric field present and similarly a part of the total electric current arises from the temperature gradient present. This interdependence is complicated and it is beyond the scope of this article to discuss this effect. The electric current as used here is current as would be measured by a meter and the thermal current is that determined by the temperature gradient and the thermal conductivity.

The Righi-Leduc effect produces a transverse temperature gradient when a longitudinal thermal current flows in a magnetic field. As in the Ettingshausen effect a transverse thermal current will flow from the cold edge to the hot edge while the temperature gradient is being built-up. The thermomotive force set up by this transverse thermal current (Righi-Leduc effect) is in opposition to the longitudinal thermal current, and is therefore able to extract energy from the thermal current and maintain itself.

As has been explained above, the IR drop, the Ettingshausen effect, the Nernst effect, and the Righi-Leduc effect all result in potentials at the probes measuring the Hall potential. Consequently, to measure Hall constant, these sources of error must be eliminated or minimized. The most common method of making Hall measurements utilizes dc magnetic fields and dc currents. The experimental setup is illustrated in Fig. 4. The probes A and B for measuring the Hall voltage are placed in contact with the sample as shown. The current leads are soldered to the ends of the sample. A dc magnetic field is applied in the *z*-direction and a current flows in the *x*-direction. If we consider the probes to be perfectly aligned on an equipotential surface at no magnetic field and we also assume that there is no longitudinal thermal current flowing, then the potential measured at A and B will be the sum of the Hall potential and the thermocouple potential due to the Ettingshausen temperature gradient.

$$V_1 = V_H + V_E.$$

where V_E, the thermocouple potential due to the Ettingshausen effect, and V_H, the Hall potential, are both dependent on the current and the magnetic field. The only apparent way to separate the two in a dc measurement is to make the probes out of the same material as the sample, and this is usually difficult if not impossible in the case of semiconductors. In making dc Hall measurements in germanium, V_E is usually considered to be of the order of 5 per cent, and is neglected.

From the following considerations it is apparent that a longitudinal thermal current will flow. When a longitudinal electric current flows, there will be a Peltier effect at the junctions where the current leads are soldered to the sample. One junction will become heated and the other cooled, and a temperature difference in the *x*-direction will be maintained so long as the current flows in the junctions. Such a temperature gradient can also occur due to nonuniform temperature of the sample or its surroundings. In general, the combination of these effects can add up to an important longitudinal temperature difference, and consequently, a longitudinal thermal current will probably exist. Because of the presence of such a thermal current, the Nernst effect and the Righi-Leduc effect should also be considered. The potential E measured at the probes A and B will be

$$E = V_H + V_E + V_N + V_{RL}.$$

V_N = potential due to Nernst effect.

V_{RL} = thermocouple potential due to the Righi-Leduc temperature gradient.

If the sample current is reversed and the potential at the probes, A and B, is remeasured, then the potential will be

$$E_1 = - V_H - V_E + V_N + V_{RL}.$$

Since the Hall and the Ettingshausen effects are directly related to the current and the Nernst and the Righi-Leduc are independent of current, only the former change sign when the current is reversed. Subtract E_1 from E;

$$E - E_1 = 2(V_H + V_E)$$

$$V_H + V_E = \frac{E - E_1}{2}. \tag{16}$$

To take account of all first-order associated effects, the IR drop must also be considered. If the probes, A and B, were not properly aligned on equipotentials at no magnetic field (as is always the case when the experiment is performed), then the voltage measured at the probes A and B would be

$$E = V_H + V_E + V_N + V_{RL} + V_{IR}.$$

V_{IR} is the potential due to IR drop.

V_{IR} is a function of the current but not the field.

V_H and V_E are dependent on both I and H.

V_N and V_{RL} are dependent on H but not on I.

If a series of four measurements is taken reversing the magnetic field and the sample current in all possible combinations,[4] then a method is available for eliminating all the undesired associated effects except the Ettingshausen.

$$(+H, +I) \; E_1 = V_H + V_E + V_N + V_{RL} + V_{IR},$$
$$(+H, -I) \; E_2 = - V_H - V_E + V_N + V_{RL} - V_{IR},$$
$$(-H, -I) \; E_3 = V_H + V_E - V_N - V_{RL} - V_{IR},$$
$$(-H, +I) \; E_4 = - V_H - V_E - V_N - V_{RL} + V_{IR},$$

which yield

$$V_H + V_E = \frac{E_1 - E_2 + E_3 - E_4}{4}. \tag{17}$$

This method has been outlined in detail because it is a standard method of making Hall measurements in the presence of the associated thermomagnetic and galvanomagnetic effects. The method is not completely satisfactory since it doesn't separate the Ettingshausen effect from the Hall effect and since a series of four measurements requires sufficient time for possible changes to occur in the experimental conditions, especially if measurements are being made as a function of temperature.

[4] The measurements must be made rapidly so that the part of the temperature gradient due to the Peltier effect does not have time to reverse after the current is reversed.

These objections are overcome by making the measurements with an ac field and an ac sample current. Let the field and the sample current be the same frequency. The Hall potential from (2) is

$$- V_H = \frac{RIH}{b}.$$

Substitute the instantaneous value of I and H in (2);

$$V_H = \frac{R}{b} I' \sin \omega t H' \sin \omega t$$

$$V_H = \frac{RH'I'}{2b} - \frac{RH'I'}{2b} \cos 2\omega t.$$

Similarly, for the Ettingshausen temperature gradient

$$\Delta T = \frac{PIH}{b} \quad \text{in equilibrium}$$

$$\Delta T = \frac{PI'H'}{2b} - \frac{PI'H'}{2b} (\cos 2\omega t)(e^{-\omega/\omega_0}). \tag{18}$$

The exponential is due to the time required to build the temperature up to equilibrium; ω_0 is a constant of the material, inversely proportional to the heat capacity and dependent on certain other properties. The other effects will be ac, and can be represented as follows:

$$V_{IR} = A \sin \omega t$$
$$V_N = B \sin \omega t$$
$$V_{RL} = C \sin \omega t.$$

A, B, and C are the appropriate constants. The signal measured at the Hall probes will be the sum of several components. If only the dc component is measured, then the potential will be the sum of the Hall voltage and the voltage due to the Ettingshausen temperature gradient, with the result that the Nernst and Righi-Leduc and the IR drop have been eliminated as a source of error and only one measurement need be made.

The Ettingshausen effect has associated a certain time delay which is required for the temperature gradient to become established. This is accounted for by the exponential in (18). For example, in bismuth the time required for the temperature gradient to reach its maximum value has been measured by one observer to be about one minute. (This corresponds to $\omega_0 = 0.1$.) This makes it apparent that an ac Ettingshausen effect would not be detectable if the period were sufficiently short compared to the response time of the effect. This is the basis of a measurement method which will separate the Hall effect and the Ettingshausen effect. If the field and the current are of different frequencies, then all the resulting signals on the Hall probes will be ac

$$I = I' \sin \omega_1 t \quad H = H' \sin \omega_2 t$$

$$V_H = \frac{RIH}{b} = \frac{RI'H'}{2b} [\cos (\omega_1 - \omega_2)t - \cos (\omega_1 + \omega_2)t]$$

$$V_E = D \frac{PI'H'}{2b}\left[\left(\exp\left[\frac{-|\omega_1 - \omega_2|}{\omega_0}\right]\right)\cos(\omega_1 - \omega_2)t\right.$$
$$\left. - \left(\exp\left[\frac{-|\omega_1 + \omega_2|}{\omega_0}\right]\right)\cos(\omega_1 + \omega_2)t\right].$$

As in the previous case, b is sample thickness and ω_0 is a constant of the material. D is a constant to account for the thermoelectric effect, and depends on the probe material and the sample material.

$$V_{IR} = A \sin \omega_1 t.$$
$$V_{H} = B \sin \omega_2 t.$$
$$V_{RL} = C \sin \omega_2 t.$$

If the difference frequency is high compared to ω_0, selecting either the sum or the difference frequency with a suitable filter, the measurement will give the Hall voltage directly. This method gives the Hall voltage free from all associated first-order effects, and requires but a single measurement to do so. The Hall constant is

$$R = \frac{2bV_H'}{I'H'}.$$

The primes indicate peak values and b is the sample thickness.

There are other possible methods for making Hall measurements; these utilize various combinations of ac and dc for the field and the sample current. The methods described above are typical and an analysis of any of the other methods would be similar.
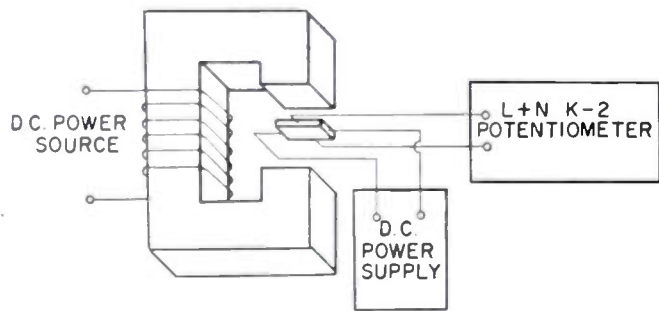
Fig. 5—Typical dc Hall effect apparatus. This is the most common and the simplest method of measuring Hall effect.

Fig. 5 illustrates schematically the apparatus for making dc Hall measurements. The dc electromagnet produces a field up to 5,000 gauss in the air gap. Usually a field of 1,000 gauss provides a Hall voltage large enough to be readily detected. The dc power supply, usually a battery, allows a sample current of 1 or 2 ma. The current should be small to produce as little I^2R heating of the sample as possible and to keep the Peltier heating at the junctions small, but it must be large enough to produce a detectable Hall voltage. A Hall voltage of about 100 μv is desirable if it is to be measured with a potentiometer. If the results of a Hall measurement are as follows,

$$V_H = 100 \ \mu v$$
$$H = 1,000 \ \text{gauss}$$

$$I = 1 \ \text{ma}$$
$$b = 1 \ \text{mm},$$

then

$$R = \frac{-V_H b}{HI} \times 10^8 = \frac{-10^{-4} \ \text{volts} \ 10^{-1} \ \text{cm}}{10^3 \ \text{gauss} \ 10^{-3} \ \text{amp}}(10^8)$$

$$R = -1,000 \frac{\text{cm}^3}{\text{coulomb}}.$$

The factor 10^8 converts voltage and current from electrostatic units to practical units. Equation (7) determines the carrier concentration;

$$n = \frac{3\pi}{8Rce} = \frac{1.17}{10^3 \times 1.6 \times 10^{-19}} = 7 \times 10^{15} \ \text{carriers/cc}.$$

In (7) the electronic charge is in electrostatic units; in the expression used above, the charge is in the practical system of units (coulombs). The carriers are electrons since the Hall constant is negative.
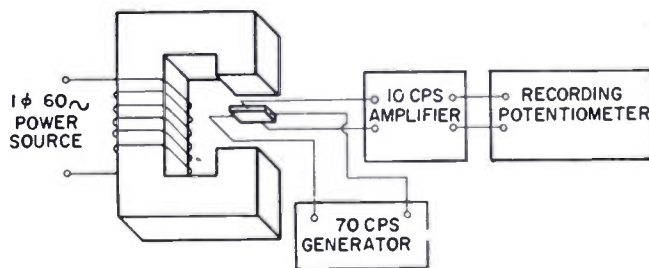
Fig. 6—Typical ac Hall effect apparatus.

Fig. 6 shows an apparatus for making ac Hall measurements. The electromagnet is supplied by a single-phase, 60-cycle power source, and produces a magnetic field in the gap up to about 5,000 gauss. The 70-cycle generator produces a current of 1 or 2 ma in the sample. The Hall-voltage component (the difference frequency) is amplified by a selective amplifier and the output fed into a measuring device, such as a recording potentiometer. The system is required to be stable so that accurate calibration can be obtained. This method is especially adapted to making measurements as a function of temperature where a continuous record of the variation of Hall voltage is desired.

The measurement of the Hall effect may be made in a variety of ways. The ac Hall apparatus is probably the best type for rapid determinations. The measurement is extremely useful in the determination of impurity concentration as such a deduction requires no assumptions as to other properties of the material.

BIBLIOGRAPHY

1. H. B. Callen, "Application of Onsager's reciprocal to thermomagnetic and galvanomagnetic effects," *Phys. Rev.*, vol. 73, p. 1349; 1948.
2. H. B. Callen, "Note on the adiabatic thermomagnetic effects," *Phys. Rev.*, vol. 85, p. 16; 1952.
3. L. L. Campbell, "Galvanomagnetic and Thermomagnetic Effects," Longmans, Green and Co., New York, N. Y.; 1923.
4. A. Sommerfeld and N. H. Frank, "The statistical theory of thermoelectric, galvanomagnetic, and thermomagnetic phenomena in metals," *Rev. Mod. Phys.*, vol. 3, p. 1; 1931.

Measurement of Minority Carrier Lifetime in Germanium*

L. B. VALDES†, ASSOCIATE, IRE

Summary—A method for measuring the lifetime of minority carriers in germanium is described. Basically, it consists of liberating the carriers optically on a flat face of a crystal and measuring the concentration of minority carriers as a function of distance from the point of liberation. The mathematical model is analyzed and experimental results are presented here.

I. THE PROBLEM

DESIGN THEORIES indicate that the resistivity and lifetime of minority carriers are essential properties of the bulk material to be used for the fabrication of transistors and other semiconductor devices. It is desirable to measure lifetime in order to understand its effect on the functions performed by a transistor and to be able to determine from theory what kind of material should be used for transistor manufacture. The role which lifetime plays in determining the reverse collector impedance and alpha under normal operating conditions and the way in which it affects other phenomena, such as minority carrier injection from the base, in both junction and point-contact transistors, must not be underestimated.[1]

However, in order for information such as this to be valuable it is necessary to be able to measure lifetime in a way which is reliable, convenient, and reproducible. The method to be described, which was developed at Bell Telephone Laboratories by Morton and Haynes, accomplishes precisely that.

II. METHODS AND ASSUMPTIONS

This measurement method, known as the Morton-Haynes method, is a scheme for measuring the diffusion length of minority carriers in single-crystal germanium. The physical model for it is shown in Fig. 1. Essentially it consists of liberating hole-electron pairs by light on the surface of a germanium crystal and measuring the concentration of minority carriers as a function of distance from the point at which they are liberated.

In the physical model the surface is shown to be illuminated with a narrow rectangle of light and a detection electrode is located a distance r from the center of the illuminated area. In order to treat this model certain assumptions must be made. It must be assumed that the illuminated rectangle is an infinitely long line, which is a good approximation if the width of the rectangle is less than one-fifth r_{min} and the length is at least four times r_{max}. If surface recombination is assumed to be negligible, the measured lifetime is the volume lifetime, and the injected carriers flow along radial lines. The electric field of the collector electrode must be small and concentrated in the immediate vicinity of the electrode; otherwise, the field will affect the flow of injected

Fig. 1—Physical model

carriers, which for simplicity should be due only to diffusion. Finally, the illuminated area and the electrode must be far from the sides of the crystal, or carriers will be reflected from those surfaces and produce a false reading. By "far" is meant distances greater than 1.5 times the diffusion length of the minority carriers.
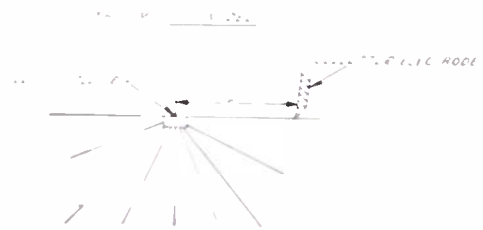
Fig. 2—Mathematical model.

The mathematical model is presented on Fig. 2. Starting from the diffusion equation for holes in n-type material

$$\nabla^2 p - \frac{p}{L_p^2} = 0,$$

transforming it to cylindrical co-ordinates, since radial flow only has to be considered and setting the boundary conditions that

$$p = p_0 i H_0^{-1}\left(\frac{ir_0}{L_p}\right) \quad \text{at} \quad r = r_0$$

and

$$p = 0 \quad \text{at} \quad r = \infty,$$

one finds the solution to be

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

$$p(r) = p_{0i} H_0{}^{-1}\left(i\,\frac{r}{L_p}\right),$$

where p = added hole density and L_p = diffusion length for holes. This solution is expressed in the form of a Hankel function of pure imaginary argument, and was obtained because the diffusion equation in cylindrical co-ordinates takes a form of Bessel's equation. See Appendix I for more details.

In order to reproduce this physical model in an experimental setup it is necessary to treat the surface of the material whose lifetime is to be measured in such a way that a low surface-recombination velocity is obtained. This is accomplished by etching the surface with CP_4 etch and then treating it with antimony oxychloride.[2]

The method used consists of illuminating the surface with a rectangle of light that is interrupted at a low audio-frequency rate. The carrier concentration is measured by a collector electrode biased in the reverse direction since the current through such an electrode is linearly related to the minority carrier density.[3] The use of chopped light permits measurement with an ac voltmeter of the voltage across a resistor in series with the collector, thus separating the minority carriers liberated optically from the steady-state concentration. To collect the carriers more readily at the electrode and increase the signal-to-noise ratio it is desirable to electrically form the point-contact electrode.

A schematic of the equipment is in Fig. 3. The light
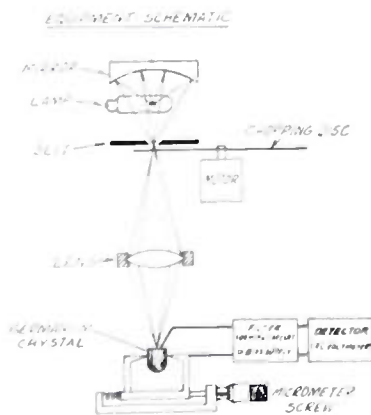
Fig. 3—Equipment schematic.

from a tungsten-ribbon lamp is focussed on a slit and interrupted as a square wave by means of a chopping disc driven by a synchronous motor. The image of the slit is projected on the surface of the germanium crystal by a lens whose position is adjustable. Baffles are provided inside the tube supporting the lens to reduce the

[2] J. R. Haynes and W. Shockley, "The mobility and life of injected holes and electrons in germanium," *Phys. Rev.*, vol. 81, no. 5, pp. 835–843; March 1, 1951.
[3] J. Bardeen, "Theory of relation between hole concentration and characteristics of germanium point contacts," *Bell. Sys. Tech. Jour.*, vol. 29, no. 4, pp. 469–495; October, 1950.
W. Shockley, G. L. Pearson, and J. R. Haynes, "Hole injection in germanium-Quantitative studies and filamentary transistors," *Bell Sys. Tech. Jour.*, vol. 28, no. 3, pp. 344–366; July, 1949.

scattered light reaching the sample. The collector electrode is placed on the surface by a micromanipulator, which also has provisions for moving the position of the crystal and electrode with respect to the stationary optical system. The electrode is adjusted to come down onto the surface at a 10° to 30° angle with respect to the normal line in order to prevent its shadow on the illuminated area.
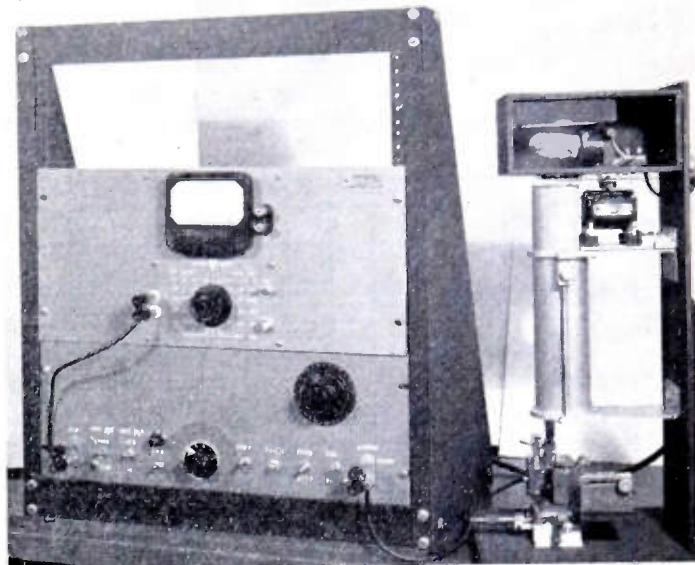
Fig. 4—Lifetime measuring equipment.

The filter, forming circuit, and bias supply are contained in a panel with switches for reversing the polarity of the bias depending on whether the sample is of n- or p-type material. Forming is carried out in steps, increasing the pulse in intensity until a suitable signal-to-noise ratio is obtained. The use of a narrow-band filter preceding the detector gives a further improvement in signal-to-noise ratio. Fig. 4 depicts the equipment.

III. RESULTS

A typical plot of the experimental results is shown in Fig. 5. The measured voltage variation across the resistor in series with the detection electrode is labelled signal voltage and used as an ordinate. The abscissa is the distance r between the collector electrode and the light. The theoretical curve is one for a value of $L_p = 94 \times 10^{-3}$ cm, corresponding to a hole lifetime of 200 μs. This scheme of matching curves is used for lifetime computations. The departure between theory and experiment occurs at small values of r/L_p, where the assumptions regarding thin line width and radial carrier flow do not hold.

The other extreme where errors may be introduced is at large distances. When the signal level decreases until it approaches approximately 3 times the noise level, the readings become unreliable. The same is true in long-lifetime samples where the carrier concentration at the bottom or the side of the sample increases appreciably and changes the diffusion gradient. See Appendix II for more detailed discussion of limitations of this method.
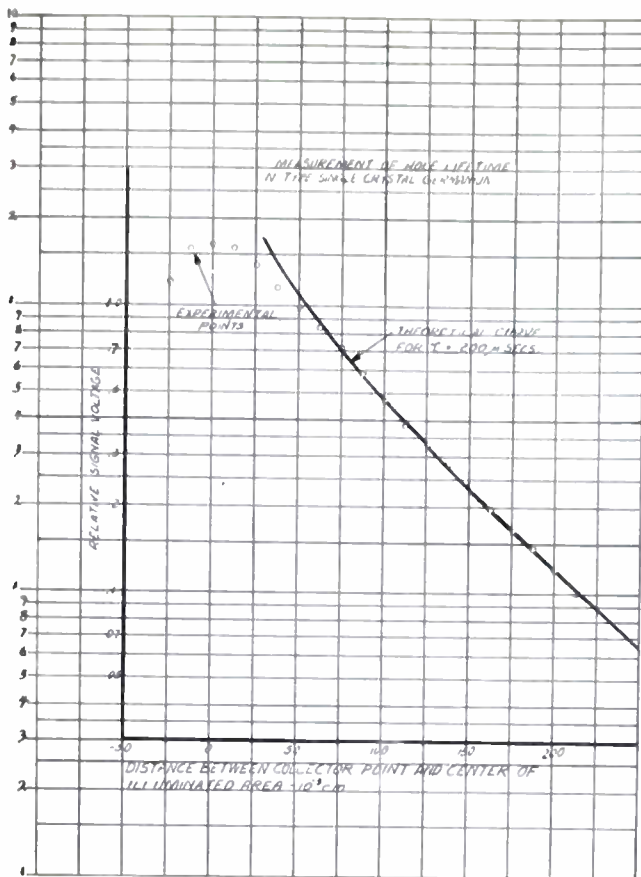
Fig. 5—Experimental results.

IV. Conclusions

This method of lifetime measurements is simple, versatile, reliable, and reasonably reproducible. It is simple because the time involved in the measurements is small and a minimum of preparation is required for the sample. The versatility is exemplified by the ability to measure lifetimes from a few microseconds to several hundred microseconds, limited in one case by the signal-to-noise ratio and in the other by the amount of surface recombination and the finite sample size. The reproducibility of the measurements made on a surface with the same surface treatment is indicated by a coefficient of variation less than 10 per cent, although the variabilities introduced by treating the surface again are generally greater and may give a coefficient of variation as large as 25 per cent.

V. Acknowledgments

Appendix I

Equations Governing the Flow of Minority Carriers in Germanium and the Calculation of Lifetime

The purpose of this section is to indicate a solution to the diffusion equation with the geometry of Fig. 2 and the assumptions outlined in the text and to discuss the

method of calculating the lifetime from the experimental data. For simplicity only, the flow of holes in n-type material will be considered since the solution applies equally well to electrons in p-type material if the appropriate diffusion constant is used.

The diffusion equation for holes is

$$\nabla^2 p - \frac{p}{L_p{}^2} = 0, \tag{1}$$

where p is the added hole density and L_p is a constant which equals the diffusion length for holes.[4] This equation can be expressed in cylindrical co-ordinates. Taking the $r=0$ axis to correspond to the illuminated line and using the assumption of radial flow, the equation becomes

$$\frac{d^2p}{dr^2} + \frac{1}{r} \frac{dp}{dr} - \frac{p}{L_p{}^2} = 0, \tag{2}$$

where p is a function $p(r)$ of the radial distance from the illuminated area only. The assumption of zero recombination velocity permits the surface of the germanium crystal to be considered perfectly reflecting, which simplifies the problem considerably. It is thus possible to solve the case of the liberation of holes along a line in the center of an infinite volume in order to get an expression for the density of holes along the surface of a supposedly semi-infinite crystal of germanium.

If, by definition, $x=r/L_p$, (2) becomes

$$\frac{d^2p}{dx^2} + \frac{1}{x} \frac{dp}{dx} - p = 0, \tag{3}$$

which is a form of Bessel's equation having a solution

$$p(r) = p_0 i H_0{}^{(1)}(ix). \tag{4}$$

This is a solution that vanishes as $r \to \infty$ and goes to infinity at $r \to 0$. The zero-order Hankel function of pure imaginary argument indicated in (4) cannot describe the actual hole density at very small values of r because (1) does not apply to the illuminated area and cylindrical flow is not realized in an illuminated area of finite width. The neglect of surface recombination also adds to the discrepancy. Errors at these small distances, however, do not invalidate the solution at larger distances and the boundary condition is determined by setting the hole concentration at some arbitrary distance r_0, where cylindrical flow exists, equal to $p_0 i H_0{}^{(1)}(ir_0/L_p)$.

Because the absolute values of $p(r)$ can not be readily obtained from theoretical considerations, the experimental method used for determining the lifetime, or diffusion length, of the unknown sample is based on a comparison of the slopes of the experimental and theo-

etical curves. Experimental data is plotted on semilog paper and the slope at any arbitrary distance r' is measured. This slope is equal to $d[\ln p(r')]/dr$, which is

$$\frac{d \ln p(r)}{dr} = \frac{d[\ln i H_0^{(1)}(ix)]}{dr} = \frac{d[\ln i H_0^{(1)}(ix)]}{dx} \frac{dx}{dr}; \quad (5)$$

and since $dx/dr = 1/L_p$,

$$\frac{d \ln p(r)}{dr} = \frac{1}{L_p} \frac{H_1^{(1)}(ix)}{i H_0^{(1)}(ix)}$$

= experimental slope on natural log paper.   (6)

These Hankel functions can be found in tabulated form.[5] Since $x = r'/L_p$, where r' is known from the point at which the slope was measured in the experimental data, the only unknown is L_p. From L_p the lifetime is computable as

$$\tau_p = L_p^2/D_p. \quad (7)$$

For calculating the lifetime of electrons in n-type material the same procedure is followed, except that the diffusion length thus obtained is L_n and the diffusion constant D_n has a different value. At room temperature the values of the diffusion constants in germanium are approximately[6]

$$D_p = 44 \text{ cm}^2/\text{sec} \qquad D_n = 92 \text{ cm}^2/\text{sec}.$$

Appendix II

Limitations of the Method

The purpose of this section is to justify the assumptions made in order to solve the problem of lifetime measurements and to indicate the order of magnitude of the errors involved.

One of the basic assumptions was to neglect surface recombination. Suppose that instead it is assumed that the surface-recombination velocity is 200 cm/sec. This happens to be a typical value obtained for holes after an n-type germanium sample has been subjected to the CP4 etch and the antimony oxychloride treatment. Surface recombination produces two important effects: a distortion of the cylindrical flow conditions by having carriers flow towards the surface and a faster decay of the carrier concentration as a function of distance. The former is more important at small values of r and the latter is more important at large values of r, where one-dimensional flow conditions would be approximated. These effects impose a maximum value of about 500 μs for reasonably accurate measurements of the carrier lifetime by this method.

The electric field of the collector electrode is probably

[5] E. Jahnke and F. Emde, "Tables of Functions with Formulae and Curves," 3rd. ed., Dover Publications, New York, N. Y., pp. 236–243; 1943.
[6] J. R. Haynes and W. Shockley, "The mobility and life of injected holes and electrons in germanium," *Phys. Rev.*, vol. 81, no. 5, pp. 835–843; March 1, 1951.

one of the most important and most neglected sources of error. Assuming radial current flow away from the electrode along the body of the semiconductor, the electric field introduced by the collector current can be calculated. With a current of 0.5 ma and 3 ohm-cm material the electric field at 0.050 cm away from the point is 0.09 volts/cm. With a hole mobility of 1,700, the transport velocity for holes due to the field at this point is 150 cm/sec. It thus appears that at reasonable distances the collector field may be neglected. Such error will, of course, necessarily appear at close spacings and, in fact, it accounts in part for the deviation between theory and experiment which is observed in Fig. 5 at distances less than 0.060 cm.

The necessity of having the illuminated area and the collector electrode far from the side boundaries of the crystal is not difficult to visualize in terms of the one-dimensional flow of carriers. Liberated minority carriers can diffuse towards the boundary of the sample and build up a higher than normal concentration, thus disturbing the normal diffusion of carriers throughout the sample. Mathematically, this is equivalent to assuming that the carriers are reflected at the boundary and eventually collected. If the distance travelled by the reflected path is less than $3L$ longer than by the straight path, the reflected carriers become a significant fraction of the total number present under the collector electrode. L is used here to indicate the diffusion length of whatever type of carrier is being considered.

The illuminated area has a width W and a length l. It was stated that this area approximates an infinitely thin and long line if

$$W \leq \tfrac{1}{5} r \text{ min}$$

and

$$l \geq 4r \text{ max}.$$

These values were arrived at from a consideration of the actual lifetime and geometry of the samples encountered in use. For instance, the width consideration can be determined by integrating the effect of several lines of light next to each other and spread over a distance W on the resulting signal a distance r away. If the minority carrier density obtained this way differs significantly from the value which is calculated on the basis of a single line of light, the spacing is too close. A more precise check on the length of the line can be made at a specific distance r by calculating the decay of the minority carriers traveling on a straight line from the end of the illuminated line to the detection point and comparing it to those diffusing towards the collector along a normal to the line. More exactly, it could be stated that if

$$e^{-\frac{\sqrt{l^2/4+r^2}-r}{L}} \ll 1$$

the illuminated area is adequately long.

Theory of Alpha for *P-N-P* Diffused Junction Transistors*

EARL L. STEELE†, ASSOCIATE, IRE

Summary—Equations are developed for the emitter and collector currents for this type of transistor, and the resulting expressions are then used to obtain the current-gain factor. The low-frequency value of the current-gain factor and its high-frequency cutoff value are shown to depend strongly on the width of the base region, the behavior at high frequencies being better for small base width. The high-frequency behavior when used in grounded emitter applications depends more directly on the lifetime of holes in the germanium and shows only a second-order dependence on the base width; the lower the lifetime the higher the frequency cutoff value.

Introduction

THE TRANSISTORS which form the basis for the discussion to follow differ from the devices made by other methods in both intrinsic properties and geometry. The general characteristics are very similar to those described in the literature,[1,2] but the new geometry brings into the picture some modified boundary conditions in the design equation. The necessary equations are developed in the following discussion.

General Description

A brief description of the transistor will be given with the aid of Fig. 1. The unit consists of emitter and collector regions of *p*-type germanium and a base region of *n*-type germanium. Ohmic contacts are made to these three regions and voltage biases applied as shown. The three regions are separated by diffused-type *p-n* junctions which, in general, are rather abrupt. These junctions arise from acceptor-type impurities diffusing
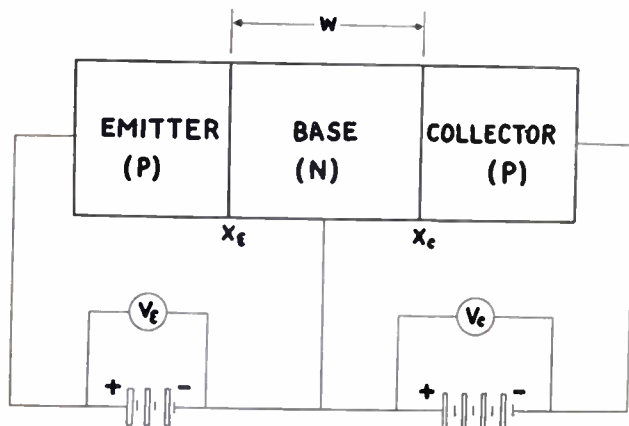
Fig. 1—Schematic diagram or *p-n-p* transistor with bias supplies shown.

[1] W. Shockley, "Theory of *p-n* junction in semiconductors and *p-n* junction transistors, *Bell Sys. Tech. Jour.*, vol. 28, p. 435; 1949.

[2] W. Shockley, M. Sparks, and G. K. Teal, "*p-n* junction transistors," *Phys. Rev.*, vol. 83, p. 151; 1951.

into relatively high-purity, single-crystal germanium.[3] The diffused material forms the barrier region behind which exists a region of *p*-type germanium, usually very thin. One of these *p*-type regions exists on either side of the *n*-type region, thus forming the transistor structure.

In this device the conduction is of the deficit type, that is, by holes. The potential-energy diagram for these holes in the above-described device is shown in Fig. 2.
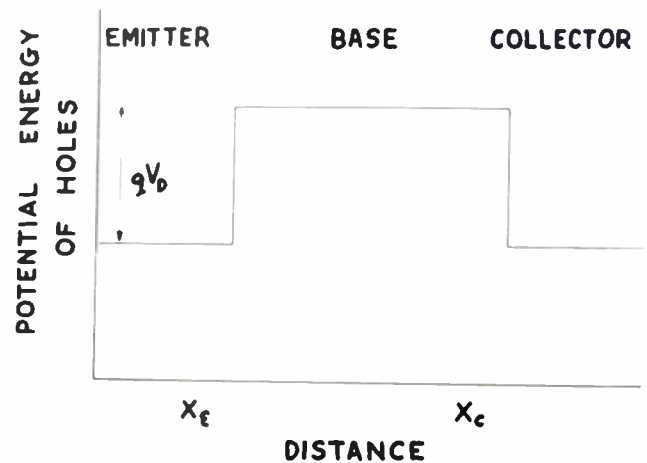
Fig. 2—Potential-energy diagram for holes with the transistor in thermal equilibrium, without bias voltages applied.

The holes in the emitter region can be treated as a gas of free particles with a Boltzmann-type energy distribution. Some of these have energies high enough to lift them up the potential hill into the base region. Now, if a voltage is applied to the emitter, making it positive relative to the base, the potential hill will be decreased and more holes can get up this hill and into the base region. This is called "hole injection." Since there is essentially no field in the base region away from the junction, the holes will drift away from the emitter junction by a diffusion process influenced only by the concentration gradient of the holes and not to any appreciable extent by a field-induced drift. As these holes drift, some will be lost by recombination with electrons, but most of them will reach the collector junction.

The collector voltage is applied so as to make the collector negative relative to the base. This causes an increase in the height of the potential wall at the collector junction. It will be seen that this forms a very good sink for holes so that once they arrive at this junction they fall through the potential drop and are collected.

Of course, at the same time as holes are being injected into the base region, electrons are being injected

[3] R. N. Hall and W. C. Dunlap, "*P-N* junctions prepared by impurity diffusion," *Phys. Rev.*, vol. 80, p. 467; 1950.

rom the base into the emitter region. As it turns out, hese electron effects are small but not negligible, and ill be taken into account in the calculations. This, however, does not alter the basic nature of the device opration.

BASIC CONSIDERATIONS

The flow of carrier through the base region can be decribed by the diffusion equations for the one-dimenional case as follows:

$$I(p) = -qD_p\frac{dp}{dx} + qp\mu_p E \tag{1}$$

$$\frac{dp}{dt} = -\frac{(p - p_B)}{\tau_B} - \frac{1}{q}\frac{dI}{dx}. \tag{2}$$

here the charge on holes (q), hole current $(I(p))$, hole ensity (p), diffusion constant (D_p), lifetime of holes in ase before recombination $(\tau_B$, hole mobility (μ_p), and lectric field (E) are related as shown. Quantity (p_B) the number of holes normally in the base region nder thermal equilibrium conditions. The physical neaning of (2) can be given as follows: The rate at hich hole density in a small region changes (dp/dt) is qual to number of holes per second lost by recombination, $((p-p_B)/\tau_B)$, plus the net number per second hich flow away as current $[(1/q)(dI/dx)]$. The recombination rate is assumed proportional to the density of njected carriers.

The assumptions which will be used in solving the bove equations can be rather simply stated.

1. The electric field (E) can be neglected in the regions away from the junctions. All the voltage applied between two regions appears only across the junction between those regions.
2. The junctions are very abrupt, we shall assume them to be step functions.
3. The density of holes in the base region is small compared to the electron density there. This means that physical parameters, such as mobility and lifetime, are not influenced by the injected carriers. Also, the conductivity in a given region is governed primarily by the majority carriers in that region and not by minority or injected carriers; that is, in the base the conductivity is due to electrons only and in the emitter and collector the conductivity is due only to holes.
4. Carrier densities are small compared to density of states in germanium and the energies are sufficiently large to enable one to use the Boltzmann approximation to Fermi statistics. This is a valid assumption for germanium at room temperature.

The boundary conditions can now be ascertained. Let he density of holes in the emitter region be (p_e); then nder thermal equilibrium conditions the hole density n the base region (p_B) can be obtained from the Boltzmann law by

$$p_B = p_e e^{-qV_D/kT}, \tag{3}$$

where the diffusion potential (V_D) is shown in Fig. 2. If an emitter voltage (V_e) is now applied as shown in Fig. 3, the hole density in the base region at the emitter junction (x_e) will be increased to (p_1), given by

$$p_1 = p_e e^{-q(V_D - V_e)/kT} = p_B e^{qV_e/kT}. \tag{4}$$

The same sort of argument can be applied at the collector junction (x_c). Thus,

$$p_2 = p_B e^{-qV_c/kT}, \tag{5}$$

where the collector voltage (V_c) is applied so that the collector is negative to the base.
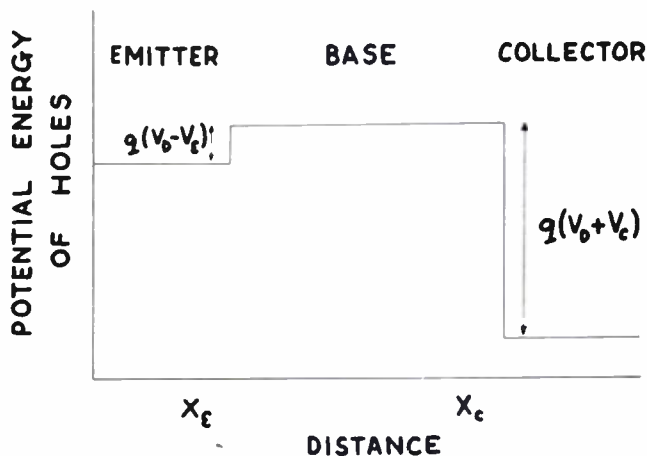
Fig. 3—Potential energy diagram for holes, with bias voltages applied to transistor.

STEADY-STATE SOLUTION OF DIFFUSION EQUATIONS

The substitution of (1) into (2), combined with the assumption of a field-free region, gives the following relation:

$$\frac{dp}{dt} = -\frac{(p - p_B)}{\tau_B} + D_p\frac{d^2p}{dx^2}, \tag{6}$$

which is to be solved subject to the boundary conditions stated by (4) and (5).

We shall first solve the steady-state problem, where $dp/dt = 0$. The time-dependent solution, which will be solved later, will yield information on frequency response.

$$\frac{d^2p}{dx^2} - \frac{(p - p_B)}{D_p\tau_B} = 0. \tag{7}$$

The solution for this is obviously of the form

$$p - p_B = A_1 e^{-x/\sqrt{D_p\tau_B}} + A_2 e^{x/\sqrt{D_p\tau_B}}. \tag{8}$$

Putting in the boundary conditions and noting that $x_c = x_e + w$ where (w) is the width of the base region, (8) becomes

$$p - p_B = (p_2 - p_B)\frac{\sinh\left(\dfrac{x - x_e}{L_B}\right)}{\sinh\left(\dfrac{w}{L_B}\right)}$$

$$- (p_1 - p_B) \frac{\sinh\left(\dfrac{x - x_e - w}{L_B}\right)}{\sinh\left(\dfrac{w}{L_B}\right)}, \quad (9)$$

where we have defined the "diffusion length" for holes in the base region as (L_B) and

$$L_B = (D_p \tau_B)^{1/2}.$$

The emitter current density due to holes $(I_e(p))$ can be obtained by combining (1) and (9) at $(x = x_e)$. This will be the hole current diffusing into the base region from the emitter junction.

$$I_e(p) = - qD_p \frac{dp}{dx}\bigg|_{x=x_e}$$

$$= \frac{qD_p p_B}{L_B}\left[(e^{qV_e/kT} - 1)\coth\left(\frac{w}{L_B}\right)\right.$$

$$\left. - (e^{-qV_c/kT} - 1)\operatorname{csch}\left(\frac{w}{L_B}\right)\right]. \quad (10)$$

The expression reduces to

$$I_e(p) = \frac{qD_p p_B}{w}\left(e^{qV_e/kT} - \frac{1}{2}\left(\frac{w}{L_B}\right)^2\right)$$

for large collector voltages $(V_c > 2 \text{ volts})$ when $[(w/L_B) \ll 1]$.

It will be noted that this emitter-current equation differs slightly from the standard diode-current equation as typified by (16). The term which accounts for the back saturation current is reduced by the factor $[(1/2)/(w/L_B)^2]$. This comes about because of the influence on the emitter by the collector.

The collector current density due to holes diffusing out of the base region into the collector can be obtained in the same manner, except that (dp/dx) is evaluated at at $(x = x_c)$.

$$I_c(p) = - qD_p \frac{dp}{dx}\bigg|_{x=x_c=x_e+W}$$

$$= \frac{qD_p p_B}{L_B}\left[(e^{qV_e/kT} - 1)\operatorname{csch}\left(\frac{w}{L_B}\right)\right.$$

$$\left. - (e^{-qV_c/kT} - 1)\coth\left(\frac{w}{L_B}\right)\right]. \quad (11)$$

Now, of course, at the emitter junction there will be electrons injected from the base into the emitter by the same mechanism as for the hole injection. This electron flow into the emitter will increase the emitter current, but will tend to decrease the efficiency of the device because this electron component of current yields nothing at the collector, and thereby cuts down the ratio of collector to emitter current.

This electron component can be computed as follows: Let (n_e) represent the density of electrons nor-

mally in the emitter. Then by an argument similar to that for (5) we get the electron density in the emitter at the junction as

$$n_1 = n_e e^{qV_e/kT}, \quad (12)$$

and similarly, for electron injection into the base from the collector,

$$n_2 = n_c e^{-qV_c/kT}. \quad (13)$$

The diffusion equation for the electrons is

$$\frac{d^2 n}{dx^2} - \frac{(n - n_e)}{L_e{}^2} = 0. \quad (14)$$

At the emitter, this can be solved subject to (12) with the additional condition that the electron density be (n_e) far in the emitter region. Physical quantities, diffusion length for electrons in the emitter region (L_e), diffusion constant for electrons (D_n), and lifetime of electrons in the emitter (τ_e), are related by the expression

$$L_e = (D_n \tau_e)^{1/2}.$$

The solution of (14) with the prescribed boundary conditions at the emitter is

$$n - n_e = n_e (e^{qV_e/kT-1})(e^{(x-x_e)/L_e}), \quad (15)$$

from which the current of electrons into the emitter can be obtained from

$$I_e(n) = qD_n \frac{dn}{dx}\bigg|_{x=x_e}$$

$$= \frac{qD_n n_e}{L_e}(e^{qV_e/kT} - 1). \quad (16)$$

Calculation of Transistor Parameters

We shall now utilize the results so far obtained to evaluate the transistor parameters concerned with (α), the current-gain factor.

A. Current Gain (α)

From the definition, the current-gain factor (α) is given by

$$\alpha = - \frac{\partial I_c}{\partial I_e}\bigg|_{V_c}. \quad (17)$$

The total emitter current involves both hole and electron flow. This is given by the sum of (10) and (16),

$$I_e = I_e(p) + I_e(n)$$

$$= \frac{qD_n n_e}{L_e}(e^{qV_e/kT} - 1)$$

$$- \frac{qD_p p_B}{L_B}\left[(e^{-qV_c/kT} - 1)\operatorname{csch}\left(\frac{w}{L_B}\right)\right.$$

$$\left. - (e^{qV_e/kT} - 1)\coth\left(\frac{w}{L_B}\right)\right]. \quad (18)$$

The total collector current is obtained from (11) and an electron component similar to (16), which yields

$$I_c = I_c(p) + I_c(n)$$
$$= \left[\frac{qD_n n_c}{L_e} - \frac{qD_p p_B}{L_B} \coth\left(\frac{w}{L_B}\right) \right] (e^{-qV_c/kT} - 1)$$
$$+ \frac{qD_p p_B}{L_B} \operatorname{csch}\left(\frac{w}{L_B}\right)(e^{qV_e/kT} - 1). \qquad (19)$$

If the term $(e^{qV_e/kT} - 1)$ is eliminated between (18) and (19) and the differentiation is carried out as indicated in (17), we obtain

$$\alpha = \cfrac{1}{\cosh\left(\dfrac{w}{L_B}\right) + \left(\dfrac{D_n}{D_p}\right)\left(\dfrac{n_c}{p_B}\right)\left(\dfrac{L_B}{L_e}\right)\sinh\left(\dfrac{w}{L_B}\right)} . \quad (20)$$

For the case of $[(w/L_B) \ll 1]$, this reduces to

$$\alpha = \cfrac{1}{\cosh\left(\dfrac{w}{L_B}\right) + \left(\dfrac{D_n}{D_p}\right)\left(\dfrac{n_c}{p_B}\right)\left(\dfrac{w}{L_e}\right)} . \quad (21)$$

In diffused-junction transistors, made with indium as acceptor material, we may conservatively estimate the following:

$$n_c/p_B \leq 10^{-6}; \quad (w/L_e) < 100; \quad (w/L_B) < 10^{-1}.$$

The denominator of (21) can be approximated as

$$1 + \frac{1}{2}\left(\frac{w^2}{L_B^2}\right) + \left(\frac{D_n}{D_p}\right)\left(\frac{n_c}{p_B}\right)\left(\frac{w}{L_e}\right),$$

and the last term is seen to be almost negligible, compared with the second term which is itself of second order.

B. Injection and Diminuation Factors

By disregarding second-order terms, the above expression for (α) can be put into the approximate form

$$\alpha = \left[\operatorname{sech}\left(\frac{w}{L_B}\right) \right]$$
$$\times \left[\cfrac{1}{1 + \left(\dfrac{D_n}{D_p}\right)\left(\dfrac{n_c}{p_B}\right)\left(\dfrac{w}{L_e}\right)} \right] = \beta\gamma. \quad (22)$$

This is similar to the form described by Shockley, et al.[2] These factors (β), (γ), can be identified as follows: The first term (β) is a diminuation factor which results from the loss of hole current due to recombination of holes with electrons while passing through the base. The second term (γ) is an injection factor and results from the fact that not all of the current injected at the emitter is in the form of holes. There is some electron current injected into the emitter which is undesirable. The factor (γ) can be defined as the fraction of the total emitter current which is in the form of holes.

C. Frequency Dependence of (α)

If we assume the injected hole density to have a small ac variation in time of the form $\exp(i\omega t)$, then the term (dp/dt) from (6) becomes $i\omega(p - p_B)$. This time variation will be satisfied if the periodic ac signal voltage applied to the emitter is small and superimposed on the dc bias. The diffusion equation simply transforms into

$$D_p \frac{d^2p}{dx^2} - \left(\frac{1}{\tau_B} + i\omega\right)(p - p_B) = 0. \qquad (23)$$

Thus, the solution for the steady-state problem can be utilized with the substitution of $(1/L_B)(1 + i\omega\tau_B)^{1/2}$ in place of $(1/L_B)$. The frequency dependence of (α) can then be written immediately. Since the diminuation factor (β) is the only term requiring significant modification, we have from (22)

$$\alpha = \beta = \operatorname{sech}\left((1 + i\omega\tau_B)^{1/2}\,\frac{w}{L_B}\right). \qquad (24)$$

A plot of this relation is shown in Fig. 4.
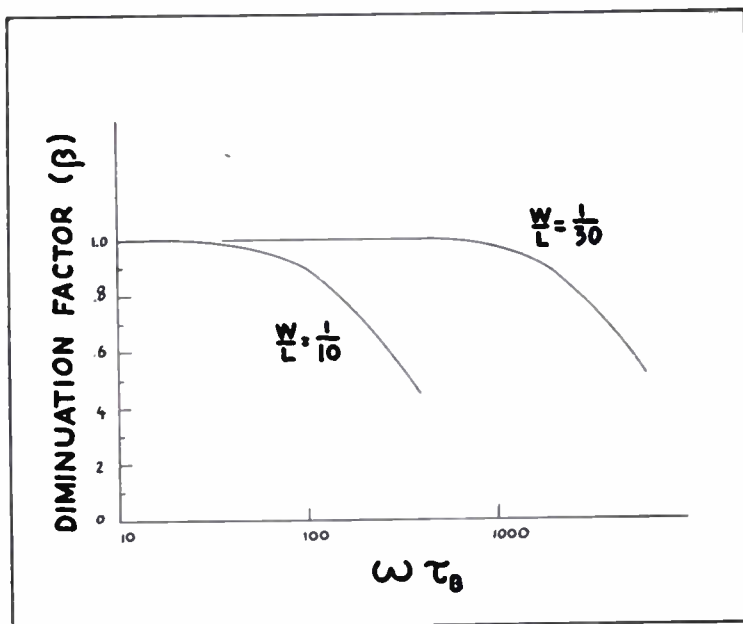
Fig. 4—Semilog plot showing the frequency dependence of diminuation factor (β) for two different values of (w/L_B).

Again, for the case of $[(w/L_B) \ll 1]$ we can expand the expression for (β) and obtain

$$\alpha = \beta = \cfrac{1}{1 + \dfrac{1}{2}(1 + i\omega\tau_B)\dfrac{w^2}{L_B^2}}, \qquad (25)$$

which yields for the frequency cutoff:

$$\omega_0 = \frac{2}{\tau_B}\left(\frac{L_B}{w}\right)^2 = \frac{2D_p}{w^2} \qquad (26)$$

after noting that $(L_B = (D_p\tau_B)^{1/2})$. For high purity, but fairly thick base regions $(w < 10^{-2}$ cm$)$, we obtain $\omega_0/2\pi = 140$ kc/s.

It can be seen from (25) that the factor, $(\alpha/(1-\alpha))$ is much more sensitive to the phase, or frequency term, than (α), above. Thus, to the same approximation as (25), we obtain

$$\frac{\alpha}{1-\alpha} = \frac{2L_B^2}{w^2}\left[\frac{1}{1+\frac{1}{12}\frac{w^2}{L_B^2}+i\omega\tau_B\left(1+\frac{1}{6}\frac{w^2}{L_B^2}\right)}\right]. \quad (27)$$

The cutoff frequency (ω_1) for the factor $(\alpha/1-\alpha)$ can be obtained from (27)

$$\omega_1 = \frac{1}{\tau_B}\left(1-\frac{1}{12}\frac{w^2}{L_B^2}\right).$$

This can be as low as a few kc/s for high lifetime base material. Since this factor enters into the gain for grounded emitter amplifiers, it is seen that the frequency response should be much poorer for this circuit. This turns out to be the case. Calculations along the same lines as indicated here by Pritchard[4] indicate essential agreement with the results.

D. Evaluation of Hole Density (p_B)

We wish now to get an expression for the hole density in the base (p_B) in terms of measurable physical parameters of the bulk germanium.

The conductivity of the base (σ_B) is given in terms of the hole mobility (μ_p), the electron mobility (μ_n), and electron density (n_B), in the form

$$\sigma_B = q(n_B\mu_n + p_B\mu_p). \quad (28)$$

In intrinsic germanium, the electron density (n_i), which is equal to the hole density (p_i) and the intrinsic conductivity (σ_i), is related in the same manner as mentioned above,

$$\sigma_i = qn_i(\mu_n + \mu_p). \quad (29)$$

By the method of detailed balancing, it can be shown that the product of the hole density and that of electron density in a given region are related by

$$n_B p_B = n_i^2. \quad (30)$$

Using the three relationships (28), (29), (30), and simplifying, we get

$$p_B = \frac{\sigma_B}{2q\mu_p}\left[1-\left\{1-4\left(\frac{\sigma_i}{\sigma_B}\right)^2\left(\frac{\mu_p}{\mu_n+\mu_p}\right)^2\left(\frac{\mu_n}{\mu_p}\right)\right\}^{1/2}\right], \quad (31)$$

which further simplifies to the following when $(\sigma_B \gg \sigma_i)$

$$p_B = \frac{1}{q\mu_p}\frac{\sigma_i^2}{\sigma_B}\left(\frac{\mu_p}{\mu_n+\mu_p}\right)^2\left(\frac{\mu_n}{\mu_p}\right). \quad (32)$$

[4] R. L. Pritchard, private communication.

This simply means that nearly all the conductivity in the base is due to electrons. This will be a good approximation for material in which the resistivity is less than 30 ohm-cm. For this resistivity, $p_B = 6\times10^{12}$ holes/cc at room temperature.

CONCLUSIONS

The equations for the p-n-p diffused-junction transistors indicate that the current gain (α) is influenced principally by the diminuation factor (β) rather than the injection factor (γ). This is due to the relatively high-purity base material used in addition to the highly doped emitter and collector regions.

The high-frequency response is influenced by the lifetime of holes in the base region as well as the width of the base region itself. For grounded emitter uses, the lifetime is the predominant influence on the frequency cutoff, while for other uses, the base-region width plays a much more important part.

LIST OF SYMBOLS

p_e, p_B, p_c Hole density (carriers/cc) in emitter, base, collector region under thermal equilibrium.

n_e, n_B, n_c Same as above but for electron densities.

D_p, D_n Diffusion constants for holes and electrons (cm²/sec).

μ_p Mobility of holes (1700 cm/s/volt/cm).

μ_n Mobility of electrons (3600 cm/s/volt/cm.)

q Charge on carriers (1.6×10^{-19} coulomb).

I_e, I_c Emitter and collector current density (amp/cm₂).

τ_B Lifetime of minority carriers (holes) in base region (sec).

τ_e Lifetime of minority carriers (electrons) in emitter region (sec).

L_B, L_e Diffusion length of holes in base and electrons in emitter (cm).

ω Angular frequency (radians/sec).

ω_0 Cutoff frequency of (α) (radians/sec).

ω_1 Cutoff frequency of $(\alpha/1-\alpha)$ (radians/sec).

w Width of base region (cm).

x_e, x_c Distance along axis at which the emitter and collector junctions are located.

V_e, V_c DC bias voltage on emitter and collector (volts).

V_D Diffusion potential.

σ_B Conductivity of base region (ohm-cm)⁻¹.

σ_i Conductivity of intrinsic germanium (ohm-cm)⁻¹.

β Diminuation factor.

γ Injection factor.

α Current gain factor.

n Electron density.

p Hole density.

Effect of Electrode Spacing on the Equivalent Base Resistance of Point-Contact Transistors*

L. B. VALDES†, ASSOCIATE, IRE

Summary—A theoretical expression for the equivalent base resistance r_b of point-contact transistors is derived here. This expression is shown to check experimental values reasonably well if the severity of some assumptions made for purposes of analysis is considered. Electrode spacing, germanium-slice thickness, and resistivity of the semiconductor are shown to be the properties that affect r_b primarily.

STATEMENT OF THE PROBLEM

A DESIGN THEORY for point-contact transistors relates the functions performed by the devices (amplification, switching properties, and so on) to the structure of the devices (resistivity of the germanium, minority carrier lifetime, electrode spacing, and so on). The functions performed by transistors are best described by the electrical characteristics of the device, which are partly expressible in the form of an equivalent circuit for the small signal (linear) properties at fixed bias conditions. Therefore, a relation that expresses the equivalent base resistance of point-contact transistors in terms of the spacing between the emitter and collector electrodes and of the resistivity of the germanium material is part of the design theory. It is the purpose of this paper to derive such a relation and to show partial agreement of experimental results with the theory.

METHOD

The small signal equivalent circuit for point-contact transistors is shown in Fig. 1. This is a T-equivalent for an active four-terminal network,[1] and the base re-
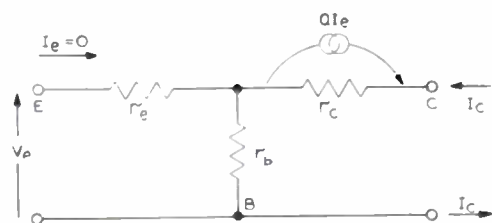
Fig. 1—Transistor equivalent circuit.

sistance r_b is actually the open-circuit feedback resistance R_{12}. By definition,

$$r_b = \frac{\partial v_e}{\partial i_c}\bigg]_{i_e} = \text{constant}, \qquad (1)$$

where v_e is the total voltage on the emitter, i_c is the total collector current, and i_e is the total current through the emitter. At a fixed dc bias condition, r_b can be measured

[1] L. C. Petersen, "Equivalent circuits of linear active four-terminal networks," *Bell Sys. Tech. Jour.*, vol. 27, no. 4, pp. 593–622; October, 1948.

by introducing a known small alternating current of amplitude I_c into the collector and measuring the alternating voltage V_e at the emitter when the emitter is essentially floating for alternating currents, the alternating emitter current I_e being zero. The base resistance is then

$$r_b = \frac{V_e}{I_c}. \qquad (2)$$

This base resistance in a transistor represents the resistance in that part of the semiconductor which lies approximately between the emitter and the bottom surface or base of the semiconductor. The resistance in the material around the collector electrode which extends as far as the emitter is not part of r_b. Also, the spreading resistance at the contact point of the emitter electrode does not appear as r_b, but any resistance between the semiconductor and the large area metallic base contact or in the base lead is part of the base resistance.

Since r_b has been defined in terms of small alternating signals, the direct-current flow does not have to be considered in the analysis. In practice the value of r_b depends on the bias conditions because of phenomena such as modulation of the conductivity of the germanium by the injection of minority carriers by the emitter. However, such effects will be neglected for purposes of analysis and the alternating-current flow lines will be assumed to be as in Fig. 2. The rms values of voltages and currents indicated in the figure are for illustrative purposes only.
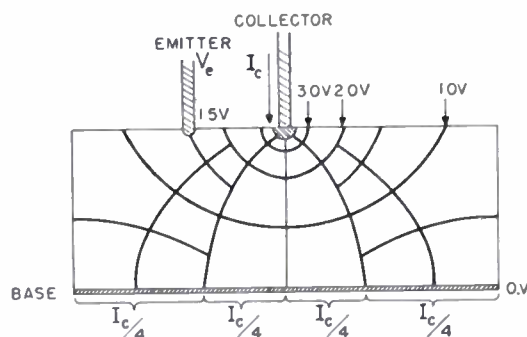
Fig. 2—Alternating-current flow lines in a die.

ASSUMPTIONS

The analytical solution for the equivalent base resistance at low frequencies can be obtained more readily if the following assumptions are made regarding the model:

1. The feedback is produced by the voltage developed by the collector current in flowing through the finite resistivity of the base region.

2. The sides of the germanium die are far from the electrodes and the effect of the sides can be neglected. This reduces the number of boundary conditions, and it will be shown later that the substitution of a slice for the die is a good assumption because the emitter and collector electrodes are generally far from the sides of the die.

3. The top surface and base of the germanium slice are parallel to each other.

4. Cylindrical symmetry for current flow exists about a line normal to the base and passing through the center of the collector region.

5. The collector barrier is hemispherical in shape and always smaller in radius r_0 than the electrode spacing and very much smaller than the slice thickness.

6. The emitter potential is the true floating potential of the point where the electrode contacts the germanium and the diameter of the contact area is very small compared to the electrode spacing.

7. The effect of widening of the space-charge layer of the collector junction is neglected.[2]

8. The resistivity ρ of the germanium is uniform throughout the area of interest.

9. Only drift current is considered in the analysis; diffusion is assumed negligible.

10. Conductivity modulation produced by minority carrier injection can be neglected.

11. The resistance of the base contact between the germanium and the base metal is negligible.

Analysis

To solve for the ohmic (majority carrier) current flow in a model, such as shown in Fig. 2, only two basic equations are needed. These are the continuity and the conductivity equations.

$$\nabla \cdot \bar{J} = 0 \qquad (3)$$

and

$$\bar{J} = \sigma \bar{E}, \qquad (4)$$

where

$\bar{J}$ = current density,
σ = conductivity = $1/\rho$,
E = electric field.

When these two equations are combined and the electric potential is introduced, Laplace's equation is obtained as shown.

$$\nabla \cdot \sigma \bar{E} = \sigma \nabla \cdot \bar{E} = -\sigma \nabla^2 V = 0 \qquad (5)$$

or

$$\nabla^2 V = 0, \qquad (6)$$

where V is the potential at any point in the semiconductor.

The general solution of Laplace's equation will have to fit the boundary conditions outlined in Fig. 3 and

derived from the list of assumptions. For convenience the origin of the system of Cartesian co-ordinates is chosen to coincide with the center of the collector formed area at the surface of the germanium. The z-axis is taken normal to the surface of the germanium. Since the slice is of width w, the $z = -w$ plane corresponds to the conducting base contact of the germanium. The emitter electrode makes contact with the semiconductor at the $z = 0$ plane (top surface), a distance away from the collector.
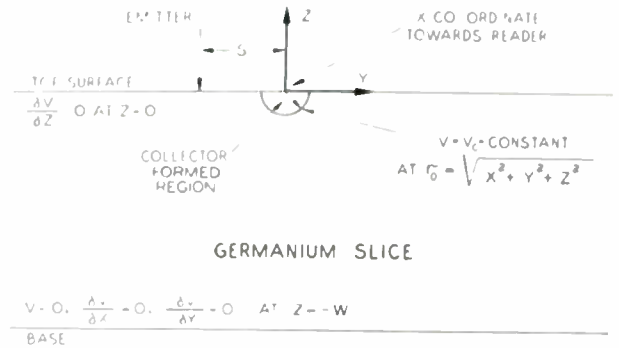
Fig. 3—Geometrical boundary conditions.

The top surface of the semiconductor is nonconducting (or reflecting) and the vertical electric field must be zero. Therefore,

$$\frac{\partial V}{\partial z} = 0 \quad \text{at} \quad z = 0, \qquad x^2 + y^2 > r_0^2. \qquad (7)$$

At the bottom surface the germanium is assumed to be in contact with a much more conductive material so that all the current flow is into the base plane and normal to the base. It follows that

$$\frac{\partial V}{\partial x} = 0 \quad \text{and} \quad \frac{\partial V}{\partial y} = 0 \quad \text{at} \quad z = -w. \qquad (8)$$

The collector barrier is hemispherical and of radius r_0. Within the region $r < r_0$ are included the region of p-type material produced in an n-type semiconductor by forming[3] and the space-charge layer associated with the collector junction. All current flow is assumed radial at $r = r_0$ so that a constant potential V_c' (not the true collector potential) must exist at that position. Therefore,

$$V = V_c' \quad \text{at} \quad r = r_0 = \sqrt{x^2 + y^2 + z^2}, \qquad (9)$$

where the potential V is measured with reference to

$$V = 0 \quad \text{at} \quad z = -w. \qquad (10)$$

Results

The solution of Laplace's equation for the model described is obtained by the method of images to be

$$V = \sum_{n=-\infty}^{n=\infty} (-1)^n \frac{\rho I_c}{2\pi} \frac{1}{\sqrt{x^2 + y^2 + (z + 2nw)^2}} \qquad (11)$$

[2] J. M. Early, "Effects of space charge widening in junction transistors," Proc. I.R.E., vol. 40, pp. 1401–1407; this issue.

[3] L. B. Valdes, "Transistor-forming effects in n-type germanium," Proc. I.R.E., vol. 40, pp. 445–448; April, 1952.

See Appendix I for details of the solution. The floating potential V_e at the emitter electrode is given by (11) as

$$V_e = V \quad \text{at} \quad s = \sqrt{x^2 + y^2}, \quad z = 0. \quad (12)$$

From (2) the equivalent base resistance r_b of a point-contact transistor is

$$r_b = \sum_{n=-\infty}^{n=\infty} (-1)^n \frac{\rho}{2\pi} \frac{1}{\sqrt{s^2 + (2nw)^2}}, \quad (13)$$

where

r_b = equivalent base resistance in ohms,
ρ = resistivity of the germanium in ohm-cm,
s = electrode spacing in cm,
w = thickness of the slice in cm.

This result is in the form of an infinite series which has been summed for values of s/w between 0.01 and 1.0. Equation (13) may then be rewritten in the form of

$$r_b = \frac{\rho}{2\pi s} F\left(\frac{s}{w}\right), \quad (14)$$
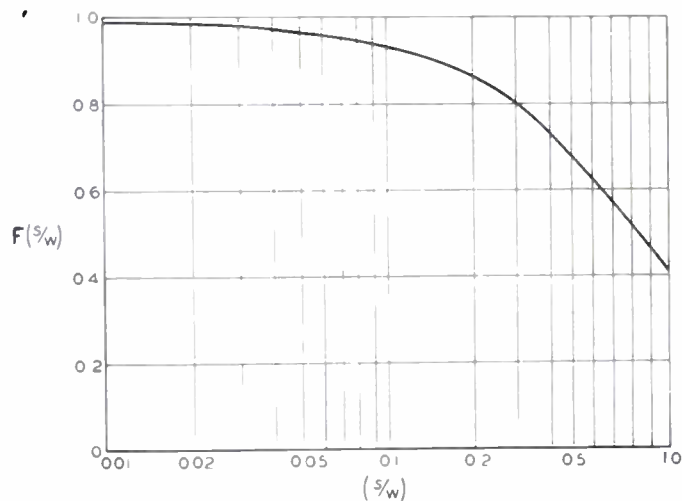
Fig. 4—Function for r_b calculation.

where the function $F(s/w)$ is the result of summing the series and is plotted in Fig. 4. This function is

$$F(s/w) = 1.00 \quad (15)$$

at values of $s/w < 0.01$.

By considering only the first few terms in the series expansion it is possible to derive an expression for r_b, which is within 1 per cent of the true solution at $s/w = 1$ and even more exact at smaller values of s/w. See Appendix II for details. It would therefore be adequate to consider that for all practical cases at $s/w < 1.0$,

$$r_b = \frac{\rho}{2\pi s}\left[1 - 0.693\frac{s}{w} + 0.113\left(\frac{s}{w}\right)^3\right], \quad (16)$$

where the symbols remain as defined for (13).

EXPERIMENTAL RESULTS

An experimental check of (13) has been obtained using a slice of 8.0 ohm-cm material from a single crystal germanium.[4] The procedure followed was that of preparing the slice in the same manner as though it were to be used for device fabrication, placing the two transistor electrodes on the surface of the slice, and making a transistor by forming the collector electrode in the same manner as would be done in a point-contact transistor such as Bell Telephone Laboratories Type M1768. Prior to forming, the equivalent base resistance was measured at bias conditions of 0.3 ma emitter direct current and −30 collector volts with respect to the base. This reading is referred to as preforming $r_b(0.3, -30)$. After forming, the transistor base resistance $r_b(1.5, -5)$ was measured. Here, again, the numbers in parenthesis designate the bias conditions, $I_e = 1.5$ ma and $V_c = -5$ volts. These two bias conditions were selected in order to correlate results with other available data.
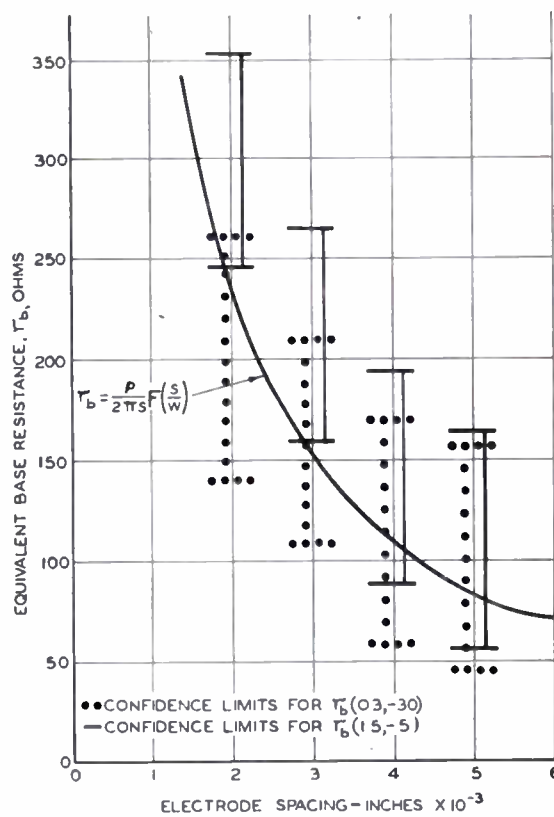
Fig. 5—Experimental results.

In this manner, measurements of preforming $r_b(0.3, -30)$ and postforming $r_b(1.5, -5)$ were obtained four times for each of the electrode spacings of 0.002 inch, 0.003 inch, 0.004 inch, and 0.005 inch. This data was treated statistically, and the 90-per cent confidence limits were calculated for the preforming and postforming base resistance for each of the four electrode spacings. These are the limits within which there is a 90-per cent probability that the true value of r_b will lie. Experimental data are plotted in Fig. 5 together with a curve obtained from (14) and Fig. 4 when $\rho = 8$ ohm-cm, $w = 0.051$ cm (0.020 inch), and s is expressed in inches $\times 10^{-3}$. This calculated curve should, in theory, agree with values of both $r_b(0.3, -30)$ and $r_b(1.5, -5)$.

[4] G. K. Teal, M. Sparks, and E. Buehler, "Single-crystal germanium," PROC. I.R.E., vol. 40, pp. 906–909; August, 1952.

Evaluation of Results

It may be said that agreement between theory and experiment is good if the assumptions that were made to simplify the analysis are considered. After forming, the value of $r_b(1.5, -5)$ is definitely higher than predicted by the formula, especially at close spacings. This may indicate that forming effects, such as traps or thermium (thermal conversion of the germanium because of heating while forming), extend beyond the emitter, that the geometry of the collector barrier is very different from the assumed spherical barrier, that conductivity modulation is affecting the measured values, or that widening of the space-charge region of the collector junction is important.

Lack of symmetry about the z-axis, such as might be the case if a large percentage of the collector alternating current flowed towards the emitter, can effectively increase the value of r_b. This asymmetry could be caused by an irregular collector barrier produced by forming which might be leaning towards the emitter side. It can also be caused by nonuniform conductivity modulation. Most of the minority carriers injected by the emitter flow towards the collector directly and reduce the resistivity of the material between emitter and collector while leaving the resistivity of the rest of the bulk germanium essentially unaffected. This causes an increase in r_b which will be more noticeable at close spacings, when the minority carriers flow more directly between electrodes and do not spread out as much throughout the semiconductor. Uniform modulation of the conductivity of the whole die or slice would tend to make the base resistance lower.

The errors introduced by conductivity modulation should be smaller in the preforming case, $r_b(0.3, -30)$, because the emitter bias at the point of measurement is smaller. Also, prior to forming, it might be expected that the focusing drift field towards the collector would be smaller and injected minority carriers spread out more. The fact that this experiment can approximately explain r_b by bulk conductivity phenomena only is an argument against surface leakage playing a major part, at least in the preforming case.

The expression for r_b presented in (13) was derived for the case of a slice and the experimental result was obtained for a slice. However, it is possible to argue intuitively that the error introduced by applying the same equation to transistor dice is small.

A typical transistor die may have a thickness of 0.020 inch and be 0.060 inch $\times$ 0.060 inch in size and the electrode spacing may be 0.003 inch. The die has side boundaries approximately 0.030 inch from the electrodes; an exact analytical solution would therefore have an infinite series of terms with respect to each of the three co-ordinate axes, whereas the slice solution only has one series of terms in z. This amounts to saying that only the zero-order terms in x and y are used in a slice or that in a die the material extends only to $x = \pm 0.030$ inch and $y = \pm 0.030$ inch and more terms are needed to take into account the other boundaries.

The error introduced by neglecting the effect of the side boundaries of a die, such as would be the case if (14) or (16) are used for checking measurements on dice, can be estimated by assuming that the series in x and y are of approximately the same form as the series in z. When the lower boundary is at $z = -0.030$ inch, the function $F(s/w) = 0.93$ for $s/w = 0.1$. Since $F(s/w) = 1$ for a semi-infinite volume, the error of neglecting the lower boundary is 7 per cent, and it may be estimated that neglecting the side boundaries introduces an error not much greater than 10 per cent for the assumed geometry. Since the analytical formula may not in some cases check experimental results accurately due to the other assumptions made for purposes of analysis, it is justifiable to use it for dice as well as for slices of a uniform-resistivity semiconductor.

Conclusions

The formula for the equivalent base resistance of transistors given in (16) is a useful approximation for many of the cases encountered in practice. Because some of the assumptions that were made to simplify the analysis are not met in practice, the formula will give lower than actual values of base resistance at small electrode spacings. Nevertheless, this formula should be considered a useful design tool in the development of new point-contact transistor types, and also aids in understanding transistor phenomena.

Acknowledgments

Appendix I

Solution of Laplace's Equation for Current Flow in a Semiconductor Between a Point and a Plane

This appendix shows the solution of Laplace's equation with the boundary conditions shown in Fig. 3. Laplace's equation expressed in spherical co-ordinates is

$$\nabla^2 V = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2} = 0. \quad (17)$$

The $\theta = 0$ axis can be made to correspond to the z-axis, and then the assumption that symmetry exists about the z-axis is equivalent to

$$\frac{\partial V}{\partial\phi} = 0 \quad \text{and} \quad \frac{\partial^2 V}{\partial\phi^2} = 0. \quad (18)$$

Rather than solving Laplace's equation for the specified boundary conditions, it is simpler to solve the

problem using the method of images. Fig. 6 shows the location of a few of the images with respect to the germanium slice and the collector current source $+I$ at $z=0$. An infinite number of images is required because there are two plane boundaries parallel to each other that must be satisfied. This can be illustrated best after solving for the case of a single-current source.
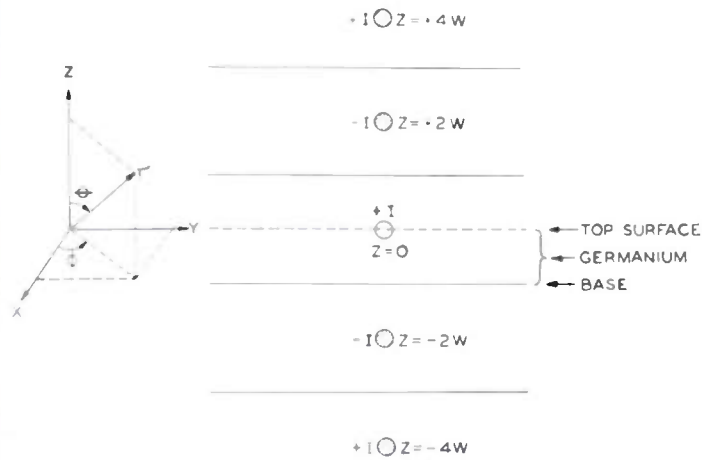
Fig. 6—Location of some images used in analysis.

With a single isolated current source, perfect spherical symmetry exists and

$$\frac{\partial V}{\partial \theta} = 0 \tag{19}$$

so that Laplace's equation reduces to

$$\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right) = 0. \tag{20}$$

To solve this equation it is assumed that I is the current entering half of the infinite volume of material and that the potential V is zero at infinity.

The result of integrating with respect to r is that

$$r^2 \frac{\partial V}{\partial r} = C_1 = \text{constant.} \tag{21}$$

This constant can be evaluated from a knowledge of the electric field E' at $r=r_0$. It is known that

$$J\rho = E = -\frac{\partial V}{\partial r}, \tag{22}$$

where J is the current density. With a given current I coming out of the current source the current density J' at $r=r_0$ can be evaluated as

$$J' = \frac{I}{\text{Area}} = \frac{I}{2\pi r_0^2}, \tag{23}$$

where the surface area of a hemisphere instead of a whole sphere has been used because it is desired to keep I as the current flowing into a semi-infinite volume of material. This is done so that the current I may later equal the collector current I_c in a slice. Therefore,

$$C_1 = r_0^2 \frac{\partial V}{\partial r} = -r_0^2 J' \rho = -\frac{\rho I}{2\pi} \tag{24}$$

and

$$r^2 \frac{\partial V}{\partial r} = -\frac{\rho I}{2\pi}. \tag{25}$$

Equation (25) can be integrated to obtain

$$V = \frac{\rho I}{2\pi r} + C_2, \tag{26}$$

where C_2 is a constant which must equal zero in order to have

$$V = 0 \quad \text{as} \quad r \rightarrow \infty.$$

This expression for potential can be stated in Cartesian co-ordinates as follows:

$$V = \frac{\rho I}{2\pi r} = \frac{\rho I}{2\pi \sqrt{x^2 + y^2 + z^2}}. \tag{27}$$

The field along each of the co-ordinate axes is given by differentiation as

$$-\frac{\partial V}{\partial x} = \frac{\rho I}{2\pi} x(x^2 + y^2 + z^2)^{-3/2} \tag{28}$$

$$-\frac{\partial V}{\partial y} = \frac{\rho I}{2\pi} y(x^2 + y^2 + z^2)^{-3/2} \tag{29}$$

$$-\frac{\partial V}{\partial z} = \frac{\rho I}{2\pi} z(x^2 + y^2 + z^2)^{-3/2}. \tag{30}$$

The co-ordinate system used in the above equations has the origin at the center of the current source. When finding the field due to an image, the co-ordinate system must be translated in position.

The boundary conditions of interest are

$$\frac{\partial V}{\partial z} = 0 \quad \text{at} \quad z = 0, \qquad x^2 + y^2 > r_0^2 \tag{7}$$

and

$$\frac{\partial V}{\partial x} = 0 \quad \text{and} \quad \frac{\partial V}{\partial y} = 0 \quad \text{at} \quad z = -w. \tag{8}$$

In order to have the vertical field zero, as required by (7), any images used in the solution must be in pairs of equal sign and equal distances away from the $z=0$ plane. This is obvious from an examination of (30). The boundary condition at the base (8), requires an equal number of images on both sides of the $z = -w$ plane. However, it is observed from (28) and (29) that for the horizontal field to be zero at any arbitrary value of x and y the images placed must be at equal distances away from the base and of equal but opposite sign.

With the principles enumerated above and the aid of Fig. 6, it is possible to show that an infinite number of images is needed to solve the problem adequately. The original current source $+I$ exists at $z=0$. To satisfy the boundary conditions at the base, $z = -w$, an image of magnitude $-I$ must be placed at $z = -2w$. This image produces a vertical component of field at $z=0$ that must be compensated by another $-I$ image at

$z = +2w$. The last image destroys the symmetry in number about the $z = -w$ plane and requires a $+I$ image at $z = -4w$. Such process continues and the reason for an infinite number is that it is impossible to have an equal number of current sources on both sides of two parallel planes and equal distances away from the planes when the number is finite.

The boundary conditions at $r = r_0$ expressed by (9) are met by the original current source which has the desired shape for a hemispherical collector barrier. The field of the images is balanced and zero at $z = 0$ so that if $r_0 \ll w$ the field due to the images at $z = -r_0$ can be neglected and (9) is satisfied.

Equation (27) expresses the potential due to each current source when the origin of the co-ordinate system is shifted. For an array of current sources along the z-axis of magnitude I and alternating sign, spaced at intervals $2w$, the potential in the semiconductor can be expressed by means of the principle of superposition as

$$V = \sum_{n=-\infty}^{n=\infty} (-1)^n \frac{\rho I}{2\pi} \frac{1}{\sqrt{x^2 + y^2 + (2nw + z)^2}}. \quad (31)$$

This equation appears in the text with I_c substituted for I as (11). The trick of dividing by the area of a hemisphere instead of a sphere in (23) makes I and I_c equal.

Appendix II

Derivation of an Approximate Expression for Base Resistance

The exact expression for r_b given in (13) can be written as

$$r_b = \frac{\rho}{2\pi} \left[\frac{1}{s} + 2 \sum_{n=1}^{n=\infty} (-1)^n \frac{1}{\sqrt{s^2 + (2nw)^2}} \right]. \quad (32)$$

For simplicity, call the summation term $S(s/w)$ and define

$$S\left(\frac{s}{w}\right) = 2 \sum_{n=1}^{n=\infty} (-1)^n \frac{1}{\sqrt{s^2 + (2nw)^2}}. \quad (33)$$

This term can then be put in the form of

$$S\left(\frac{s}{w}\right) = \sum_{n=1}^{n=\infty} (-1)^n \frac{1}{nw} \frac{1}{\sqrt{1 + \left(\frac{s}{2nw}\right)^2}}, \quad (34)$$

which is useful because $s/2nw \ll 1$, particularly as n increases.

The expression under the square-root radical can be expanded as

$$\left[1 + \left(\frac{s}{2nw}\right)^2\right]^{-1/2} = 1 - \frac{1}{2}\left(\frac{s}{2nw}\right)^2$$
$$+ \frac{1 \cdot 3}{2 \cdot 4}\left(\frac{s}{2nw}\right)^4 - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\left(\frac{s}{2nw}\right)^6 + \cdots \quad (35)$$

For an approximate solution it is enough to consider only the first two terms of this expansion. When (35) is substituted into (34), the following is obtained:

$$S\left(\frac{s}{w}\right) \approx \frac{1}{w} \sum_{n=1}^{n=\infty} (-1)^n \frac{1}{n}\left[1 - \frac{1}{2}\left(\frac{s}{2nw}\right)^2\right] \quad (36)$$

or

$$S\left(\frac{s}{w}\right) \approx \frac{1}{w} \sum_{n=1}^{n=\infty} \frac{(-1)^n}{n}$$
$$- \frac{1}{w} \sum_{n=1}^{n=\infty} (-1)^n \frac{1}{2n}\left(\frac{s}{2nw}\right)^2. \quad (37)$$

The series expansion for the natural logarithm of 2 is

$$\ln 2 = - \sum_{n=1}^{n=\infty} \frac{(-1)^n}{n} = 0.693 \quad (38)$$

so that the first term in (37) can be expressed as a logarithm by means of (38). The second term can also be evaluated in a somewhat more complicated manner. To a first approximation.

$$\frac{1}{w} \sum_{n=1}^{n=\infty} (-1)^n \frac{1}{2n}\left(\frac{s}{2nw}\right)^2 = \frac{s^2}{w^3} \sum_{n=1}^{n=\infty} (-1)^n \frac{1}{8n^3}$$
$$\approx \frac{s^2}{w^3}\left(-\frac{1}{8} + \frac{1}{8(2)^3}\right) \approx \frac{s^2}{w^3} \frac{(-7)}{64}. \quad (39)$$

This expansion has been evaluated more accurately,[5] and the correct value is

$$\sum_{n=1}^{n=\infty} (-1)^n \frac{1}{8n^3} = 0.113. \quad (40)$$

Equations (38), (39), and (40), when substituted into (37), result in

$$S\left(\frac{s}{w}\right) = -0.693 \frac{1}{w} + 0.113 \frac{s^2}{w^3}. \quad (41)$$

From (32), (33), and (41) the approximate expression[6] is obtained as

$$r_b \approx \frac{\rho}{2\pi s}\left[1 - 0.693 \frac{s}{w} + 0.113\left(\frac{s}{w}\right)^3\right]. \quad (42)$$

At $s/w = 1$ this expression gives

$$r_b = \frac{\rho}{2\pi} 0.42, \quad (43)$$

whereas the exact expression is

$$r_b = \frac{\rho}{2\pi} 0.418. \quad (44)$$

For smaller values of s/w the result obtained from (42) is even more accurate.

[5] A. Uhlir, personal communication.
[6] This expression was first suggested by R. M. Ryder.

Transistor Operation: Stabilization of Operating Points*

RICHARD F. SHEA†, SENIOR MEMBER, IRE

Summary—This paper treats the problem of supplying emitter and collector bias currents to a transistor amplifier, and derives the mathematical relationship between the values of resistors and voltages required and the desired operating points, in order to achieve a desired degree of stability of these operating points.

I. INTRODUCTION

IN THE utilization of transistors, one of the major problems is that of supplying relatively constant current to the various elements. In the case of the junction transistor, the problem is further complicated by the fact that operation is feasible at very low voltages and currents; hence, it is desirable to make the greatest possible use of this potential efficiency.

Working against this is the effect of variation of the collector current at zero emitter current, both between different units and with temperature in any one unit. This variation with temperature may be as high as ten to one, and with some circuit configurations may produce a variation of this order of magnitude in collector current. With resistance-coupled amplifiers this may result in the operating point shifting in such manner as to materially reduce the gain. In power amplifiers, the shift may produce an increase in internal heating, with the result that the process may be cumulative.

This paper tries to derive the formulas governing the dependence of collector current on this zero-current value and to indicate the degree of stabilization attainable with moderate increase in power dissipation.
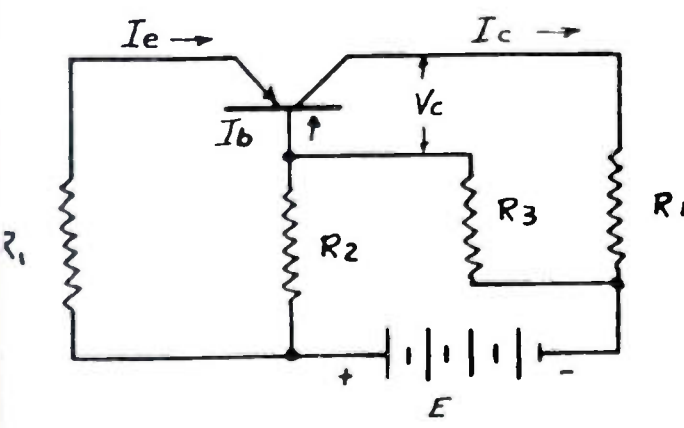
Fig. 1

II. SINGLE-BATTERY OPERATION

The circuit of Fig. 1 shows a transistor with one battery of voltage E and resistors R_1, R_2, and R_3 regulating the currents to emitter, collector, and base. R_L is the dc resistance of the load. The direction of currents are assumed as shown by the arrows. We thus assume;

1. V_c, the collector-base voltage, is maintained within the region where it has negligible effect on collector current.

2. The current amplification, $\alpha = (\partial I_c / \partial I_e) V_c$, is constant over the operating range.

3. V_e, the emitter-base voltage, is zero. In most actual cases it will be less than 0.1 volt.

If we solve the mesh equations for this circuit, we get

$$I_c = \frac{I_{c0}\left(1 + \dfrac{R_1}{R_2} + \dfrac{R_1}{R_3}\right) + \dfrac{\alpha E}{R_3}}{1 - \alpha + \dfrac{R_1}{R_2} + \dfrac{R_1}{R_3}}, \qquad (1)$$

$$I_e = \frac{I_c - I_{c0}}{\alpha}, \qquad (2)$$

$$I_b = \frac{I_{c0} - I_c(1 - \alpha)}{\alpha}, \qquad (3)$$

where I_{c0} is the collector current for $I_j = 0$.

By differentiating I_c with respect to I_{c0} we derive a stability factor S.

$$S = \frac{1 + \dfrac{R_1}{R_2} + \dfrac{R_1}{R_3}}{1 - \alpha + \dfrac{R_1}{R_2} + \dfrac{R_1}{R_3}}. \qquad (4)$$

For stable operation, S should be as low as possible, consistent with reasonable values of E and of power dissipated in the stabilizing resistors.

Expressing I_c in terms of S,

$$I_c = SI_{c0} + \frac{E}{R_3}(S - 1). \qquad (5)$$

In the usual operation there is a desired operating point which dictates V_c, the collector-base voltage, and I_c. For this desired operating point and the available supply E, we can calculate the values of the resistors for any desired value of S, or conversely, of S for any fixed set of resistance values.

$$R_1 = \frac{\alpha(E - V_c - R_L I_c)}{I_c - I_{c0}}, \qquad (6)$$

$$R_2 = \frac{S - 1}{\dfrac{(1 - S + \alpha S)(I_c - I_{c0})}{\alpha(E - V_c - R_L I_c)} - \dfrac{(I_c - SI_{c0})}{E}}, \qquad (7)$$

$$R_3 = \frac{E(S - 1)}{I_c - SI_{c0}}. \qquad (8)$$

The total power supplied by E is

$$R_{dc} = EI_e + \frac{(V_c + R_L I_c)(I_c - SI_{c0})}{S - 1}. \qquad (9)$$

Fig. 2 shows a series of curves illustrating the interdependence of the above values. For these curves an operating point corresponding to $V_e = 5.0$ volts and $I_e = 1.0$ ma was chosen, for a transistor having $\alpha = 0.9$ and $I_{c0} = 0.05$ ma.
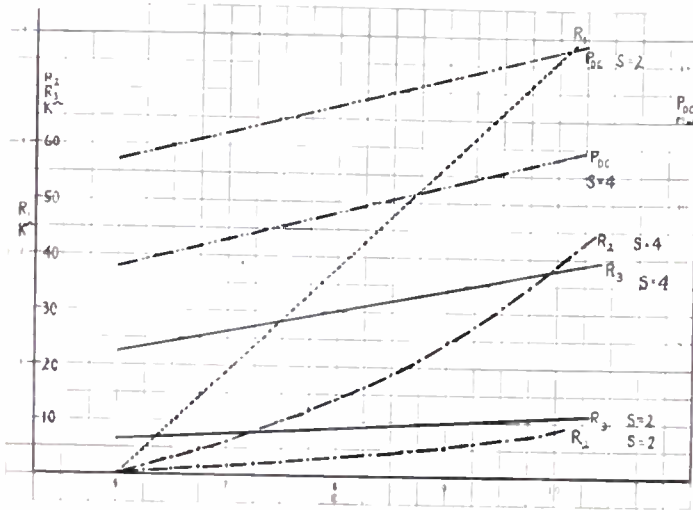
Fig. 2

As an example of the use of these curves, let us take a grounded-emitter configuration with a minimum permissible loading of the base input circuit of 15,000 ohms (This loading is the net parallel resistance of R_2 and R_3.)

Inspection of Fig. 2 shows that for R_2 in parallel with R_3 equal to 15,000 ohms a voltage greatly in excess of 10 volts is required for an S of 2; hence, this degree of stability can only be obtained at the expense of considerable waste power. For $S = 4$, however, a voltage E of 9.15 volts gives $R_2 = 27k$, $R_3 = 34.5k$, $R_1 = 3,000$, and $P_{dc} = 14.6$ mw. This indicates that 5 mw is dissipated in the transistor collector and 1 mw in the load resistance, and 8.6 mw loss in R_1, R_2, and R_3 is required for the above degree of stabilization. In passing, it should be pointed out that without the stabilization resistors the S factor becomes $1/1 - \alpha$ or, in this example, 10; hence, we have improved our stability by a factor of $2\frac{1}{2} \div 1$ at the above cost in power.

III. Separate Emitter Bias Supply

In Fig. 3 is shown an arrangement producing comparatively constant emitter current, but requiring a tapped battery or separate emitter supply.

Equations for I_c, I_e, and I_b and S, R_1, R_2, E_1, E_2, and P_{dc} are as follows:

$$I_e = \frac{1}{R_1 + R_2(1 - \alpha)} [I_{c0}(R_1 + R_2) + \alpha E_1]. \quad (10)$$

$$I_e = \frac{I_c - I_{c0}}{\alpha}. \quad (11)$$

$$I_b = \frac{I_{c0} - I_c(1 - \alpha)}{\alpha}, \quad (12)$$

$$S = \frac{R_1 + R_2}{R_1 + R_2(1 - \alpha)}, \quad (13)$$

$$I_c = S\left(I_{c0} + \frac{\alpha E_1}{SR_1 + R_2}\right), \quad (14)$$

$$R_1 = \frac{E_1[1 - S(1 - \alpha)]}{I_c - SI_{c0}}. \quad (15)$$

$$R_2 = \frac{E_1(S - 1)}{I_c - SI_{c0}}, \quad (16)$$

$$E_2 = V_c + R_L I_c - R_2\left[\frac{I_c(1 - \alpha) - I_{c0}}{\alpha}\right], \quad (17)$$

$$P_{dc} = E_2 I_c + \frac{E_1(I_c - I_{c0})}{\alpha}. \quad (18)$$

For comparison with the example previously given for the single battery circuit, we now have for the same operating conditions

$$R_2 = 15,000 \text{ ohms} \quad \text{and for} \quad S = 4.$$
$$E_1 = 4.0^1,$$
$$K_1 = 3,000 \text{ ohms},$$
$$E_2 = 5.17^1,$$
$$P_{dc} = 9.39 \text{ mw.}$$
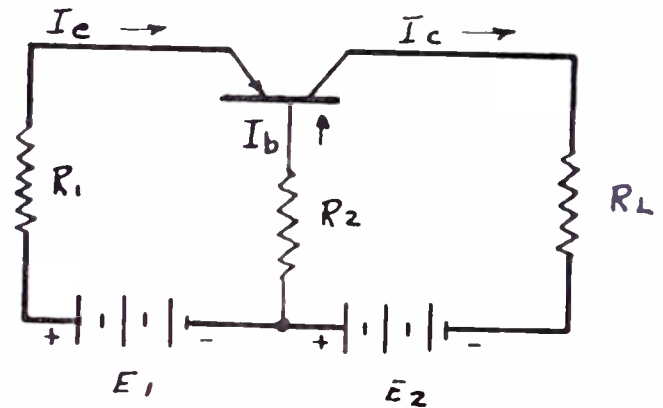
Fig. 3

From this it can be seen that the total battery voltage required is the same as before but that the power requirement is less by about 5 mw. Therefore, we must pay for the privilege of using a single battery by either greater power dissipation or reduced stability.

A caution regarding battery life is in order here. In deciding on the method of biasing the transistor, it must be borne in mind that the battery voltage will decrease with life, to some arbitrary end-of-life value, usually 50 to 75 per cent of the initial voltage. Unless the above computations are made for initial and terminal voltages and satisfactory conditions set up at both extremes, it is possible to reach a condition of inoperation at some time prior to normal end-of-life. This is particularly important where separate emitter and collector batteries having different capacities are used. With different decay rates one can reach a condition where the net collector-to-base voltage drops to zero.

IV. Effect of Emitter Bias

One of the assumptions originally made was that the emitter-base voltage was zero. In most cases this ap-

proximation results in little error; however, as R_1 approaches zero, this will cause an appreciable error. In most cases it has been noted that the dc emitter voltage is approximately proportional to emitter current; consequently, it may be considered that a minimum R_1 exists in the emitter circuit, usually of the order of 100 ohms. With high emitter current, however, the dc emitter resistance may drop considerably below this value, and consequently it may be found desirable to include some value of external resistance in series with the emitter to minimize current drift due to this effect. An indication of adequate emitter resistance is the rapid stabilization of collector current, which will achieve a steady value in a matter of less than a minute; whereas

with insufficient emitter resistance the current may increase gradually for a considerable time, or even rise rapidly to excessive level, particularly at higher operating dissipations.

V. Conclusion

From the above equations and examples it is evident that either single- or dual-battery operation of transistors is feasible, the former at added cost in battery power, and that adequate stabilization against temperature drift by either means is achieved only at the expense of over-all efficiency, unless efficient constant-current sources are developed.

Oscilloscopic Display of Transistor Static Electrical Characteristics*

N. GOLDEN† AND R. NIELSEN†

Summary—An oscilloscope capable of displaying transistor output families is described. Static electrical characteristics are displayed in such a manner as to permit easy evaluation of alpha, "four-pole" collector resistance and feedback resistance.

I. Introduction

APPLICATION of transistors to practical use requires a statement of the electrical nature of the transistor which is commonly made through static electrical characteristic curves.

Point-by-point measurement of static characteristics is time-consuming and tedious. The usefulness of the static characteristics is sufficient in application, and in production of transistors, so that development of a more rapid method of obtaining static characteristics was undertaken at Sylvania Electric Products, Inc.

The equipment described below presents the output family curves on a calibrated oscilloscope in such a manner as to permit rapid determination of significant transistor parameters.

II. Current Biases for the Transistor

Electrical description of the transistor is accomplished through families of characteristic curves in a manner similar to the description of the vacuum tube. Four quantities must be related in describing the transistor in static relationships. These four quantities (emitter voltage and current and collector voltage and current) may be graphed in several manners, but the uses to which the graphs are put determine the desirable plots.

Design of the present equipment is aimed at three transistor parameters: (1) alpha, the negative of the ratio of collector current to emitter current at constant collector voltage; (2) collector resistance (in the four-pole transistor equivalent circuit); and (3) the four-pole

feedback resistance. Graphs of V_e versus I_e and V_e versus I_c yield the desired quantities. Alpha and collector resistance may be found on the plot of V_e versus I_c, and the feedback resistance is the slope of the curve of V_e versus I_c.
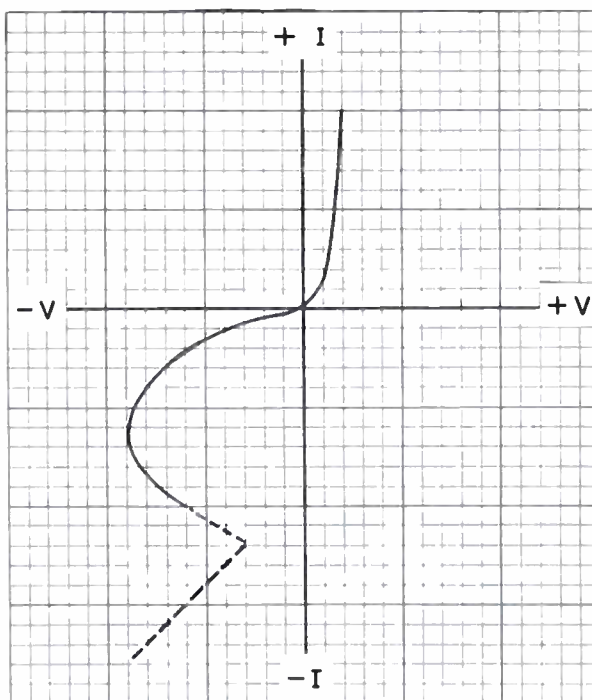
Fig. 1 – Current-voltage plot for rectifying point contact.

Similarity of vacuum-tube and transistor description may be extended through the concept of duality. Table I displays dual relationships of vacuum tube and transistor in small-signal analysis. It is apparent that current in the transistor is the dual of minus the vacuum-tube plate cathode voltage.

Small-signal stability in the vacuum tube is obtained when input or output terminals are short circuited.

Duality continues to hold so that the transistor reveals itself stable for small-signal analysis with open-circuited terminals. This fact indicates the need of biasing sources with high internal impedance, for obtaining static characteristic curves. An ideal current generator possesses infinite internal impedance; thus the equipment should operate with emitter and collector currents as determined by suitable current generators.

A second factor indicating need of currents independently determined at emitter and collector is resultant of a semiconductor phenomenon. When extended to sufficiently great potentials, a plot of current versus voltage for a semiconductor reveals a multivalued curve. Fig. 1 illustrates this characteristic wherein three values of current may exist at voltage *V*. Choosing current as an independent variable in the plot of the *E-I* characteristic eliminates the difficulty of multiple values.

III. Circuit Description

Fig. 2 is a block diagram of the curve tracer functions. No switching is shown for the scope calibrator. Switch S_1 permits display of V_e versus I_c or V_c versus I_c.
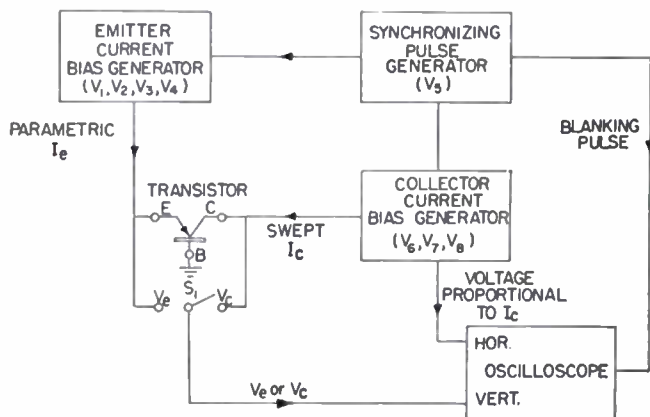
Fig. 2—Blocked diagram of curved tracer.

The curve tracer's fundamental function is sweeping collector current from zero to approximately 20 ma. By changing emitter bias currents during successive collector sweeps, a family of curves is obtained. Emitter bias current is the parametric variable, then, for plots of V_e versus I_c and V_c versus I_c.

Successive collector sweeps and emitter biases are controlled in time by the synchronizer generator. Persistence of vision and trace persistence on the oscilloscope puts a lower limit on the rate at which successive sweeps may occur. The present equipment utilizes a synchronizer generator frequency of approximately 1,000 cps to display a group of 16 curves.

The value of emitter current is controlled by four flip-flop stages. Curves may be displayed between zero and 1.5 ma in 0.1-ma steps, or zero and 3 ma in 0.2-ma steps to the emitter.

Fig. 3 is the curve-tracer schematic circuit diagram. Tube V_5 is a free-running multivibrator operating from ground to negative voltages on its plates. From one plate of V_5 a trigger is coupled off to V_1 the first of four flip-flop stages utilizing V_1 through V_4. Flip-flop plates operate between ground and positive 105 volts.

With this value of plate voltage on an "off" flip-flop, a large resistor (relative) to the emitter terminal of a grounded-base transistor comprises a practical current generator. Successive flip-flop stages feed 0.1, 0.2, 0.4, and 0.8 ma to the emitter or double these values when S_7 is thrown. Addition of these currents through the emitter terminal of the transistor yields successive bias currents of zero to 1.5 ma or zero to 3 ma.

Connected directly to the plate of V_5 one grid of V_6 alternately allows charge and discharge of capacitor C_{18}. The sawtooth waveform thus derived is applied through a cathode follower utilizing the second half of V_6 to the collector current-sweep generator. Current generator, composed of V_7 and V_8 and associated components, supplies sweep current to transistor collector circuit.

A pulse derived from V_5 applied to the oscilloscope beam blanks the scope beam during the change of the flip-flops and the collector sweep-current flyback. Switch S_5 allows emitter current to be calibrated. This is accomplished by setting S_1 through S_4 to appropriate positions and adjusting the variable resistors between flip-flop plates and emitter. Proper setting of S_5 and S_6 allows calibration of vertical- and horizontal-scope amplifiers. Two positions on S_6 allow display of V_e versus I_c or V_c versus I_c.

As designed, the curve-tracer sweep-current generator can supply zero to 15 ma at voltages from zero to 200 volts. These limits are ample for present point-contact transistors. With improvement in transistors to allow greater voltages, modification of the current generator will make available larger currents at higher voltage.

IV. Use and Utility

The rapidity with which transistor static electrical characteristics are displayed is the curve tracer's only asset. It permits rapid comparison of individual transistors with each other or some standard. An oscilloscope camera achieves permanent records of the characteristic curves. The result is a test equipment capable of rapid, accurate measurements of the fundamental transistor electrical characteristics.

TABLE I

VACUUM TUBE	TRANSITOR DUALITY
V_p	$-I_c$
I_p	$-V_c$
V_g	$-I_e$
I_g	$-V_e$

Studies of the forming of point-contact transistors by electrical pulses have been facilitated by curve tracer. Rapid visualization of pulse effect is possible and has led to rapid definition of pulsing technique.

In life-test studies, photographs of transistor curves before and after test permit evaluation of aging characteristics with minimum time wasted on obtaining curves. The time and labor saved in this manner is of appreciable value during extensive life-test studies.

As a production tool, the curve tracer is adapted to many needs other than its use in life tests, and proves to be of use at several steps in production process.
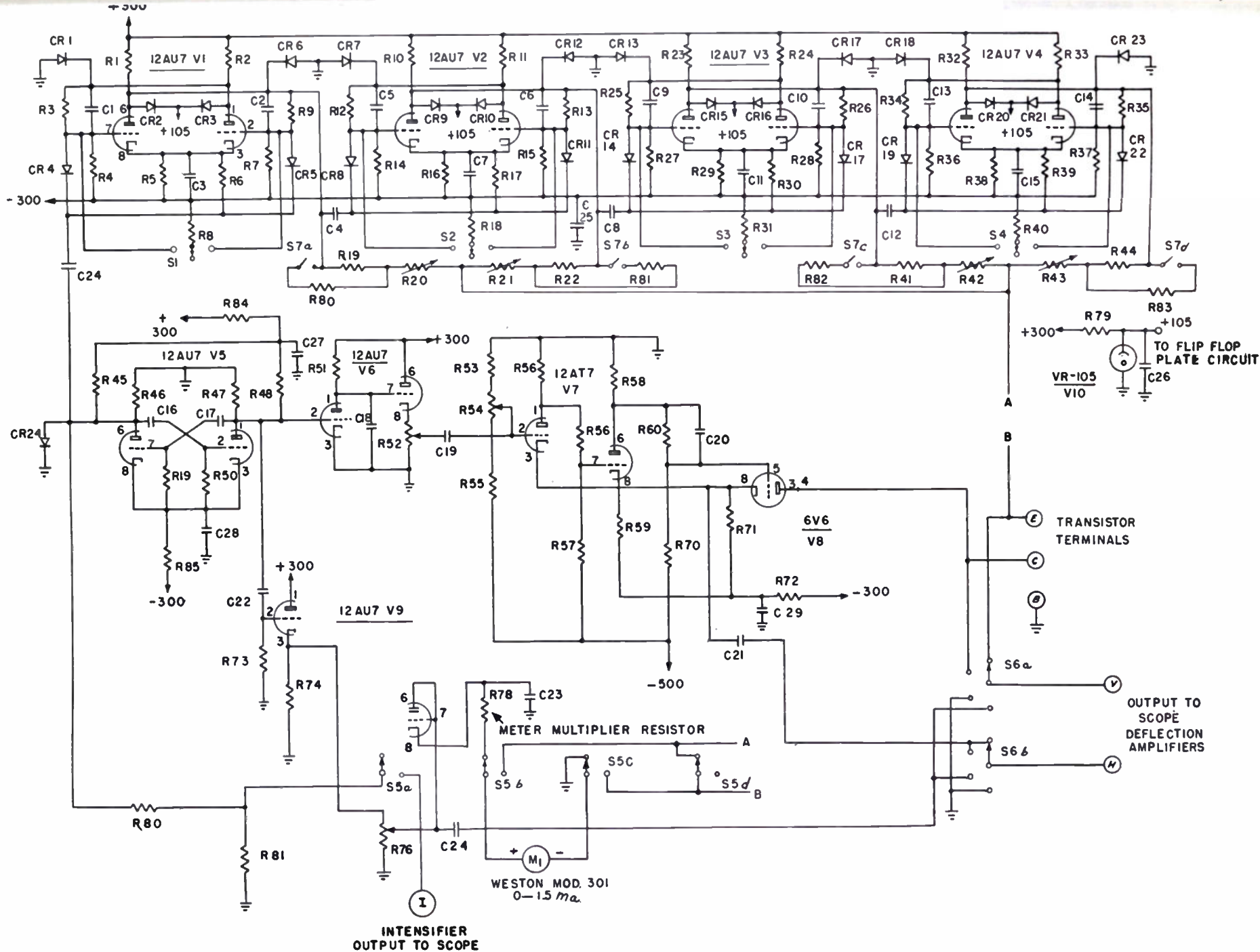
Fig. 3—Schematic diagram of curved tracer.

On the Theory of Noise in *P-N* Junctions and Related Devices*

RICHARD L. PETRITZ†

Summary—A discussion of possible sources of noise in *p-n* junctions is given. The sources are classified into two groups, the first group constituting noise resulting from fluctuations inherent in the electronic system of a *p-n* junction, the second group constituting noise caused by a fluctuating quantity which exerts a control over the average flow of current. Noise of the first group is investigated, and it is shown to be a result of concentration fluctuations of the minority carrier. The model used for a *p-n* junction is characterized by a distribution of localized energy states for the minority carrier in addition to the valence and conduction-band states. The local concentration of the minority carrier, $p(x, t)$, fluctuates by body and surface recombination and by diffusion current flow. Fluctuations in $p(x, t)$ constitute an infinite-dimensional random process. A portion of the paper is devoted to reducing the problem to one in a finite number of variables while retaining the physical features of a *p-n* junction over the frequency range $0 \rightarrow 1/2\pi\tau p$. A noise theory is developed on the basis of the resulting lumped parameter description of a *p-n* junction. The results of the noise theory are summarized in an equivalent circuit and an appropriate noise generator. Noise characteristics of *p-n* junction rectifiers and transistors are analyzed. The available noise power of a *p-n* rectifier is voltage dependent, the equation resembling that of Weisskopf for point-contact rectifiers. The noise figures of *p-n* transistors are found to be of the order of unity (zero db) and independent of size and current density. A comparison of point-contact and *p-n* rectifiers and transistors is made, the Weisskopf noise formula being used for the point-contact devices. A relation between the noise spectrum and the admittance of a *p-n* junction is obtained. The mean lifetime, τ_p, of the minority carrier characterizes the spectrum. The $1/f$ spectrum is not obtained. It is concluded that noise resulting from fluctuations inherent in the electronic systems of *p-n* and point-contact rectifiers and transistors constitute only a part of the measured noise; another source of noise of class II is required to account for, one, the noise figures of *p-n* and point-contact transistors being appreciably greater than unity, two, the large difference between noise figures of *p-n* and point-contact transistors, and three, the $1/f$ portion of the measured noise.

1. Introductory Remarks and a Survey of Previous-Noise Investigations

NYQUIST[1] HAS PROVED that the mean-square voltage fluctuation in a frequency interval f, $f + \Delta f$ appearing at the terminals of a resistive device in thermal equilibrium is $4kTR\Delta f$, where k is Boltzmann's constant, T is the absolute temperature, and R is the resistance. Nyquist's derivation is made by the use of statistical mechanics and thermodynamics and is independent of the atomic model of the resistance. This formula has been experimentally[2] verified over all frequencies that are accessible to electronic measurement technique. The high-frequency form of this equation is still open to question,[3] but does not concern us in this paper.

When steady currents flow through metals, semicon-

ductors, and related devices, Nyquist's law need not hold and experiments[4] indicate appreciable deviations from it in semiconductors, rectifiers, and the like. Metallic resistors generally show no appreciable deviations from the Nyquist law in the region of Ohm's law. An exception to this is thin metallic films, a study of which (by Bernamont[5]) is an important chapter in the history of investigations of noise under nonthermal equilibrium conditions.

Noise in bulk semiconductors, rectifiers, and transistors has, when a steady current flows, the following qualitative characteristics:

1. The noise power per frequency interval Δf increases over the Nyquist value. The increase is generally, although not always, proportional to the square of the average current. A more precise statement of this is that the excess noise power increases as the square of the current, the excess noise being defined by subtracting the Nyquist noise power from the measured noise power.

2. The spectrum is no longer flat, but increases towards lower frequencies. The increase is described by a $1/f^\kappa$ law, where κ is approximately one. We will use the expression, "the $1/f$ law," to categorically describe this spectrum.

3. The excess noise is not strongly temperature dependent.

4. While the characteristics (1-3) are exhibited by point-contact and *p-n* junction rectifiers and transistors, a point-contact device is considerably noisier than the corresponding *p-n* junction device. Point-contact transistors have, for example, a noise figure of about 60 db, while the noise figure of *p-n* junction transistors is 10–20 db.

There have been two general approaches to explaining the noise characteristics of metals, semiconductors, and related devices in nonequilibrium situations. One approach has been to relate the noise to fluctuations inherent in the electronic system of the component. We shall call these class I fluctuations. The second approach has been to relate the noise to a fluctuating quantity which modulates the average flow of current. We shall call these class II fluctuations.

Before going on we will define what we mean by the electronic system of metals and semiconductors. In the free electron theory of solids, metals and semiconductors can be considered to be composed of two thermodynamic systems—the electronic system and the lattice system. The electronic system is basically that system of energy levels, electrons, and holes that we are accustomed to using in studying electronic properties of metals and semiconductors. The state of the electronic system is defined by stating the distribution of electrons

and holes over the electronic energy states. At thermal equilibrium this distribution is given by the Fermi-Dirac function.

The lattice system is that part of a solid concerned primarily with such things as structure, lattice vibrations, and the like. It plays a relatively minor role in certain electronic properties, such as the electronic specific heat; but in others, such as the electrical resistance, it is of considerable importance.

Going back to noise, one general approach to studying noise in metals, semiconductors, and related devices has been to examine fluctuations inherent in the electronic system itself. We use the word inherent because fluctuations must occur in all thermodynamic systems. In the case of the electronic system of metals and semiconductors these fluctuations are principally velocity fluctuations of the charge carriers and fluctuations in the number of charge carriers. These can be lumped under the general description of fluctuations in the distribution of electrons and holes over the electronic energy states. It is convenient to separate them because in a scattering process a particle's velocity is changed (that is, it moves to a different momentum state); but the number of electrons and holes in the conducting state is not changed. When an electron and hole recombine, the number of charge carriers is changed. The calculation of the transition probabilities for scattering and recombination processes is a quantum mechanical problem, and we shall not be concerned with this here. Since the electrical resistance is closely related to these same transition probabilities, it is perhaps not surprising that noise resulting from fluctuations in the electronic system can involve the electrical resistance. In fact, when thermal equilibrium exists, this noise must reduce to the Nyquist formula. In the case of metals and bulk semiconductors the Nyquist noise is related to velocity fluctuations of the charge carriers. This is not always the case though; we shall show in this paper that, for example, the Nyquist law is related to concentration fluctuations of the minority carrier in p-n junctions. We will discuss this point in more detail later in the paper.

Weisskopf[6] has suggested using the term "shot noise" for noise associated with fluctuations of the type discussed above. Because this term is so well identified with a particular type of electronic noise, we shall use instead the term "electronic noise" to designate noise associated with fluctuations inherent in the electronic system of metals, semiconductors, and related devices.

A natural starting point for theoretical studies of noise in metals, semiconductors, and related devices is to fully investigate "electronic noise." This has been done by a number of people, and it seems worth while to briefly survey these efforts, particularly since it helps one to understand why the second broad class of theories have been developed.

A kinetic theory of the Nyquist relation has been given by Weisskopf[6] and Bakker and Heller[7] for bulk metals and semiconductors. They show that it is directly related to velocity fluctuations of the charge carriers.

The noise caused by fluctuations in the number of charge carriers has been investigated by Bernamont[5] and others,[8] principally in an effort to explain the $1/f$ law. Bernamont first proposed this as the source of the current dependent noise in thin metallic films. The first application of this idea to semiconductors was apparently made by Davydov and Gurevich.[8]

It is shown in the papers of footnote references 5 and 8 that the power spectrum obtained from a recombination process with a single mean lifetime is of the form $[1/1+(\omega\tau)^2]$. This does not fit a $1/f$ law over any appreciable frequency range. The idea of superimposing spectra of the $[1/1+(\omega\tau)^2]$ type using a distribution of lifetimes seems to have occurred independently to a number of investigators.[5,8] The first to have done this is again Bernamont.[5] A $1/f$ law can be approximated over a large frequency interval if a sufficiently large distribution of lifetimes is used. A serious difficulty with this procedure is that the $1/f$ law extends to very low frequencies; in fact, the measurements of Montgomery[9] indicate no deviation from it at $1/10$ cps. This makes it necessary to introduce very long lifetimes in the distribution, and arguments for such long electronic lifetimes have not been given.

Schottky's[10] theory of the "shot effect" in vacuum tubes and Weisskopf's[6] theory of a "shot-effect" noise in point-contact rectifiers are important studies of "electronic noise."

The difficulty of explaining the $1/f$ law at low frequencies on the basis of electronic fluctuations is one of the reasons that a second broad attack on the problem of nonthermal equilibrium noise has developed. This second approach has been to relate the noise to a fluctuating quantity which exerts a control over the average flow of current. A convenient way of describing this effect is in terms of the resistance, that is, one describes it as a fluctuation in the resistance. We will call this second class of noise "modulation noise."

The first of these investigations was made by Schottky[11] in his theory of flicker noise. He proposed that fluctuations in the ionic concentration near the surface of the cathode would cause fluctuations in the work function, the emission current being changed accordingly. Using a simple jump mechanism so that an ion jumps into and out of the surface region leads to a "jump spectrum," that is, the $[1/1+(\omega\tau)^2]$ spectrum. This basic idea has been extended by a number of investigators; a partial list is given in the references.[12]

A superposition of ionic lifetimes has also been proposed[12] for explanation of the $1/f$ law. Van der Ziel[13] gives an excellent discussion of the physical meaning of the procedure and also shows the relation of it to Macfarlane's[12] earlier theory.

Studies of the noise spectrum resulting from ions diffusing in and out of the surface region rather than jumping in and out have been made by Miller, Richardson, Macfarlane,[12] and Petritz. The spectrum in this case depends on the number of dimensions in which diffusion takes place. Diffusion in one dimension approximates a $1/f$ law over about three decades and goes as $1/f^{1/2}$ at

low frequencies and as $1/f^{3/2}$ at high frequencies. Van der Ziel,[13] Richardson,[14] and others[8,12] have pointed out that a $1/f$ law cannot hold over the entire frequency range because $\int_{f_2}^{f_1} df/f = \log f_1/f_2$ diverges at both the low- and high-frequency limits. Therefore, it can be expected that at sufficiently low frequencies the law must be $1/f^\kappa$, where $\kappa < 1$, and at high frequencies $1/f^l$, $l > 1$. A one-dimensional ionic diffusion mechanism has therefore promise of explaining the so-called $1/f$ spectrum.

Perhaps its most serious difficulty is connected with the temperature dependence of the $1/f$ noise. Experimental evidence[4] indicates that the $1/f$ noise is rather temperature insensitive. Ionic diffusion is quite temperature sensitive, the diffusion coefficient[15] being of the form $D = D_0 \exp(-E/kT)$, where E is the activation energy of the diffusion process. Very low values of E would be required to explain the observed temperature dependence of the noise. This is particularly objectional in a p-n junction because in the interior of solids the diffusion activation energies of atoms are known[16] to be of the order of electron-volts. The activation energies for surface diffusion can be much lower.

There are a number of other quantities whose fluctuations can affect the average flow of current in metals and semiconductors; Van der Ziel,[13] Richardson,[14] and others[12] have discussed some of them.

Richardson[14] has made a general mathematical study of fluctuations in the electrical resistance resulting from coupling the resistance to a diffusional medium, the diffusing medium being particles or heat. He has analyzed several special physical models and obtains a $1/f$ spectrum from a contact between relatively large areas of rough surfaces covered with diffusing layers. While such a model may describe quite well the physical situation in a granular resistance, it seems to us that single-crystal p-n junctions grown from the melt do not have this characteristic.

The lattice system of metals and semiconductors is a thermodynamic system, and its temperature is subject to fluctuations in accordance with the general laws of statistical mechanics. The resistance of metals and semiconductors is dependent on the lattice temperature through the mobility and the effective number of charge carriers. The effect of lattice temperature fluctuations on the resistance $R(T)$ of homogeneous metals has been investigated[16] in connection with bolometers as radiation detectors. Their importance decreases with size of the resistor as can be seen by elementary calculation

$$R(T) = \rho(T)L/A, \qquad (1)$$

where $\rho(T)$ is the resistivity, A the cross-sectional area, and L the length of the sample.

$$\Delta R = \left[\frac{\partial \rho}{\partial T}\bigg|_{(T)} \right] \frac{L}{A} \Delta T$$

$$\frac{\langle \Delta R^2 \rangle}{R^2} = \frac{\left[\frac{\partial \rho}{\partial T}\big|_{(T)} \right]^2}{\rho^2} \langle \Delta T^2 \rangle, \qquad (2)$$

where $\langle \Delta R^2 \rangle$ is the mean-square fluctuation in the resistance resulting from the mean-square lattice temperature fluctuation $\langle \Delta T^2 \rangle$. Now from statistical mechanics[17] $\langle \Delta T^2 \rangle = k\langle T \rangle^2/C_v \, AL$, where C_v is the specific heat of the body and AL the volume. Therefore, we have

$$\frac{\langle \Delta R^2 \rangle}{R^2} = \frac{\left[\frac{\partial \rho}{\partial T}\big|_{(T)} \right]}{\rho^2} \frac{k\langle T \rangle^2}{C_v \, AL}, \qquad (3)$$

and the size dependence is clearly indicated. For normal-size resistors this effect is of negligible importance as compared with electronic noise.

Recently we[18,19] pointed out that the current in rectifiers and transistors is controlled by conditions in the immediate neighborhood of the contact difference of potential. Another way of stating this is to say that the resistance is concentrated in the immediate neighborhood of the contact. Because of this, the effective L that appears in (3) can be quite small and the corresponding mean-square fluctuation in the resistance quite large.

This brief survey of the large amount of experimental and theoretical work that has been done in connection with noise in metals, semiconductors, and related devices should not be considered to be complete. It is presented primarily for the purpose of giving the reader some insight as to why we have chosen to investigate certain particular mechanisms as causing noise in p-n junctions.

In choosing one or more mechanisms as a basis for studying noise in a particular device, such as a p-n junction, it seems to us that in addition to describing the noise of the device itself, it is desirable that the mechanism should also be able to account for noise in other devices that exhibit similar noise characteristics. For example, granular resistances, point-contact devices, and p-n junction devices all exhibit similar noise characteristics, and it is reasonable to seek one or more common causes for this noise. It should be stressed that because of the possible large differences in physical structure of these devices on the atomic and crystalline level it is not at all necessary that the same mechanisms be responsible for the noise in each. Our point is that the similarity in the noise characteristics leads one to seek common causes.

A natural starting point for studying noise in p-n junctions is to investigate noise inherent in the electronic system of the device. This will be the program of this paper. We are preparing another paper in which modulation noise is investigated, and in particular, a comprehensive study of noise resulting from lattice temperature fluctuations will be given.[18,19]

As mentioned earlier, the two basic fluctuations that can occur in the electronic system of a semiconductor are velocity fluctuations of the charge carriers and fluctuations in the number of charge carriers. We will show that electronic noise in p-n junctions arises principally from concentration (number) fluctuations of the minority carrier.

In view of the large number of previous studies of noise in semiconductors resulting from fluctuations in the number of charge carriers, it seems desirable to state where our work differs from that previously done.

In bulk semiconductors the current flow is proportional to the electric field strength, and current flow by diffusion is negligible. It is then not necessary to consider the local concentration, $p(x, t)$, of the carrier but only the total number of carriers in the sample. Effects due to fluctuations in the local concentration are negligible.

Now in a *p-n* junction, current flow is principally a result of concentration gradients, and it is therefore necessary to consider the local concentration $p(x, t)$ of the minority carrier. The number of random variables, $p(x_1, t)$, $p(x_2, t)$, $p(x_i, t)$, and the like is then infinite, resulting in a rather complicated mathematical problem. A portion of this paper is devoted to reducing the problem to one in a finite number of variables, while retaining the physical features of a *p-n* junction over the frequency interval $0 \rightarrow 1/2\pi\tau p$.

Basically then our work differs from that listed in footnote reference 8 in that current flow is by diffusion instead of being proportional to the electric field strength. While the physical problem is not more complicated, the mathematical problem seems to be more so.

We have attempted to emphasize the physical aspects of the problem in the text and have placed the more mathematical aspects in an appendix. The results of the study are summarized in an equivalent circuit (see Fig. 5) with appropriate noise generators. A brief discussion of how the results can be used to calculate electronic noise characteristics of *p-n* junction devices is given in Section VII.

II. Model of *p-n* Junction

The voltage-current relationships for a *p-n* junction have been derived by Shockley[20] in his classic paper. We shall be concerned with deriving the electronic noise characteristics at such a junction. We consider hole injection into an *n* type region and will use Shockley's notation in so far as possible. The basic equations describing the behavior of holes in an *n* type region are

$$\frac{\partial p(x, t)}{\partial t} = -\frac{(p - p_n)}{\tau_p} - \frac{1}{q} \operatorname{div} \overrightarrow{I_p} \quad (4)$$

$$\overrightarrow{I_p} = A\left[q\mu_p p\overrightarrow{E} - qD_p \operatorname{grad} p\right] \quad (5)$$

$$\cong -AqD_p \operatorname{grad} p, \quad (6)$$

where $p(x, t)$ is the hole concentration at x at time t, I_p is the hole current, μ_p is the hole mobility, p_n is the hole concentration at thermal equilibrium, τ_p is the mean lifetime of a hole, and D_p is the diffusion constant.

The following assumptions[29] are normally made in the analysis of *p-n* junctions:

a. Donors and acceptors are fully ionized (a good assumption for germanium at room temperature).

b. Density of minority carriers is much smaller than density of majority carriers in each region.

c. The net rate of recombination in any region is linear in the deviation of the minority carrier density from its thermal equilibrium value. (Assumptions (b) and (c) permit the use of linear equations in dealing with the currents arising from carrier injection.)

d. Space charge is negligible except at the space-charge regions in the *p-n* junctions themselves.

e. Current flow is primarily a result of concentration gradients (6).

In formulating a noise theory for a device such as a *p-n* junction, it is important to recognize that there is a fundamental relation between the macroscopic equations of the system ((4), (5), (6) are the macroscopic equations of a *p-n* junction) and the equations describing fluctuation phenomena of the system. Such a relation is that, on the average, the decay of a fluctuation follows the ordinary phenomenological macroscopic laws. This relation is a fundamental assumption in Onsager's Principle of Microscopic Reversibility and the papers of reference[21] give a careful discussion of it. When the macroscopic laws are nonlinear, this relation exists only when small deviations are considered in the macroscopic equations. Now (4) has already been linearized by considering small deviations from the normal concentration. Therefore, we must formulate our noise theory so that (4) appears as an equation describing the average decay of a concentration fluctuation.

We have mentioned this point because it has an important bearing on the manner in which the distribution of localized impurity states is incorporated into our noise theory. This is closely related to the meaning of τ_p in (4), and we will discuss this now.

If we consider a small volume element $\Delta V = A\Delta x$, the number of holes in this region can change in two ways, first by body and surface recombination and second by current flow between neighboring regions. We analyze briefly the recombination processes in order to properly incorporate them into the model.

Recombination can occur in the body of the material and at the surface. Considering body recombination first, there may be several processes by which a hole can be removed from the valence band. Considering, for example three of these, we have

$\alpha_c = 1/\tau_c$ transition probability per unit time for a hole to recombine with an electron,

$\alpha_1 = 1/\tau_1$ transition probability per unit time for a hole to fall to the level E_1,   (7)

$\alpha_2 = 1/\tau_2$ transition probability per unit time for a hole to fall to the level E_2.

The levels E_1 and E_2 represent impurity states and are shown in Fig. 1. Now if these events are competing processes, we have

$$\alpha = \alpha_c + \alpha_1 + \alpha_2 = \frac{1}{\tau_c} + \frac{1}{\tau_1} + \frac{1}{\tau_2} = \frac{1}{\tau_b} \quad (8)$$

as the transition probability per unit time that a hole is removed from the valence band by body recombination,

without stating into which state it falls. The mean lifetime for body recombination (τ_b) is shortened because of the competing processes. This is exactly the same effect as is shown by an atom in an excited state; the mean lifetime of the state is shortened by having more than one way to decay.
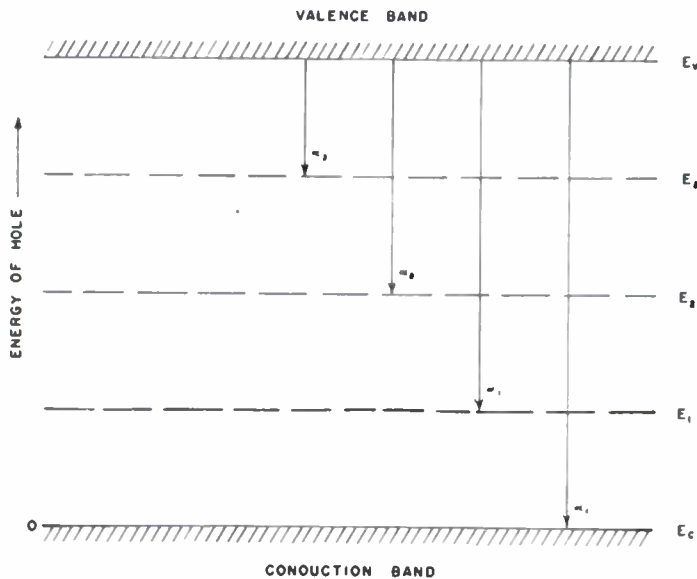
Fig. 1—Energy levels for holes in an *n*-type semiconductor.

Shockley, Read, and Hall[21a] have discussed the transition probabilities defined in (7) so we will not discuss them except to make the following remark: Addition of the transition probabilities as an aid implies that the localized states of different energies are homogeneously distributed throughout the solid so that the probability for a hole to encounter a particular specie of localized states is proportional to the density of said specie. If the material is inhomogeneous so that localized states of energy E_1 are localized within a region V_1, and localized states of energy E_2 are localized in another region V_2, the states E_1 and E_2 are not competing and one does not add the transition probabilities as in (8).

For an inhomogeneous model the individual regions are considered separately, and (4) and (6) are used for each region with an appropriate mean lifetime. The impedance of an inhomogeneous model resembles that of a network of series and parallel impedances. The noise spectrum of a homogeneous and of an inhomogeneous model will differ in an important way. This will be discussed further in Section VIII. We assume a homogeneous model in this paper.

Considering surface recombination, it is clear that surface states are spatially localized away from a majority of the body states. Therefore, the recombination at surface states cannot be considered as a competing process with body recombination and one is forced to use the three-dimensional form of (4).

The three-dimensional equation with surface recombination has been studied by Shockley.[23] He reaches the conclusion that when recombination on the surface is small compared to body recombination the effect of sur-

face recombination is simply to modify the effective lifetime. He defines an effective lifetime τ_p as

$$1/\tau_p = 1/\tau_b + 2s/w, \tag{9}$$

where s is the recombination rate on the surface and $2w$ is the width of sample. The meaning of this is that one can use the one-dimensional form of (4) and incorporate the effect of surface recombination on the lifetime of a hole by "pseudo" trapping states homogeneously distributed throughout the body of the material.

We shall assume in our study of noise in *p-n* junctions that surface recombination can be accounted for in this manner. The geometry of existing junctions is in favor of this, while in long, thin filaments it may be necessary to use the three-dimensional equations.

For a *p-n* junction we define

$\alpha_s = w/2s = 1/\tau_s$ transition probability per unit time for a hole to be trapped in (10) the "pseudo" body state E_s.

Then we have

$\alpha = (\alpha_c + \alpha_1 + \alpha_2 + \cdots) + \alpha_s$ transition probability per unit time for a hole to be removed (11) from the valence band,

$1/\tau_p = \alpha = 1/\tau_b + 1/\tau_s$ mean lifetime of a hole. (12)

We are, therefore, using for our model of a *p-n* junction one in which the effects of body and surface recombination can be effectively described by a mean (shortened) lifetime τ_p. Impedance,[22] photoconductive, and lifetime[24] measurements indicate that a *p-n* junction is described quite well in this manner.

The lifetime of holes in some of these localized states can be very long, that is, they can be metastable states. However, we have not incorporated into our model any effects of these metastable states on the emission characteristics of nearby localized states, such as discussed by Shockley.[2] Such a procedure would lead to a nonlinear equation in place of (4) since the holes would not be independent of one another.

III. Formulation of the Noise Theory

In studying the noise of *p-n* junctions, (4) is interpreted as describing the average behavior of $p(x, t)$ where $p(x, t)$ is now considered to be random variable. Fluctuations in the hole concentration will occur about the average value at each point x. If we divide space from $x = 0$ to infinity into small intervals of volume $\Delta V = A\Delta x$, we can define the number of holes in a region x_i, $x_i + \Delta x_i$ as $m(x_i) = p(x_i)A\Delta x_i$ (see Fig. 2). This number can fluctuate by two main processes. The first is by body and surface recombination. The transition probability for this to occur is given by (11). The second method is by current flow between neighboring regions. It is clear that we have a problem concerning an infinite number of random variables, namely $m(x_i)$, $m(x_i + \Delta x_i) \cdots$. We are therefore studying a problem which

is considerably more complicated mathematically than that of a system with a finite number of random variables such as in lumped circuits. We have not as yet been able to solve the problem in this form.
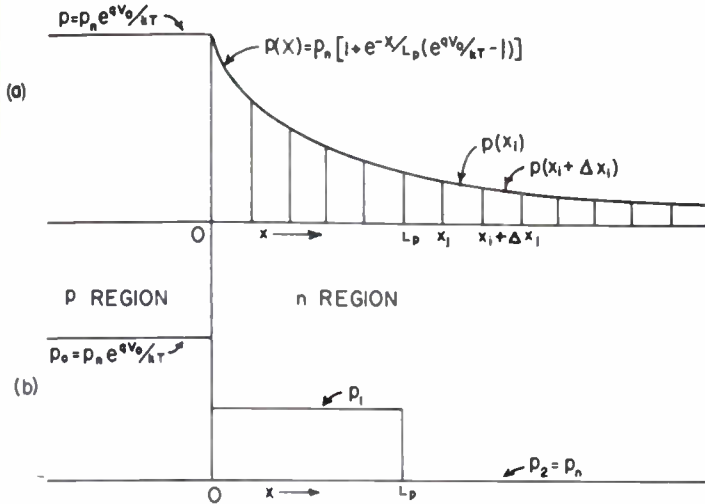
Fig. 2—Hole concentration as a function of position in a *p-n* junction. (a) Distributed parameter solution. (b) Lumped-parameter solution.

Now it is generally possible to approximate the behavior of a distributed parameter system over a limited frequency range by a lumped parameter system. Lumped circuit theory is the classic example of this, the Maxwell equations being partial differential equations. Shockley[20] has shown that for frequencies below $f_p = 1/2\pi\tau_p$ the admittance of a *p-n* junction can be represented quite well by lumped circuit elements. For frequencies higher than f_p the equivalent circuit is that of a transmission line. It therefore seems reasonable to expect that the low-frequency noise characteristics of *p-n* junctions could be obtained by a lumped parameter theory.

We are particularly interested in the frequency range $0 - f_p$ because we should like to see if the $1/f$ spectrum can be explained by electronic noise. It therefore seemed to be of interest to treat the low-frequency range by a suitable lumped parameter approximation. A future theory based on (4) will contain the low- as well as the high-frequency character of the spectrum. Of course, if one wants to work hard enough, a lumped parameter approximation can be continually improved by increasing the number of elements; in this way the high-frequency solution could in principle be attained. We will not do this, but will restrict our attention to low-frequency range which we believe has considerable experimental and theoretical interest. Our program in this paper is, therefore, to

1. obtain a suitable lumped parameter approximation to (4) to be used for describing the voltage-current characteristics of *p-n* junctions in the frequency range $0 - f_p$, and to
2. use this as a basis for a study of the noise.

IV. Lumped Parameter Theory of a *p-n* Junction

Our procedure will be, first, to translate (4) into an infinite set of total differential equations. Next we will use results of Shockley's[20] exact solution to guide us in a reduction of this infinite set of total differential equations to a finite set. After this we will solve for the voltage-current characteristics of this finite set.

We divide space into regions of finite length Δx_i (Fig. 2). Integrating (4) over the volume $\Delta V_i = A\Delta x_i$, we have

$$\iiint_{\Delta V} \frac{\partial p}{\partial t} dV = - \iiint_{\Delta V} \frac{(p - p_n)}{\tau_p} dV$$
$$- \frac{1}{q} \iiint_{\Delta V} \text{div } \overrightarrow{I_p} dV. \quad (13)$$

Applying Gauss's Law, we have

$$\iiint_{\Delta V} \text{div } \overrightarrow{I_p} dV = \iint_{S > \Delta V} \overrightarrow{I_p} \cdot \overrightarrow{n} \, da. \quad (14)$$

Now current flow is assumed to take place only in the x direction, and is uniform over the cross-sectional area. Therefore,

$$A \int_{\Delta x_i} \frac{\partial p}{\partial t} dx = \frac{-A}{\tau_p} \int_{\Delta x_i} (p - p_n) dx$$
$$- \frac{A}{q} [j(x_i) + j(x_i + \Delta x_i)] \quad (15)$$

$$I_p(x_i) = Aj(xi). \quad (16)$$

Defining

$$p(x_i) = \frac{\int_{\Delta x_i} p(x) dx}{\Delta x_i} \quad (17)$$

as the mean density in the region x_i, $x_i + \Delta x_i$, we have

$$\Delta x_i \frac{dp(x_i)}{dt} = - \frac{\Delta x_i}{\tau_p} [p(x_i) - p_n(x_i)]$$
$$- \frac{1}{q} [j(x_i) + j(x_i + \Delta x_i)]. \quad (18)$$

Equation (18) is the continuity equation for the number of holes in the region bounded by the planes x_i, $x_i + \Delta x_i$. The first term on the right-hand side represents the effects of recombination, while the second represents the net current flow across the two boundary planes at x_i and $x_i + \Delta x_i$. Equation (18) represents an infinite set of equations since the current flow is determined by conditions in the regions on either side of x_i as well as in it. We can write down an approximate expression for the current using (6):

$$I(x_i) \cong AqD_p \left[\frac{p(x_i) - p(x_i - \Delta x_i)}{\Delta x_i} \right]$$

$$I(x_i + \Delta x_i) \cong AqD_p \left[\frac{p(x_i) - p(x_i + \Delta x_i)}{\Delta x_i} \right]. \quad (19)$$

Equation (19) becomes exact in the limit $\Delta x_i \to 0$.

We now want to reduce this infinite set of total differential equations to a finite set that will provide a reasonably accurate description of a *p-n* junction for

frequencies $<f_p$. To do this we examine the solution of the exact problem.

The steady-state solution for $p(x)$[20] is (Fig. 2(a))

$$p(x) = p_n[1 + e^{-x/L_p}(e^{qV_0/T} - 1)], \quad (20)$$

where $L_p = (D\tau_p)^{1/2}$ is the diffusion length for holes in the n-region and V_0 is the dc component of the applied voltage. Within a distance L_p, $[p(x)-p_n]$ has fallen to $1/e$ its value at $x=0$; and because of the rapid decay of an exponential function, $p(x)$ approaches the value p_n within a relatively few lengths of L_p. The injected steady-state hole current decays in much the same way.

It therefore would appear that for describing the steady-state characteristics one could divide the space roughly into three divisions as follows:

1. The region to the left of $x=0$ with a constant density equal to $p_0 = p_n \exp(qV_0/kT)$.
2. The region to the right of, for example, $x = 5L_p$ with the constant density $p_2 = p_n$.
3. The region from $x=0$ to, for example, $x = 5L_p$ to be subdivided in accordance with the accuracy to which we want to carry the approximation.

Examination of the low-frequency admittance[20] $A_p = G_p + iS_p$ ((48) and (49)) enables us to simplify the problem still further. A_p is the admittance of two lumped circuit elements—a resistance and capacitance in parallel. The conductance G_p is that of a layer L_p cm in length with a hole conduction corresponding to the density $p_n \exp(qV_0/kT)$.

The reactive part of the admittance corresponds to a capacity roughly of size AL_p.[20] We conclude that a first approximation to the distributed parameter system can be attempted using the three regions mentioned above, modifying statements 2 and 3 to read

2. The region to the right of $x = L_p$ has the constant density $p_2 = p_n$.
3. The region from $x=0$ to $x = L_p$ is not to be further subdivided, letting p_1 be the mean density of holes in the region $x=0$ to $x = L_p$ and $m_1 = AL_p p_1$ be the total number of holes in this region. Fig. 2(b) shows this divisioning.

Our infinite set of differential equations then reduces to the single equation,

$$L_p A \frac{dp_1}{dt} = -\frac{AL_p}{\tau_p}(p_1 - p_n) + \frac{I_{01}}{q} - \frac{I_{12}}{q}, \quad (21)$$

where

I_{01} = net current flow from region 0 to 1. This is the current measured in an external circuit.

I_{12} = net current flow from region 1 to 2. I_{12} will be less than I_{01} due to recombination within region 1.

Now I_{01} is the result of diffusion into and out of region 1 across $x=0$, and will depend on the concentration of holes in regions 0 and 1. We can write down an approximate expression using (19).

Let i_{10} = flow of current from 1 to 0.

i_{01} = flow of current from 0 to 1.

$$I_{01} = i_{01} - i_{10} \cong qAD_p\left(\frac{p_0 - p}{L_p}\right)$$

$$= \alpha A D_p\left(\frac{p_0 - p_1}{L_p}\right) \quad (22)$$

$$i_{10} = \frac{\alpha A q D_p p_1}{L_p} \quad (23)$$

$$i_{01} = \frac{\alpha A q D_p p_0}{L_p}, \quad (24)$$

where α is a numerical constant of order unity. Using $D_p = L_p^2/\tau_p$, we have $i_{10} = \alpha A q L_p p_1/\tau_p$ and $-(i_{10}/m_1) = (\alpha q/\tau_p)$, and this leads to the simple interpretation that the diffusion current per particle from $1 \rightarrow 0$ is $\alpha q/\tau_p \cong dQ/dt$, where τ_p is the lifetime or duration of a current pulse.

We will carry α as a parameter and will show that none of the qualitative results of the noise analysis depends on assigning a numerical value to it. Its value can be estimated by requiring $I_{12}/I_{01} = 1/e$ in accordance with the exact result.[20]

We have reduced the partial differential equation to one total differential equation (21), and the currents are

$$I_{01} = \frac{\alpha q D_p}{L_p}(p_0 - p_1), \qquad I_{12} = \frac{\alpha q D_p}{L_p}(p_1 - p_2),$$

where

$$p_0 = p_n e^{qV_0/kT} \quad \text{and} \quad p_2 = p_n. \quad (25)$$

Equation (21) will form the basis of our noise theory, p_1 being then considered a random variable. Before going into noise theory we will solve for the voltage-current characteristics of the junction on the basis of (21).

We will do this for two reasons: First, because we will need the impedance of the junction in the noise theory, and second, to show that such an approximation does describe quite well the junction at frequencies below f_p. We will consider improvements over this approximation later in the paper.

IV. Voltage-Current Characteristics of a p-n Junction Based on the Lumped-Parameter Approximation

Equation (21) along with the relations of (25) is a nonlinear equation relating I_{01} and V_0, the applied potential. We can solve the steady-state problem exactly, and will use a small-signal analysis for obtaining the impedance characteristics.

Considering first the steady-state solution (using capital letters to indicate steady-state values),

$$\frac{dp_1}{dt} = 0, \quad P_0 = p_n e^{qV_0/kT}, \quad P_2 = p_n \quad (26)$$

$$P_1 = p_n \frac{[1 + \alpha(e^{qV_0/kT} + 1)]}{(1 + 2\alpha)} = p_n e^{qV_1/kT} \quad (27)$$

$$I_{01} = \frac{\alpha q D A}{L_p}(P_0 - P_1)$$

$$= \frac{\alpha(1+\alpha)AqD}{(1+2\alpha)L_p}\, p_n[e^{qV_0/kT}-1] \tag{28}$$

$$I_{01} = I_s[e^{qV_0/kT}-1] \tag{29}$$

$$I_s = \frac{\alpha(1+\alpha)A\, qDp_n}{(1+2\alpha)\, L_p}, \tag{30}$$

where I_s is the reverse saturation current. Equation (29) agrees in form with the exact solution[20] and differs from it only by a numerical factor of order unity.

The current flow into region 2 is

$$I_{12} = \frac{\alpha q D_p A}{L_p}(P_1 - P_2)$$

$$= \frac{\alpha^2}{(1+2\alpha)}\frac{qD_pP_n}{L_p}(e^{qV_0/kT}-1) \tag{31}$$

and

$$I_{12}/I_{01} = \alpha/1+\alpha. \tag{32}$$

The value of α can be fixed by requiring $I_{12}/I_{01}=1/e$, from which $\alpha = 1/e - 1 \cong 1/2$.

The small-signal characteristics can be found in the usual manner.

$$p_0 = p_n e^{q(V_0+\Delta V_0)/kT} = p_n e^{qV_0/kT}[1+(q/kT)\Delta V_0(t)] \tag{33}$$

$$p_1 = p_n e^{q(V_1+\Delta V_1)/kT} = P_1 + \Delta p_1(t) \tag{34}$$

$$p_2 = P_2 = p_n \text{ (the right side being held at ground}$$
$$\text{potential).} \tag{35}$$

The equation for $\Delta p_1(t)$ then is, by (21),

$$\frac{d\Delta p_1(t)}{dt} = \frac{\alpha q p_n e^{qV_0/kT}}{\tau_p kT}\Delta V_0(t) - \frac{(1+2\alpha)}{\tau_p}\Delta p_1. \tag{36}$$

Assuming a sinusoidal voltage for $\Delta V_0(t)$,

$$\Delta V_0 = \Delta V_{0\omega}e^{i\omega t}, \qquad \Delta p_1 = \Delta p_{1\omega}e^{i\omega t}. \tag{37}$$

Substituting (37) in (36), we obtain

$$\Delta p_{i\omega} = \frac{\alpha}{(1+2\alpha)}\frac{q\Delta V_{0\omega}}{kT}p_n e^{qV_0/kT}\left(\frac{1}{1+i\omega\tau'}\right) \tag{38}$$

$$\tau' = \tau_p/(1+2\alpha) = \tau_p/2 \quad \text{for} \quad \alpha = 1/2. \tag{39}$$

The injected ac hole current is

$$\Delta I_{01} = \frac{\alpha q D_p A}{L_p}(\Delta p_0 - \Delta p_1) = \Delta I_{01\omega}e^{i\omega t} \tag{40}$$

$$\Delta I_{01\omega} = \Delta V_{0\omega}\frac{\alpha q^2 A D_p}{L_p kT}p_n e^{qV_0/kT}\left[1-\frac{\alpha/(1+2\alpha)}{1+i\omega\tau'}\right] \tag{41}$$

$$\Delta I_{01} = A_{0g}\Delta V_0, \qquad A_{0g} = G + iS \tag{42}$$

$$G = \frac{1}{R} = \frac{\alpha(1+\alpha)\,q^2 A D_p\, p_n e^{qV_0/kT}}{(1+2\alpha)\,L_p kT} \tag{43}$$

The equivalent circuit is shown in Fig. 3(a) and 3(b), where

$$\frac{1}{R_{01}} = \frac{1}{R}\frac{(1+2\alpha)}{[1+2\alpha-\alpha e^{q(V_0-V_1)/kT}]} \tag{44}$$

$$\frac{1}{R_{13}} = \frac{1}{R_{1T}}+\frac{1}{R_{12}} = \frac{1}{R}\frac{(1+2\alpha)}{[\alpha e^{q(V_0-V_1)/kT}]} \tag{45}$$

$$C = q^2 A L_p P_1/kT \tag{46}$$

$$Z_{0g} = \frac{1}{A_{0g}} = R_{01} + \frac{R_{13}}{1+i\omega C R_{13}} \tag{47a}$$

$$1/Z_{0g} = A_{0g} = \frac{1}{R}\left[\frac{1+i\omega\tau_p/1+\alpha}{1+i\omega\tau_p/1+2\alpha}\right]. \tag{47b}$$

It is convenient to combine R_{1T} and R_{12} since in this approximation they are in parallel.
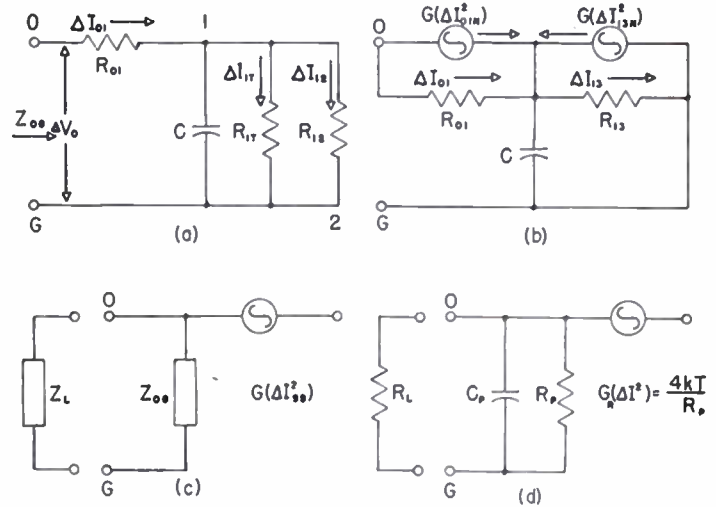
Fig. 3—(a) Equivalent circuit of a *p-n* junction lumped-parameter solution. (b) Equivalent circuit with noise generators. (c) Representation of (b) by Norton's Theorem. (d) Low-frequency equivalent circuit by Shockley's solution with a Nyquist noise generator.

This equivalent circuit is in qualitative agreement with Shockley's,[20] shown in Fig. 3(d), his values of R_p and C_p being

$$\frac{1}{R_p} = \frac{Aq^2 p_n D}{kTL_p}e^{qV_0/kT} \tag{48}$$

$$C_p = q^2 p_n e^{qV_0/kT}L_p/2kT. \tag{49}$$

$R = R_{01} + R_{13}$ and C are in good agreement with R_p and C_p. The location of C in Fig. 3(b) is not correct, but it would require many more elements to push it forward to being nearly directly across $0-g$. We shall be satisfied temporarily with this representation and go on to the noise studies, after which methods of improving the result will be considered.

V. Evaluation of Electronic Noise in a *p-n* Junction on the Basis of the Lumped-Parameter Approximation

We are now ready to use the lumped-parameter approximation of a *p-n* junction as a basis for studying its noise characteristics. Equations (28) and (31) describe the average behavior of current flow as related to the average density of holes. We now consider effects of fluctuations in the hole density on the current observed

in an impedance Z_L connected across 0-g, Fig. 3(b); p_1 is considered to be a Markoffian random variable. We shall use the mathematical techniques of Markoffian random processes, but will attempt to give a word description of our procedure, restricting the mathematical details to Appendix A. For the interested reader the theory of Markoffian random processes is presented in a most interesting manner in the paper of Uhlenbeck and Wang.[25]

The end result of our analysis is to obtain an expression for the power spectrum of the short-circuit noise current generator $G(\Delta I^2_{ss})$ at terminals 0-g, Fig. 3(c). Knowing this current generator and the impedance Z_{0g}, one can write down the expression for the noise observed in any load impedance by the use of conventional circuit techniques (Norton's theorem).

We arrive at this result by first finding expressions for the power spectrum of the noise current generators, $G(\Delta I^2_{01N})$ and $G(\Delta I^2_{13N})$, to be associated with resistors R_{01} and R_{13}, respectively, as shown in Fig. 3(b).

If we were only concerned with thermal equilibrium situations, we could write down the appropriate noise generators by the Nyquist result. That is,

$$G(\Delta I^2_{01N}, V_0 = 0)df = \left(\frac{4kT}{R_{01}}\right)df \quad (50)$$

$$G(\Delta I^2_{13N}, V_0 = 0)df = \left(\frac{4kT}{R_{13}}\right)df. \quad (51)$$

It is just the deviations from this that we are interested in, particularly the dependence on the applied voltage and other parameters.

We find the power spectrum of the noise generators by deriving the power spectrum $G(\Delta m^2)$ of $m = AL_p p_1$. Knowing $G(\Delta m^2)$, we can find $G(\Delta I^2_{01N})$ and $G(\Delta I^2_{13N})$.

The power spectrum of m is determined by how m can change in an infinitesimal time Δt. We have discussed in Section II the appropriate transition probabilities for body and surface recombination, and they are given in (7), (10), and (11). Now since the holes are considered to be statistically independent (this is inherent in using a linear equation such as (4)), the probability for m to decrease by unity in an infinitesimal time Δt by body and surface recombination will be

prob. $[m$ to change to $m - 1$ in $\Delta t]$

$$= m(\alpha_c + \alpha_1 + \alpha_2 + \cdots + \alpha_s)\Delta t = m\Delta t/\tau_p. \quad (52)$$

Now when no current is flowing, the law of detailed balance requires that the probability that m increase by unity in Δt by body and surface excitation be

prob. $[m$ to increase by one in $\Delta t] = \dfrac{m_n}{\tau_p} \Delta t \quad (53)$

$$m_n = AL_p p_n.$$

Similar arguments can be made for m to change by diffusion of particles across the planes $x = 0$, $x = L_p$. We have then for $x = 0$,

prob. $\begin{bmatrix} m \to m - 1 \text{ in } \Delta t \text{ by diffusion out} \\ \text{across } x = 0 \end{bmatrix}$

$$= \frac{\alpha m \Delta t}{\tau_p} \quad (54)$$

prob. $\begin{bmatrix} m \to m + 1 \text{ in } \Delta t \text{ by diffusion in} \\ \text{across } x = 0 \end{bmatrix}$

$$= \frac{\alpha A L_p P_0 \Delta t}{\tau_p}. \quad (55)$$

Note that we are neglecting fluctuations in P_0 itself. This is in accordance with the procedure of assigning a fixed boundary condition $P_0 = p_n \exp(qV_0/kT)$ to the region immediately to the left of $x = 0$. We can account for the noise in the p region by the Nyquist theorem since it is ohmic in character. Noise resulting from concentration fluctuations of the majority carrier should be small when the acceptor states are nearly all ionized.

Similar postulates are made for diffusion across $x = L_p$. These postulates characterize the random process completely. The details of obtaining the appropriate noise currents are given in Appendix A. It is there shown that the average (more precisely the conditional average) behavior of m is described by

$$\frac{d}{dt}\langle m(m_0, t)\rangle = -(1 + 2\alpha)\langle m(m_0, t)\rangle$$

$$+ \frac{AL_p}{\tau_p}[\alpha P_0 + p_n + \alpha P_2], \quad (56)$$

where $\langle m(m_0, t)\rangle = AL_p \langle p_1(t)\rangle$. We note the relation of this equation to (21) and see that our noise theory is very closely related to the average behavior of the system.

The appropriate noise currents to be associated with the resistances R_{01} and R_{13} are

$$G(\Delta I^2_{01N})df = \frac{4kT}{R_{01}}\left[\frac{(1 + 2\alpha)e^{-q(V_0 - V_1)/kT} - \alpha}{(1 + \alpha)}\right]df \quad (57)$$

$$G(\Delta I^2_{13N})df = \left(\frac{4kT}{R_{13}}\right)df. \quad (58)$$

We note that at $V_0 = V_1 = 0$ both noise currents reduce to the Nyquist value. However, under biased conditions $G(\Delta I^2_{01N})$ differs in an important way from the Nyquist value.

The power spectrum of the short-circuit noise generator $G(\Delta I^2_{ss})$, Fig. 3(c), is found by circuit analysis to be

$$G(\Delta I^2_{ss})df = \frac{4kT}{R}\left[\frac{(1 + \alpha) + \alpha e^{qV_0/kT}}{(1 + 2\alpha)e^{qV_0/kT}}\right]F(\omega\tau_p, \alpha)df \quad (59)$$

$$F(\omega\tau_p, \alpha) = \left[\frac{1 + (\omega\tau_p)^2/(1 + 2\alpha)(1 + \alpha)}{1 + \left(\dfrac{\omega\tau_p}{1 + 2\alpha}\right)^2}\right]. \quad (60)$$

For low frequencies, $F(\omega\tau_p < 1, \alpha) = 1$. For $V_0 = 0$,

$$G(\Delta I^2_{ss}, V_0 = 0)df = \frac{4kT}{R} F(\omega\tau_p, \alpha)df, \quad (61)$$

which is the Nyquist formula at low frequencies. For a large reverse bias, $V_0 \ll 0$

$$G(\Delta I^2_{ss}, V_0 \ll 0)df$$

$$= \frac{4kT}{R}\left(\frac{1 + \alpha}{1 + 2\alpha}\right) e^{qV_0/kT}F(\omega\tau_p, \alpha)df. \quad (62)$$

For a large forward bias, $V_0 \gg 0$

$$G(\Delta I^2_{ss}, V_0 \gg 0)df = \frac{4kT}{R}\left(\frac{\alpha}{1 + 2\alpha}\right)F(\omega\tau_p, \alpha)df. \quad (63)$$

Before discussing these formulas, we make a simple check on their qualitative validity.

VI. Use of Nyquist Criteria for Checking and Simplifying the Lumped-Parameter Result

We can apply the Nyquist theory to Shockley's distributed parameter solution to check our lumped-parameter solution. Shockley[2] has shown that the admittance of a p-n junction is

$$A_p = \frac{qp_n\mu_p}{l_p} e^{qV_0/kT}(1 + i\omega\tau_p)^{1/2} \quad (64)$$

for $\omega\tau_p \ll 1$. This reduces to (48) and (49). The equivalent low-frequency circuit is shown in Fig. 3(d). Now the Nyquist theory states that in thermal equilibrium the resistance R_p has a noise current associated with it whose power spectrum is

$$G_R(\Delta I^2)df = \left(\frac{4kT}{R_p}\right)df. \quad (65)$$

We can therefore calculate the noise power in a load resistor R_L by the circuit in Fig. 3(d). If the load is matched, $R_L = R_p$

$$G(P_a)df = \frac{kT\,df}{1 + \left(\frac{\omega\tau_p}{4}\right)^2}. \quad (66)$$

Now a similar calculation based on our lumped-parameter result at zero applied voltage gives

$$G(P_a)df = kTB(\omega\tau_p, \alpha)df, \quad (67)$$

where $B(\omega\tau_p, \alpha)$ is a spectrum involving the characteristic time τ_p in a somewhat more complicated manner than (66) because of the difference in location of the condenser in the lumped-parameter solution. The lumped-parameter solution is in good agreement with (66) in that the available power at low frequencies is kT and the spectrum is flat to frequencies of the order $1/2\pi\tau_p$, at which point it begins to fall off. It is plotted in Fig. 4.

We can use the Nyquist solution to improve and simplify our lumped-parameter solution by the following argument:

Examination of the expression for $G(\Delta I_{ss})$, (59), shows that the spectral character of $G(\Delta I_{ss})$ is independent of the applied voltage, that is, $F(\omega\tau_p, \alpha)$ is independent of V_0. Similarly, the spectral character of Z_{0g}, (47b), is independent of V_0, even though R_{01}, R_{13}, and C are all voltage dependent. It therefore seems reasonable to adapt the spectral characteristics of the Nyquist solution. That is, we change our circuit to agree with the Nyquist value at $V_0 = 0$. The result of this is to shift the capacitor across 0–g. The noise generator to be associated with this circuit is

$$G_p(\Delta I^2_{ss})df = \frac{4kT}{R_p}\left[\frac{(1 + \alpha) + \alpha e^{qV_0/kT}}{(1 + 2\alpha)e^{qV_0/kT}}\right]df. \quad (68)$$
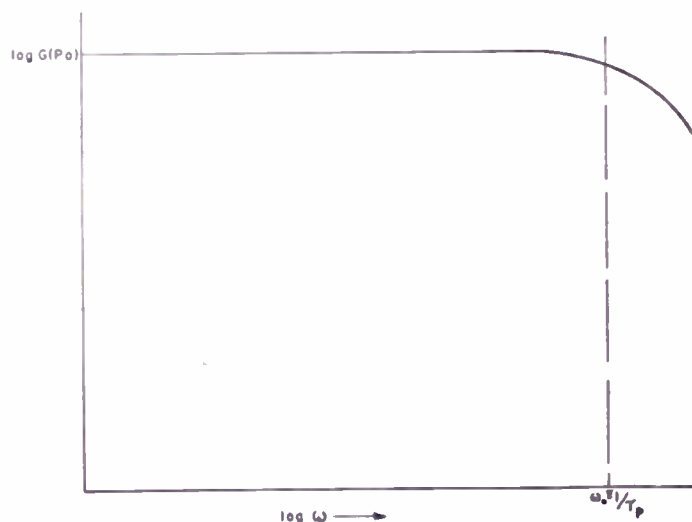
Fig. 4—Spectrum of the available noise power of a p-n junction.

Noise analyses made with this circuit will agree with the Nyquist solution at $V_0 = 0$. It gives the voltage dependence for the short-circuit current generator as found by our lumped-parameter analysis.

While we have no reason to doubt the validity of this procedure, we should like to remark that for values of R_L of the order R_p a noise analysis based on Figs. 3(c) and 3(d) (with $G_p(\Delta I_{ss})$) will be nearly identical. The chief advantage of Fig. 3(d) is its simplicity.

VII. Equivalent Circuit for Analysis of Electronic Noise in p-n Junctions and Related Devices

In Appendix B the effects of electron conduction in the P region are included and also the capacity of the transition region is added. We then arrive at Fig. 5 as an equivalent noise circuit that can be used as a basis for analyzing noise in p-n junction rectifiers, transistors, and related devices. The expressions for the parameters in Fig. 5 are given in (104) and (105) of Appendix B.

The analysis of electronic noise in a p-n junction transistor is a straightforward application of Fig. 5. For each junction the circuit of Fig. 5 is connected, the quiescent operating point of the junction being speci-

fied. The effects of noise in the base region and other ohmic regions can be incorporated by the appropriate Nyquist generators. It does not seem likely that concentration fluctuations in the regions where the current is carried by the majority carrier will produce appreciable deviations from the Nyquist value, particularly at room temperature, since the donors and acceptors are nearly completely ionized.
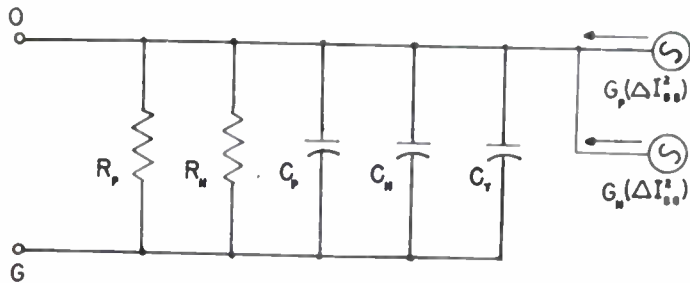
Fig. 5—Equivalent circuit of a *p-n* junction with noise generators.

Straightforward application of the small-signal transistor equations can then be made. the noise generators, of course, constituting the signals.

VIII. Discussion

We will use (59) and the equivalent circuit of Fig. 3(c) as a basis for analyzing certain properties of *p-n* junctions and related devices.

An important consideration is the dependence of the available noise power of *p-n* junctions and the noise figure of *p-n* junction transistors on certain parameters, such as bias voltage, physical size, current densities, and the like. It has been pointed out that point-contact transistors operate with very high current densities, and it is of interest to see if concentration fluctuations of the minority carrier lead to a dependence of noise figure on current density, and so on.

The available noise power is the noise power delivered to a matched resistance. From (59) and Fig. 3(c) we have

$$G_{p\text{-}n}(P_a)df = kT\left[\frac{(1+\alpha)+\alpha e^{qV_0/kT}}{(1+2\alpha)e^{qV_0/kT}}\right]df \quad \omega\tau_p \ll 1. \quad (69)$$

We note that $G_{p\text{-}n}(P_a)$ is dependent only on the bias voltage, and does not exhibit any dependence on size or correlation with current density. This is in contrast to the situation in bulk semiconductors where noise due to number fluctuations has a definite relation to the current. For bulk semiconductors we have

$$G_{b.s.}(P_a)df = \frac{K\tau_p(I/A)^2}{[1+(\omega\tau_p)^2]}df + kTdf, \quad (70)$$

where I/A is the mean current density and K is a constant not involving the current or size of the sample. A derivation of this can be found in footnote reference 8. The second term in (70) arises from velocity fluctuations of the holes while the first term arises from fluctuations in the number of holes in the valence band. At zero voltage the concentration fluctuations contribute no noise, the Nyquist noise being related to velocity fluctuations.

In a *p-n* junction the noise has been entirely related to concentration fluctuations; we have neglected effects of velocity fluctuations. The basic difference in (69) and (70) arises because current flow in bulk semiconductors is proportional to the electric field and flow resulting from concentration gradients is neglected, while we have assumed the opposite situation for a *p-n* junction.

Although concentration fluctuations inherent in the electronic system of bulk semiconductors can be used to explain the correlation of noise phenomena with the average current, we now conclude that such a relation does not exist in *p-n* junctions.

We carry this one step further and calculate the noise figure of a *p-n* junction transistor in Appendix C and find the noise figure to be approximately unity (zero db), independent of bias voltages, size, current densities, and the like.

It is of interest to compare these results with those for point-contact devices. Weisskopf's[6] basic formula for the available noise power of a point-contact rectifier is

$$G_{p.c.}(P_a)df = \frac{kT}{2}\left[1 + e^{-qV_0/kT}\right]df. \quad (71)$$

Equations (69) and (71) are qualitatively the same, the difference apparently arising from recombination; since, if recombination is neglected, it is easily shown that $(1+2\alpha)\to2\alpha$ and $(1+\alpha)\to\alpha$ and (69) reduces to (71).

The coefficients involving α will be precisely determined in an exact solution of the problem, and our result (69), should be considered as giving only the correct order of magnitude when $\alpha = 1/e - 1$. None of the qualitative features of our analysis will depend on assigning a precise value to α.

We see by (71) that there is no dependence on size or correlation with density of the available noise power for point-contact rectifiers and a calculation similar to the one in Appendix C leads to a noise figure of the order of unity. We have based these calculations on noise generated at the junctions, neglecting noise generated in the base and other homogeneous regions. It does not seem likely that noise generated in the homogeneous regions can account for the 40-db difference in noise figure between *p-n* and point-contact transistors.

We conclude that there is no dependence on size or current density of the noise figure of transistors if the transistor is subject only to fluctuations inherent in the electronic system and, therefore, that the 40-db difference in noise figure between *p-n* and point-contact types must be related to a fluctuation of Class II.

Some remarks in addition to those found in Section VI will now be made concerning the spectrum.

A first point of interest is that the spectrum of the noise voltage appearing across an external load is dependent on the impedance of the load. This illustrates the close relation of the noise to the impedance of the junction and of the load. This is in contrast to the noise in bulk semiconductors caused by fluctuations in the number of charge carriers, the spectrum, (70) being de-

termined by τ_p independent of the load (except for capacitive load effects).

However, in so far as the junction itself is concerned, the spectral character of the short-circuit current generator and of Z_{v_g} are determined by τ_p and α as discussed in Section VI. Also the available noise power as calculated by the Nyquist solution contains τ_p in the characteristic jump spectrum, $1/[1+(\omega\tau_p/4)^2]$.

It is somewhat tempting at this point to use the distribution of localized states (Fig. 1) as a basis for superposing spectra of the $1/1+(\omega\tau_i)^2$ type, where $\tau_i = 1/\alpha_i$ is the mean lifetime associated with each localized state E_i. On the basis of the model we have used, such a procedure would not be correct. We have already incorporated the distribution of localized states into (12), the net effect being to shorten the lifetime of a hole in the valence band.

The inherent relation between the macroscopic equations of the system and the corresponding noise equation must be preserved if a noise theory is to be developed that is consistent with a given model. Therefore, if one is to superpose spectra at this point, it would also be necessary to change the macroscopic equation of the system. Such a superposition would correspond to an inhomogeneous model with localized regions of localized (trapping) states, the regions having different transition probabilities for recombination and emission.

The macroscopic electronic properties of p-n junctions appear to be accurately [23,24] described by (4), and therefore we conclude that electronic noise cannot explain a $1/f$ spectrum even when a distribution of localized states is present. Thus a second source of noise must exist in p-n junctions that is the cause of the $1/f$ spectrum.

It is perhaps worth mentioning here that in noise studies based on the ionic modulation mechanism the question of how a distribution of ionic lifetimes should be incorporated into the model arises. If the ionic states are competing in the sense described in Section II, then the net effect will be to shorten the lifetime of the mobile ion. The correct procedure in the case of ions cannot be established as easily as in the electronic problem because analogous experiments to impedance, lifetime, and photoconductive studies have not been made. Thus there is no independent check on the procedure. The procedure of adding the $1/1+(\omega\tau_i)^2$ spectra has been closely related to the model in the ionic studies referred to in footnote reference 12. These comments are not to be taken as a criticism of that work, but merely as a point to be considered generally.

The inherent relation between the macroscopic equation of a p-n junction and the corresponding noise equation can be used as a guide in the interpretation of noise measurements. For example, the characteristic frequency, $f_p = 1/2\pi\tau_p$, is involved in the admittance and in the photoconductive response as well as in the noise spectrum. Now a $1/f$ noise spectrum does not exhibit any correlation with impedance and photoconductive studies of p-n junctions. On the basis of this, one would conclude that the second source of noise in semiconductors, rectifiers, and transistors completely overshadows

the electronic noise. However, this is not quite the case since deviations from the $1/f$ law have recently been reported, and there are other experiments that do indicate a correlation between noise and certain macroscopic electronic properties of bulk semiconductors and p-n junction devices. We mention these experiments briefly in order to show that evidence exists that electronic noise is being observed experimentally, and to show that the $1/f$ law does not have the universal character that is sometimes attributed to it.

The correlation studies of Montgomery and Shockley[26] are evidence for the appearance of the lifetime, τ_p, in the noise measurements. While the over-all spectrum reported by Montgomery is of the $1/f$ variety, which does not, of course, contain τ_p, the correlation studies do show that τ_p is present in the noise.

A deviation from the $1/f$ law has been reported by Van der Ziel and Herzog[2] in measurements on germanium filaments. They have fitted the spectrum with the law

$$G(f) = \frac{AI^2}{f} + \frac{BI^2}{1+(\omega\tau_p)^2}, \tag{72}$$

where τ_p is the lifetime of the minority carrier, I is the current, and A and B are suitable parameters. The value of τ_p found in the noise measurements was in reasonably good agreement with the lifetime of holes as found by Montgomery,[26] and also agrees in order of magnitude with the lifetime as determined by Haynes.[24]

Van der Ziel and Anderson[2] have recently reported a correlation between the impedance and the noise spectrum of a p-n junction biased in the forward direction. In the paper they suggest that such a correlation might be expected to exist, and they also use the Weisskopf equation for interpreting the measurements.

A deviation from the $1/f$ spectrum has been observed in noise measurements[27] made on photoconductive films by Miss Lummis and myself. A definite correlation of the noise spectrum with the photoconductive response time was observed over a temperature range 75–300° K.

Conclusion

On the basis of the results of this and other papers and considering the over-all experimental studies, we conclude that noise inherent in the electronic system of p-n junction devices, point-contact devices, and bulk semiconductors can account for only a portion of the measured noise in these devices and that there is a second important source of noise. This second source is of the Class II, or modulation, noise. The modulation noise has, to a large extent, obscured the electronic noise when currents flow through the device. In particular, we conclude that the modulation noise is necessary to explain

1. the noise figures of transistors being appreciably larger than unity,
2. the large difference between noise figures of point-contact and p-n junction transistors,
3. the $1/f$ portion of the noise spectrum.

An important practical as well as scientifically interesting question to be answered in examining noise sources of Class II is to determine whether or not this noise is fundamental to the device or whether improvements in materials and design can reduce its magnitude or even completely eliminate it.

We will discuss this matter more fully in the paper under preparation that was mentioned in Section I.

ACKNOWLEDGMENT

APPENDIX A

We will outline here the method by which we obtain the power spectrum of the noise current generators $G(\Delta I^2_{01N})$ and $G(\Delta I^2_{13N})$, to be associated with R_{01} and R_{13} in Fig. 3(b). We set up the Kolmogoroff-Fokker-Planck[28] (K.F.P.) equation for the p-n junction (in our lumped parameter approximation) from which the power spectrum of m, $G(\Delta m^2)$ is obtained. We then use $G(\Delta m^2)$ to find expressions for the noise current generators. We are in effect transforming from the K.F.P. equation to the Langevin[28] equation of the system. Uhlenbeck and Wang[25] present the connection between these two formulisms. The Langevin equations are very useful for noise calculations. In this formulism the ordinary "circuit equations" are used, and the fluctuation character of the system enters through noise voltage or current generators. In a p-n junction, (21) will be the Langevin equation after the proper noise current generators have been added.

We assume the fluctuations in m to constitute a stationary Markoffian[29] random process. A one-dimensional Markoffian random process is characterized by the probability function

$P(m_0, t_0/m_1 t_1) =$ conditional probability of m at time t_1
         if $m(t = t_0) = m_0$.

For stationary processes,

$$P(m_0, t_0/m_1 t_1) = P(m_0/m, t) \quad t = t_1 - t_0$$

and

$$\text{limit } P(m_0/m, t) = W(m) = \text{stationary probability of } m$$
$$t \to \infty.$$

The basic equation of the random process is the K.F.P. equation.

$$\frac{dP}{dt}(m_0/m, t) = \sum_{s;\, s \neq m} P(m_0/s, t) Q(s, m) - P(m_0/m, t) \sum_{r;\, r \neq m} Q(m/r). \quad (73)$$

A random process is characterized by postulating how the system can change in an infinitesimal time interval Δt; that is, a "Stosszahlansatz" is made

$$P(r/s, \Delta t) = Q(r/s)\Delta t + \text{order } (\Delta t)^2. \quad (74)$$

We have discussed in Sections II and V the appropriate transition probabilities, and have from (52) and (55)

$$
\begin{aligned}
&P(m/m - 1, \Delta t) \\
&= (\alpha_c + \alpha_1 + \alpha_2 + \cdots + \alpha_s) m \Delta t \\
&= \frac{m}{\tau_p} \Delta t \\
&P(m/m + 1, \Delta t) = (A L_p p_n / \tau_p)\Delta t \\
&P(m/m - 1, \Delta t) = \frac{\alpha m}{\tau_p} \Delta t \\
&P(m/m + 1, \Delta t) = \alpha_s L_p P_0 \Delta t / \tau_p \\
&P(m/m - 1, \Delta t) = \alpha m \Delta t / \tau_p \\
&P(m/m + 1, \Delta t) = \alpha A L_p P_2 \Delta t / \tau_p
\end{aligned}
$$

by body and surface recombination and emission · by diffusion across $x = 0$ (75) · by diffusion across $x = L_p$.

Using (75) and (74) in (73), we find the K.F.P. equation for this random process to be

$$
\begin{aligned}
\frac{dP}{dt}(m_0/m, t) =\ & P(m_0/m - 1, t) \frac{A L_p}{\tau_p}(p_n + P_0 + P_2) \\
& + P(m_0/m + 1, t)\left(\frac{1 + 2\alpha}{\tau_p}\right)(m + 1) \\
& - P(m_0/m, t)\left[\frac{A L_p}{\tau_p}(p_n + P_0 + P_2)\right. \\
& \left. + \frac{(1 + 2\alpha)}{\tau_p} m\right]. \quad (76)
\end{aligned}
$$

The power spectrum of m can be found without solving (76). (It can be solved, but we will not use its solution.) We do this by finding the correlation function and using the Wiener-Khintchine (W.K.) theorem[30] to obtain the power spectrum from the correlation function. Proceeding,

$$
\begin{aligned}
\rho(m, t) &= \langle m(0) m(t) \rangle - \langle m \rangle^2 \text{ correlation function of } m \\
&= \sum_{m_0} \sum_m W(m_0) P(m_0/m, t) m_0 m - \langle m \rangle^2 \\
&= \sum_{m_0} W(m_0) m_0 \langle m(m_0, t) \rangle - \langle m \rangle^2 \quad (77)
\end{aligned}
$$

$$G(\Delta m^2) = 4 \int_0^\infty \rho(m, t) \cos 2\pi f t\, dt \text{ power spectrum}$$

of m (W.K. Theorem),   (78)

where

$$\langle m(m_0, t) \rangle = \sum_m P(m_0/m, t)m \quad \text{conditional av. of } m$$

$$\langle m(m_0, 0) \rangle = m_0 \qquad \langle m(m_0, \infty) \rangle = \langle m \rangle$$
$$\langle m \rangle = \sum_m W(m)m \quad \text{stationary average of } m. \tag{79}$$

From the K.F.P. equation we obtain a differential equation for $\langle m(m_0, t) \rangle$:

$$\frac{d}{dt} \langle m(m_0, t) \rangle = \sum_m \frac{dP}{dt}(m_0/m, t)m$$

$$= -\frac{(1 + 2\alpha)}{\tau_p} \langle m(m_0, t) \rangle$$

$$+ \frac{AL_p}{\tau_p}[p_n + P_0 + P_2] \tag{80}$$

$$P_0 = p_n e^{qV_0/kT}. \qquad P_2 = p_n \text{ as in (26)}.$$

Equation (80) bears a close relation to (21), as discussed in Section II.

The stationary average value of m is by (80):

$$\frac{d}{dt} \langle m(m_0, \infty) \rangle = 0$$

$$\langle m(m_0, \infty) \rangle = \langle m \rangle = \frac{AL_p p_n}{\tau_p}\left[\frac{1 + \alpha(e^{qV_0/kT} + 1)}{(1 + 2\alpha)}\right]$$

$$= AL_p P_1 \tag{81}$$

in agreement with (27).

It is convenient to use the deviation from the equilibrium value;

$$\Delta m = m - \langle m \rangle, \quad \Delta m_0 = m_0 - \langle m \rangle. \tag{82}$$

Equation (80) then is

$$\frac{d}{dt} \langle \Delta m(\Delta m_0, t) \rangle = -\frac{(1 + 2\alpha)}{\tau_p} \langle \Delta m(\Delta m_0, t) \rangle, \tag{83}$$

whose solution is

$$\langle \Delta m(\Delta m_0, t) \rangle = \Delta m_0 e^{-t/\tau'}$$
$$\tau' = \tau_p/1 + 2\alpha, \text{ as in (39)}. \tag{84}$$

From which, using (77), we find the correlation function to be

$$\rho(m, t) = \langle \Delta m^2 \rangle e^{-t/\tau'} \tag{85a}$$
$$\langle \Delta m^2 \rangle = \sum_m W(m)(m - \langle m \rangle)^2 = \langle m \rangle. \tag{85b}$$

That $\langle \Delta m^2 \rangle = \langle m \rangle$ can easily be shown by writing down the second-moment equation from the K.F.P. equation.

The power spectrum of m is found using (78) and (85a).

$$G(\Delta m^2) = \langle \Delta m^2 \rangle 4 \int_0^\infty e^{-t/\tau'} \cos \omega t\, dt \tag{86}$$

$$= 4\tau'\langle \Delta m^2 \rangle \left[\frac{1}{1 + (\omega\tau')^2}\right].$$

It is interesting to note that no noise voltage or current generators enter the K.F.P. equation or (80) for the conditional average.

We can only sketch briefly the transformation to the Langevin formulism, referring interested readers again to reference[25]. The Langevin equation (or circuit equation) is (see Fig. 3(b))

$$q\frac{d}{dt}\Delta m(t) - \Delta I_{01} + \Delta I_{13} = \Delta I_{01N} + \Delta I_{13N}; \tag{87}$$

from (40) and a similar equation for ΔI_{13}, we have

$$\Delta I_{01}(t) = -\frac{\alpha q}{\tau_p}\Delta m(t), \quad \Delta I_{13} = \frac{(1 + \alpha)q}{\tau_p}\Delta m(t)$$
$$D_p = L_p^2 \tau_p, \qquad \Delta m = AL_p\Delta p_1. \tag{88}$$

Note that $\Delta P_0 = 0$ since there is no applied ac voltage. V_0 is involved only in the stationary conditions (81).

$\Delta I_{01N}(t)$ and $\Delta I_{13N}(t)$ are the current noise generators of R_{01} and R_{13}, respectively. Our problem is to find expressions for the power spectrum of $\Delta I_{01N}(t)$ and $\Delta I_{13N}(t)$.

To find $G(\Delta I^2_{01N})$ and $G(\Delta I^2_{13N})$ we translate (87) to the frequency plane by considering $\Delta m(t)$ for a very long time T_0 and taking $\Delta m(t) = 0$ outside the time interval. Then we can develop the resulting function into a Fourier Integral;

$$\Delta m(t) = \int_{-\infty}^{+\infty} df A(f)e^{i2\pi ft}. \tag{89}$$

Similarly, for the noise current generators,

$$\Delta I_{01N}(t) = \int_{-\infty}^{+\infty} df B(f)e^{i2\pi ft}$$
$$\Delta I_{13N}(t) = \int_{-\infty}^{+\infty} df C(f)e^{i2\pi ft}. \tag{90}$$

Substituting in (87), we have

$$\int_{-\infty}^{+\infty} df\left[\left(i\omega + \frac{1 + 2\alpha}{\tau_p}\right)qA(f) - B(f) - C(f)\right] = 0. \tag{91}$$

Now taking the limit as $T_0 \to \infty$ and defining,

$$G(\Delta m^2) = \lim_{T_0 \to \infty} \frac{2}{T_0}|A(f)|^2 \quad \text{power spectrum of } m$$

$$G(\Delta I^2_{01N}) = \lim_{T_0 \to \infty} \frac{2}{T_0}|B(f)|^2 \quad \text{power spectrum of } \Delta I_{01N} \tag{92}$$

$$G(\Delta I^2_{13N}) = \lim_{T \to \infty} \frac{2}{T_0}|C(f)|^2 \quad \text{power spectrum of } \Delta I_{13N}.$$

We get from (90), (91), and (92)

$$\left[\omega^2 + \left(\frac{1 + 2\alpha}{\tau_p}\right)^2\right]q^2 G(\Delta m^2) = G(\Delta I^2_{01N}) + G(\Delta I^2_{13N}). \tag{93}$$

Now we know $G(\Delta m^2)$ from our K.F.P. formulism, and it is given in (86).

If there were only one noise generator to be found, the solution of (93) would be trivial. With two generators to be evaluated and only one equation the solution is somewhat more involved. The solution to this apparent difficulty is that the mean-square fluctuation in $m(\Delta m^2)$ must be independent of the frequency spectrum of the fluctuations. Therefore, if we consider that one branch is of very high resistance, for example $R_{13} \rightarrow \infty$, then the fluctuations in m occur through R_{01}. We then go back to the K.F.P. equation and find the power spectrum of m for this case. It will not be the same as for R_{13} finite, although $\langle \Delta m^2 \rangle$ will be independent of the spectrum. Considering then $R_{13} \rightarrow \infty$, that is, we assume temporarily that recombination and diffusion across $x = L_p$ are considered small compared to diffusion across $x = 0$. We then have

$$\frac{d}{dt} \langle \Delta m(\Delta m_0, R_{01}, t) \rangle = - \frac{\alpha}{\tau_p} \langle \Delta m(\Delta m_0, R_{01}, t) \rangle \quad (94)$$

and

$$G(\Delta m^2, R_{01}) = \frac{4(\tau_p/\alpha)\langle \Delta m^2 \rangle}{1 + \left(\frac{\omega \tau_p}{\alpha}\right)^2}. \quad (95)$$

We see that the spectrum has changed, but the mean-square fluctuation in m

$$\int_0^\infty G(\Delta m^2, R_{01}) df = \langle \Delta m^2 \rangle \quad (96)$$

is independent of the new time constant τ_p/α.

The appropriate Langevin equation (on the frequency plane) with $R_{13} = \infty$ is

$$[\omega^2 + \alpha^2/\tau_p^2] q^2 G(\Delta m^2, R_{01}) = G(\Delta I^2_{01N}). \quad (97)$$

Solving for $G(\Delta I^2_{01N})$, using (95), we have

$$G(\Delta I^2_{01N}) = 4\alpha q^2 \langle \Delta m^2 \rangle / \tau_p. \quad (98)$$

Similarly, we find

$$G(\Delta I^2_{13N}) = \frac{4q^2(1 + \alpha)}{\tau_p} \langle \Delta m^2 \rangle. \quad (99)$$

The analogous property of RC circuits in thermal equilibrium is that the mean-square voltage fluctuation across a condenser is $\langle e_c^2 \rangle = kT/C$, independent of the spectrum of e_c. Note that we are not assuming thermal equilibrium but only that the fluctuation in m constitutes a stationary random process, so that $\langle \Delta m^2 \rangle$ is a well-defined quantity.

We can check our expressions for $G(\Delta I^2_{01N})$ and $G(\Delta I^2_{13N})$ in two ways: First, using them together in (93) should yield the correct spectrum and mean-square fluctuation for m, and second, in thermal equilibrium they should reduce to the Nyquist values.

Substituting (98) and (99) in (93), we have

$$G(\Delta m^2) = \frac{G(\Delta I^2_{01N}) + G(\Delta I^2_{13N})}{\left(\frac{1 + 2\alpha}{\tau_p}\right)^2 q^2 \left[1 + \left(\frac{\omega \tau_p}{1 + 2\alpha}\right)^2\right]} \quad (100)$$

$$= \frac{4\left(\frac{\tau_p}{1 + 2\alpha}\right)\langle \Delta m^2 \rangle}{1 + \left(\frac{\omega \tau_p}{1 + 2\alpha}\right)^2}, \quad (101)$$

which agrees with (86). Note also that

$$\int_0^\infty G(\Delta m^2) df = \langle \Delta m^2 \rangle.$$

The second check can be made by substituting from (43), (44), (45), (81), and (85b) into (98) and (99). This leads to expressions for $G(\Delta I^2_{01N})$ and $G(\Delta I^2_{13N})$ in terms of R_{01}, R_{13} and the quiescent voltages.

$$G(\Delta I^2_{01N}) = \frac{4kT}{R_{01}} \left[\frac{(1 + 2\alpha)e^{q(V_0 - V_1)/kT} - \alpha}{(1 + \alpha)}\right] \quad (102)$$

$$G(\Delta I^2_{13N}) = \frac{4kT}{R_{13}}, \quad (103)$$

and at $V_0 = V_1 = 0$ we have the Nyquist result. Our analysis has carried us beyond the Nyquist formula since (102) and (103) can be used when steady currents are flowing, and it is clear that deviations from the Nyquist law occur.

We are now in a position to use the equivalent circuit, Fig. 3(b), with the noise current generators (102) and (103), to find such quantities as the available noise power in a load resistance connected across terminals 0-g and the power spectrum of the noise observed at this point.

Appendix B

Inclusion of the Effects of Electron Conduction in the p Region

Our analysis has only considered hole conduction in the n-region. Shockley has shown that electron conduction in the p-region is simply an additive effect to the hole current. The noise resulting from concentration fluctuations of electrons is statistically independent of the hole fluctuations and the analysis would follow the same lines as in Sections II–VI and Appendix A. We then have for our equivalent noise generator

$$G_N(\Delta I^2_{ss}) df = \frac{4kT}{R_N} \left[\frac{1 + \alpha + \alpha e^{qV_0/kT}}{(1 + 2\alpha)e^{qV_0/kT}}\right]$$

$$1/R_N = \frac{A q^2 n_p D_{NC} e^{qV_0/kT}}{kT L_N} \quad (104)$$

$$C_N = q^2 n_p \frac{L_N}{2kT} e^{qV_0/kT}.$$

Adding the noise generators and the admittances, we have

$$G_N(\Delta I^2_{ss}) + G_p(\Delta I^2_{ss})$$

$$= 4kT\left[\frac{1 + \alpha + \alpha e^{qV_0/kT}}{(1 + 2\alpha)e^{qV_0/kT}}\right]\left(\frac{1}{R_p} + \frac{1}{R_N}\right)\Delta f$$

$$\frac{1}{R_p} = \frac{Aq^2 p_n D_p e^{qV_0/kT}}{kTL_p} \qquad (105)$$

$$C_p = q^2 p_n e^{qV_0/kT}L_p/2kT$$

$C_T =$ (see sections 2 and 3 of footnote reference 20 for

     expressions for C_T).

here we have added the capacitance of the transition region, C_T. Finally, we have Fig. 5 as an equivalent circuit that can be expected to describe the electronic noise of a *p-n* junction over the frequency range $\rightarrow 1/2\pi\tau_p$.

Appendix C

Noise Figure of a p-n Junction Transistor and of a Point-contact Transistor

The noise figure of a *p-n* junction transistor as determined by electronic fluctuations can be accurately evaluated by use of the equivalent circuit of Fig. 5 for each junction and by following the procedure outlined in Section VII.

We will make a qualitative estimate of the noise figure by assuming that the emitter noise current reaches the collector circuit without amplification or loss. The power gain of the transistor is then simply the ratio of the impedance of the collector circuit to that of the emitter circuit. We assume the collector circuit to be biased reverse, and the emitter forward.

We calculate the noise figure by first finding the noise power in a matched load in the collector circuit resulting from the noise of the two *p-n* junctions. (We neglect other ohmic noise sources as discussed in Section VII.) Using as a signal the Nyquist noise of a resistor equal to R_e, the emitter resistance, we then find the noise power in the load. The ratio of these two powers is the noise figure.

Proceeding, we have in the collector circuit, using 105),

$$G(\Delta I_e^2) + G(\Delta I_c^2) = \frac{4kT}{R_e}\left[\frac{(1 + \alpha) + \alpha e^{qV_e/kT}}{(1 + 2\alpha)e^{qV_e/kT}}\right]$$
$$+ \frac{4kT}{R_c}\left[\frac{(1 + \alpha) + \alpha e^{qV_c/kT}}{(1 + 2\alpha)e^{qV_c/kT}}\right] \quad (106)$$

and

$$R/R_e \sim e^{q(V_e - V_c)/kT} = \text{power gain}. \qquad (107)$$

Now assuming $qV_e/kT \gg 1$ and $qV_c/kT \ll 1$, we have

$$G(P_a) = [G(\Delta I_e^2) + G(\Delta I_c^2)]R_c/4 \qquad (108)$$

$$G(P_a) = kT\left[\frac{\alpha}{(1 + 2\alpha)}e^{q(|V_c| + |V_e|)/kT} + \frac{(1 + \alpha)}{(1 + 2\alpha)}e^{q|V_c|/kT}\right]. \quad (109)$$

Assume now a current "signal" at the emitter equal to

$$G_s = 4kT/R_e. \qquad (110)$$

The "signal" power in the load is

$$G_s(P_a) = \frac{4kT}{R_e} \times \frac{R_c}{4} = kTe^{q(|V_c| + |V_e|)/kT} \qquad (111)$$

and the noise figure (N.F.) is

$$\text{N.F. } (p\text{-}n) = \frac{G(P_a)}{G_s(P_a)} = \left[\frac{\alpha + (1 + \alpha)e^{-|qV_c|/kT}}{1 + 2\alpha}\right]$$

$$\cong \frac{\alpha}{1 + 2\alpha} = 1/e + 1 \text{ for } \alpha = 1/e - 1. \quad (112)$$

A similar analysis for a point-contact transistor can be made using Weisskopf's equation (see Section VIII, (71)), and we have

$$\text{N.F.}(p.c.) \cong 1/2. \qquad (112)$$

Bibliography

1. H. Nyquist, "Thermal agitation of electric charge in conductors," *Phys. Rev.*, vol. 32, p. 110; 1928.
2. J. B. Johnson, "Thermal agitation of electricity in conductors," *Phys. Rev.*, vol. 32, p. 97; 1928. For experimental investigations, see the following:
   J. L. Lawson and G. E. Uhlenbeck, "Threshold signals," MIT Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, N. Y., vol. 24; 1950.
   E. B. Moulin, "Spontaneous Fluctuations of Voltage," Oxford University Press, New York, N. Y.; 1938.
   P. H. Miller, Jr., "Noise spectrum of crystal rectifiers," Proc. I.R.E., vol. 35, p. 252; 1947.
   C. J. Christensen and G. L. Pearson, "Spontaneous resistance fluctuations in carbon microphones and other granular resistances," *Bell Sys. Tech. Jour.*, vol. 15, p. 197; 1936.
   G. B. Herzog and A. Van der Ziel, "Shot noise in germanium single crystals," *Phys. Rev.*, vol. 84, p. 1249; 1951.
   J. Bernamont, "Variations of resistance of a metallic conductor of small volume," vol. 198, pp. 1755, 2144; 1934. Also, "Fluctuations in resistance of thin films," *Proc. Phy. Soc.* (London); 1937. And also, "Fluctuations in voltage at the barriers of a metallic conductor of small volume carrying a current," *Ann de Chemie et de Phys.*, vol. 7, p. 71; 1937.
   H. C. Montgomery and W. Shockley, "Noise in germanium related to fluctuations in hole concentration," *Phys. Rev.*, vol. 78, p. 646A; 1950.
   W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y., p. 342; 1950.
   H. C. Montgomery, "Background noise in transistors," *Bell Lab. Rec.*, p. 400; 1950.
   R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, p. 367; 1949.
   R. L. Wallace, Jr. and W. J. Pietenpol, "Some Circuit Properties and Applications of *n-p-n* Transistors," *Bell Sys. Tech. Jour.*, vol. 30, p. 530; 1951. Also, Proc. I.R.E., vol. 39, p. 753; 1951.
   A. Van der Ziel and R. L. Anderson, "On the Shot Effect in *P-N* Junctions," paper presented at the Semiconductor Device Conference; June, 1952. The authors kindly gave me a prepublication copy.
   F. L. Lummis and R. L. Petritz, "The relation between response time and noise in photoconductors; II Experiment," *Phys. Rev.*, vol. 86, p. 660A; 1952.
   H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," MIT Rad. Lab. Series, McGraw-Hill, New York, N. Y., no. 15, chap. 6; 1950.
3. J. L. Lawson and E. Uhlenbeck, *op. cit.*, sec. 4, 5.
4. P. H. Miller, Jr., *op. cit.*
   C. J. Christensen and G. L. Pearson, *op. cit.*
   G. B. Herzog and A. Van der Ziel, *op. cit.*
   J. Bernamont, *op. cit.*
   H. C. Montgomery and W. Shockley, *op. cit.*
   W. Shockley, *op. cit.*
   H. C. Montgomery, *op. cit.*
   R. M. Ryder and R. J. Kircher, *op. cit.*
   R. L. Wallace, Jr. and W. J. Pietenpol, *op. cit.*
   A. Van der Ziel and R. L. Anderson, *op. cit.*
   F. L. Lummis and R. L. Petritz, *op. cit.*
   H. C. Torrey and C. A. Whitmer, *op. cit.*
5. J. Bernamont, *op. cit.*

6. V. Weisskopf, N.D.R.C., MIT Rad. Lab. Series, No. 133; May 12, 1943. An outline of this paper will be found in H. C. Torrey and C. A. Whitmer, *op. cit.*
7. C. J. Bakker and G. Heller, "Brownian motion in electric resistances," *Physica*, vol. 6, p. 262; 1939.
8. L. Brillouin, "Theory of electrons in metals," *Helv. Phys. Acta*, vol. 7, supp. 2, p. 47; 1934.
   B. I. Davydov and G. Gurevich, "Voltage fluctuations in semiconductors," *Jour. Phys.* (URSS). vol. 7, p. 138; 1943.
   M. Surdin, "Fluctuations in the thermionic current and the flicker effect," *Jour. Phys. Radium*, vol. 10, p. 188; 1939. Also, "A theory of electrical fluctuations in semiconductors," *Jour. Phys. Radium*, vol. 12, p. 783; 1951.
   J. H. Gisolf, "On spontaneous current fluctuations in semiconductors," *Physica*, vol. 15, p. 825; 1949.
   H. C. Montgomery and W. Shockley, *op. cit.*
   R. L. Petritz and A. J. F. Siegert, "On the theory of noise in semiconductors," *Phys. Rev.*, vol. 79, p. 215A; 1950.
   R. L. Petritz, "On Noise in Semiconductors Caused by Fluctuations in the Number of Conduction Electrons," Thesis, Northwestern Univ., Evanston, Ill., 1950.
   W. Shockley, *op. cit.*
   A. Van der Ziel, "On the noise spectra of semiconductor noise and flicker effect," *Physica*, vol. 16, p. 359; 1950.
   H. F. Matare, "Statistical Fluctuations in Semiconductors," *Z. Naturf.*, vol. 4a, p. 275; 1949. Also, "Background noise in semiconductors," *Jour. Phys. Radium*, vol. 10, p. 364; 1949, and vol. 11, p. 130; 1950. Also, "Randschichtwechselwerkung und Statistische Schwankungen beim Dreielektroden Kristall," *Zeit fur Phys.*, vol. 131, p. 82; 1951.
9. H. C. Montgomery, *op. cit.*
10. W. Schottky, "Uber Spontane Stromschwankungen in Verschiedenen Elektrizitatsteitern," *Ann. Phys.* (Lpz.), vol. 57, p. 541; 1918.
11. W. Schottky, "Small shot effect and flicker effect," *Phys. Rev.*, vol. 28, p. 74; 1926.
12. G. G. Macfarlane, "A theory of contact noise in semiconductors," *Proc. Phys. Soc.* (London), vol. 63B, p. 807; 1950. Also, "A theory of flicker noise in valves and impurity semiconductors," *Proc. Phys. Soc.* (London), vol. 59, p. 336; 1947.
   W. Miller, "Theory of Frequency Spectrum of Noise Due to Density Fluctuations," BuShips Contract Nohs 34144, Tech. Report No. 11; May 11, 1948.
   J. M. Richardson, "The linear theory of fluctuations arising from diffusion mechanisms," *Bell Sys. Tech. Jour.*, vol. 29, p. 117; 1950.
   F. K. DuPré, "A suggestion regarding the spectral density of flicker noise," *Phys. Rev.*, vol. 78, p. 615; 1950.
   R. L. Petritz, "On the diffusion theory of noise in rectifiers and transistors," *Phys. Rev.*, vol. 87, p. 189A; 1952.
   V. Weisskopf, *op. cit.*
   M. Surdin, *op. cit.*
   A. Van der Ziel, *op. cit.*
   H. F. Matare, *op. cit.*
13. A. Van der Ziel, *op. cit.*
14. J. M. Richardson, *op. cit.*
15. F. Seitz, "The Modern Theory of Solids," McGraw-Hill Book Co., Inc., New York, N. Y., p. 494; 1940.
16. J. M. W. Milatz and H. A. Van der Velden, "Natural limit of measuring radiation with a bolometer," *Physica*, vol. 10, p. 369; 1943.
17. R. H. Fowler, "Statistical Mechanics," Cambridge University Press, Cambridge, England; 1936.
18. R. L. Petritz, "A theory of contact noise," *Phys. Rev.*, vol. 87, p. 535; 1952.
19. R. L. Petritz, "Theory of Noise in P-N Junctions Resulting from Lattice Temperature Fluctuations," paper presented at the Semiconductor Device Conference, Champaign, Ill.; June, 1952.
20. W. Shockley, "The theory of P-N junctions in semiconductors and P-N junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, p. 435; 1949.
21. H. B. G. Casimir, "On Onsager's principle of microscopic reversibility," *Rev. Mod. Phys.*, vol. 17, p. 343; 1945.
21a. W. Shockley and W. T. Read, Jr., "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, p. 835; September, 1952.
   R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, p. 387; July, 1952.
   L. Onsager, "Reciprocal relations in irreversible processes I," *Phys. Rev.*, vol. 37, p. 405; 1931 and "II," *Phys. Rev.*, vol. 38, p. 2265; 1931.
   S. R. DeGroot, "Thermodynamics of Irreversible Processes," Interscience Pub., New York, N. Y.; 1951.
22. 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, p. 637; 1951.
   W. Shockley, M. Sparks, and G. K. Teal, "P-N transistors," *Phys. Rev.*, vol. 83, p. 151; 1951.
23. W. Shockley, *op. cit.*, Appendix V.
24. J. R. Haynes and W. Shockley, "Investigation of hole injection in transistor action," *Phys. Rev.*, vol. 75, p. 691; 1949.
25. M. C. Wang and G. E. Uhlenbeck, "On the theory of brownian motion II," *Rev. Mod. Phys.*, vol. 17, p. 323; 1945.
26. H. C. Montgomery and W. Shockley, *op. cit.* W. Shockley "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y.; 1950.
27. F. L. Lummis and R. L. Petritz, *op. cit.*
28. S. Chandrasekhar, "Stochastic problems in physics and astronomy," *Rev. Mod. Phys.*, vol. 15, p. 50; 1943. See also, footnote reference 25 for a discussion of the K.F.P. and Langevin equations.
29. See footnote reference 25 for the definition of a Markoffian random process.
30. N. Wiener, "Generalized harmonic analyzer," *Acta. Math.*, vol. 55, p. 117; 1930. A. Khintchine, "Correlation theory of stationary stochastic processes," *Math. Ann.*, vol. 109, p. 604; 1934.

An Experimental Investigation of Transistor Noise*

E. KEONJIAN†, SENIOR MEMBER, IRE, AND J. S. SCHAFFNER†, ASSOCIATE, IRE

Summary—Transistor noise is discussed in general and methods of measurement are presented. Experimental results show that noise figures for point-contact transistors are approximately 50 db; for junction transistors they may be as low as 10 db at 1,000 cycles. The effect of the variation of the dc operating point on noise figures and other parameters is discussed.

It was found that the noise figure of point-contact transistors is relatively independent of the dc operating point, while for junction transistors it may vary considerably with the collector voltage and to some extent with the emitter current.

INTRODUCTION

IN RECENT YEARS considerable work has been done on the theory and applications of transistors. However, in current publications devoted to transistors, noise considerations receive little or no attention despite the fact that these considerations are of first importance, for example, if transistors are used in a sensitive amplifier.

In this paper, noise in transistors is discussed and methods are given for measuring noise figures and open-circuit noise voltages together with some experimental

results for General Electric point-contact and *p-n-p* junction transistors.

I. THE EQUIVALENT CIRCUIT OF A TRANSISTOR WITH NOISE

At the present time, two types of transistors have been developed: point contact[1] and junction.[2,3] Point-contact transistors have at present (March, 1952) a better frequency response but a high noise level, while junction transistors have a considerably lower noise level but also a poorer frequency response. Representative noise figures for these two types of transistors are 50 db and 15 db, respectively.

It is convenient to describe the noise of a transistor by adding two noise-voltage generators with mean-square voltages E_{ne} and E_{nc} in the emitter and collector leads (Fig. 1). It is then assumed that the other elements of the equivalent circuit are noiseless.
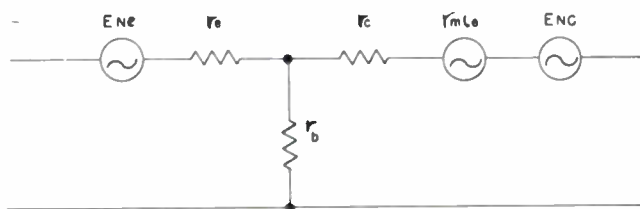
Fig. 1—Transistor amplifier with noise voltage generators.

Representative values for E_{nc} and E_{ne} for point-contact transistors are given by Ryder and Kircher as $100\mu V$ and $1\mu V$, respectively, measured at a frequency of 1,000 cycles for a bandwidth of one cycle. Corresponding voltages for junction transistors are smaller.

Transistor noise differs from thermal noise insofar as it is not white noise. The noise power per cycle bandwidth due to a transistor varies approximately inversely with frequency.

$$P_n \cong k/f. \qquad (1)$$

(A more precise equation would be $P_n = kf^{-1.1}$.)

The total noise power in a frequency band limited by the frequencies f_2 and f_1 is then

$$P_n = \int_{f_1}^{f_2} \frac{k}{f} df = k(\ln f_2 - \ln f_1) = k \ln \frac{f_2}{f_1}. \qquad (2)$$

Corresponding equation for thermal noise would be

$$P_n = k'(f_2 - f_1). \qquad (3)$$

Noise voltages and powers are usually given for a bandwidth of one cycle at a frequency of 1,000 cycles.

II. THE NOISE FIGURE

Fig. 2 shows a transistor amplifier. It is assumed that

[1] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, p. 367; 1949.
[2] R. L. Wallace, Jr. and W. J. Pietenpol, "Some circuit properties and applications of *n-p-n* transistors," PROC. I.R.E., vol. 39, pp. 753–767; July, 1951.
[3] J. S. Saby, "Recent developments in transistors and related devices," *Tele-Tech.*, vol. 10, p. 32; December, 1951.

both the source impedance and the load impedance are resistive. Neglecting thermal noise originating in R_L, the noise figure F may be defined as the total noise power in the output divided by that portion of the output noise that results from thermal agitation in R_g. It is customary to express F in db.
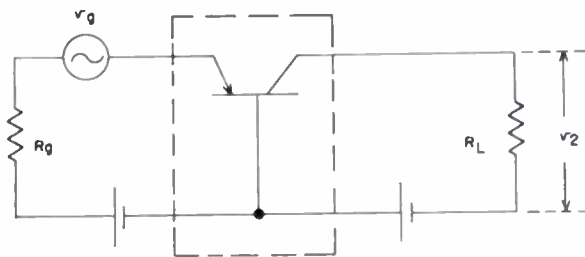
Fig. 2—Transistor amplifier.

If the mean-square value of the noise voltage measured across R_L is V_{nc}^2, then the total noise power in the output is

$$R = \frac{V_{nc}^2}{R_L}. \qquad (4)$$

This is to be divided by that portion of the noise power in the output that results from thermal noise in R_g.

This noise may be represented by a noise-voltage generator in series with R_g. As is well known, the mean-square value of this voltage is

$$E^2 = 4kTR_g(f_2 - f_1), \qquad (5)$$

where k is Boltzmann's constant

$$k = 1.347 \times 10^{-23} \text{ (wattsec per °K)}$$

and T is the absolute temperature in degrees K.

If the ratio of the output voltage v_2 to the signal voltage v_g is $A_1 = v_2/v_g$, (Fig. 2), then the noise power in the output due to thermal agitation in R_g is

$$P_n' = \frac{4kTR_g(f_2 - f_1)}{R_L} \times A_1^2. \qquad (6)$$

The noise figure is, therefore

$$F = \frac{V_{nc}^2}{4kTR_g(f_2 - f_1)A_1^2}. \qquad (7)$$

The noise figures for the three basic types of amplifier circuits (grounded emitter, grounded base, and grounded collector) may be expressed mathematically in terms of equivalent circuit parameters and open-circuit voltages as follows:

Grounded Emitter

$$F = 1 + \frac{1}{4kTR_g(f_2 - f_1)}\left[E_{ne}^2\left(\frac{R_g + r_m + r_b}{r_m - r_e}\right)^2 \oplus E_{nc}^2\left(\frac{R_g + r_e + r_b}{r_m - r_e}\right)^2\right]. \qquad (8)$$

Grounded Base

$$F = 1 + \frac{1}{4kTR_g(f_2 - f_1)}$$
$$\left[E_{ne}^2 \oplus E_{nc}^2 \left(\frac{R_g + r_e + r_b}{r_m + r_b} \right)^2 \right]. \tag{9}$$

Grounded Collector

$$F = 1 + \frac{1}{4kTR_g(f_2 - f_1)} \left[E_{ne}^2 \left(\frac{R_g + r_e + r_b}{r_c} \right)^2 \right.$$
$$\left. \oplus E_{nc}^2 \left(\frac{R_g + r_b}{r_c} \right)^2 \right]. \tag{10}$$

The addition sign $\oplus$ in these equations expresses "addition with attention to any correlation between E_{ne} and E_{nc}." If no correlation exists, then this operation may be replaced by a simple addition.

These equations show that the noise figures are independent of R_L. With respect to R_g, they have a minimum. The values of R_g at which this minimum occurs are

(for grounded base and grounded emitter)

$$R_g \simeq \left[(r_e + r_b)^2 \oplus \frac{E_{ne}^2}{E_{nc}^2} (r_m + r_b)^2 \right]^{1/2} \tag{11}$$

(for grounded collector)

$$R_g \simeq \left[r_b^2 \oplus \frac{E_{ne}^2}{E_{nc}^2} r_c^2 \right]^{1/2}. \tag{12}$$

Experimental curves for the variation of F and R_g are plotted in Fig. 3. They show that the dependence of the noise figure on R_g is not critical.
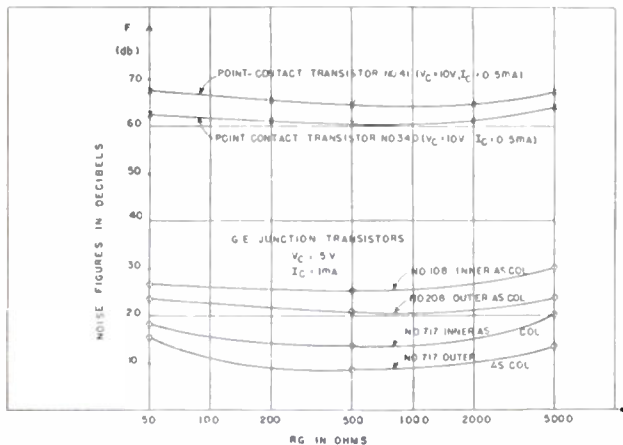
Fig. 3—The noise figure as a function of R_g.

Therefore, it is customary to give only the minimum value of the noise figure with respect to R_g. For the measurements of this report, R_g was selected arbitrarily as 500 ohms.

The symbol F_0 will be used for the noise figure, measured at a frequency of 1,000 cycles for a bandwidth of one cycle. The minimum noise figure for a bandwidth of one cycle at any other frequency is

$$F \simeq F_0 \frac{1,000}{f},$$

which is obtained as follows:

Since V_{nc}^2 is proportional to $\ln f_2/f_1$ and the thermal noise due to R_g is proportional to $(f_2 - f_1)$, then the noise figure for an arbitrary bandwidth limited by the frequencies f_2 and f_1, is

$$F = F_0 \, 1,000 \frac{\ln f_2/f_1}{f_2 - f_1}. \tag{13}$$

If $f_2/f_1 \cong 1$, then

$$\ln \frac{f_2}{f_1} \cong \frac{f_2 - f_1}{f_1}$$

and

$$F \simeq F_0 \cdot \frac{1,000}{f_1}.$$

The noise figure may, therefore, be considered fairly independent of the bandwidth if this bandwidth remains relatively small.

III. The Equivalent Noise—Voltage

For some applications, it is desirable to express the noise of a transistor amplifier by an "equivalent" noise-voltage generator in series with R_g (Fig. 4). If this noise
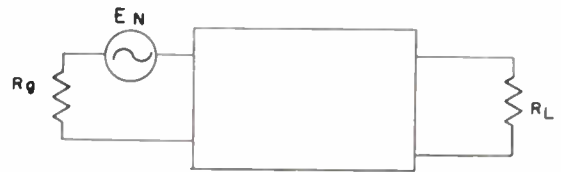
Fig. 4—Amplifier with equivalent noise source.

generator is to produce a voltage V_{nc} across R_L, then its rms value must be

$$E_n = V_{nc}/A_1 \tag{14}$$

This expression may be combined with (7) to give

$$E_n^2 = F4kTR_g(f_2 - f_1). \tag{15}$$

If F is replaced by F_0, then, (13),

$$E_n^2 = F_0 4kTR_g \, 1,000 \ln \frac{f_2}{f_1};$$

or, if the base of the logarithm is changed to ten and numerical values are substituted for k and T,

$$E_n^2 \simeq 3.6 \times 10^{-17} \cdot R_g \cdot F_0 \log \frac{f_2}{f_1} \text{ (volts}^2\text{)}.$$

The available power at the input due to E_n is (the available power is defined as the power transferred to a matched load).

$$P_0 = \frac{E_n^2}{4R_g} = 0.9 \times 10^{-17} \times F_0 \times \log \frac{f_2}{f_1} \text{ (watts). } \tag{16}$$

This has been plotted in Fig. 5. Equation (16) is approximately

$$P_0 \cong 0.9 \times 10^{-17} \times F_0 \cdot \log\left(1 + \frac{\Delta f}{f}\right) \text{ (watts)}, \quad (17)$$

where Δf is the bandwidth and f the center frequency.

Equation (16) shows that the noise power depends on the ratio of the frequencies f_2 and f_1 and not on their magnitudes. Equation (17) shows that for a constant bandwidth the noise power will decrease as the center frequency increases.
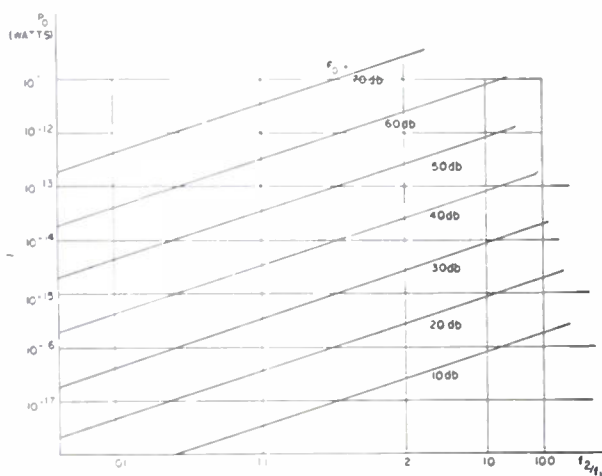
Fig. 5—The available "equivalent noise power."

For many amplifiers only the noise of the first stage will contribute significantly to the noise in the output. The signal-to-noise ratio at the output of a transistor amplifier is then equal to the ratio of the available signal power in the input to the available "equivalent noise power" P_0 (16).

$$\frac{S}{N} = \frac{V_g^2}{4R_g P_0} = \frac{V_g^2}{3.6 \cdot 10^{-17} \cdot R_g \cdot F_0 \cdot \log f_2/f_1}, \quad (18)$$

where V_g is the rms value of the signal voltage.

Using this equation it is possible to determine the maximum permissible noise figure of a transistor for the first stage of an amplifier.

This is illustrated by the following example: Assume that the available signal power is 10^{-8} watts in a band from 50 cycles to 5 kc and that the signal to noise ratio is 40 db. P_0 must, therefore, be less than 10^{-12} watts. For $f_2/f_1 = 100$ this corresponds to a maximum permissible noise figure of 48 db (Fig. 5).

IV. Noise Measurements

A. Measurements of Open-Circuit Noise Voltages

To measure the open-circuit emitter and collector noise voltages, the transistor under test is biased to the desired dc operating point through chokes having practically infinite impedances at all frequencies of interest. The emitter and collector noise voltages are measured by means of a General Radio wave analyzer, having a bandwidth of four cycles Q connected through a cathode follower to the input and output terminals of the transistor.

The meter of the wave analyzer showed slow random fluctuations that made the reading of the noise voltages difficult. In order to facilitate the measurements, it was found convenient to increase the time constant of the dc amplifier of the wave analyzer and to place a large capacitance (2,000 μF) across the meter.

B. The Measurement of Noise-Figures

For the measurement of noise figures, techniques similar to those used for vacuum-tube amplifiers may be employed. One of the most common of these methods is the following: A noise generator in series with R_g is connected to the input of the transistor. (Fig. 6.)

With the generator turned off, the noise power at the output is measured. Then with the noise generator turned on, its voltage level is adjusted so that the power at the output of the transistor is doubled.

If the corresponding rms voltage of the noise generator is E_0, then the noise figure is

$$F = \frac{V_0^2}{4kTR_g(f_2 - f_1)}.$$

The noise figures may also be measured by the following method:

The noise generator in Fig. 6 is replaced by a signal generator and the rms value of the noise-voltage across R_L, corresponding to a bandwidth $(f_2 - f_1)$, is measured.
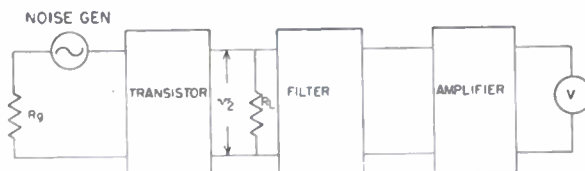
Fig. 6—Circuit for the measurement of noise figures.

Then the ratio $A_1 = v_2/v_g$ of the transistor amplifier is measured by conventional methods where v_g is the output voltage of the signal generator. From these two measurements, the noise figures may be determined as

$$F = \frac{V_n^2}{4kTR_g(f_2 - f_1) \cdot A_1^2}.$$

For the measurements reported in this paper the second method was used. The source impedance R_g was selected as 500 ohms, the load impedance as 10,000 ohms. The measurements were carried out at a frequency of 1,000 cycles and a bandwidth of 4 cycles.

Equations (8) to (10) show that the noise figures for grounded-base and grounded-emitter amplifiers are approximately equal. The average difference for a large number of transistors was found to be 1 db. Therefore, it is unimportant which type of amplifier is used for the measurement of the noise figures.

Fig. 7 shows the variation of the noise figures with frequency for two transistors. These curves agree quite well with (13). No measurements were made at frequen-
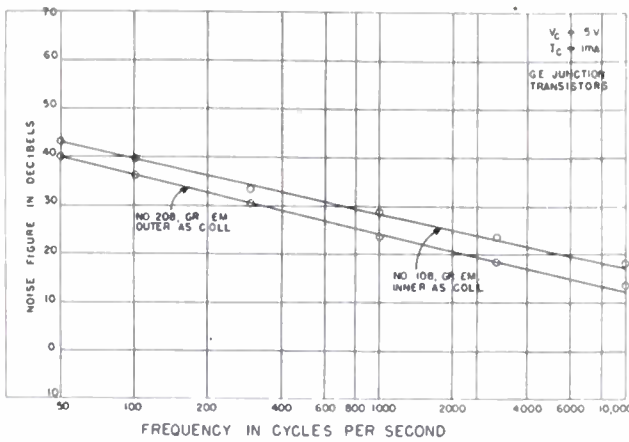
Fig. 7—Transistor noise figure versus frequency.

cies higher than 15 kc since no equipment for such measurements was available. This was true also for the measurement of the open-circuit noise voltages.
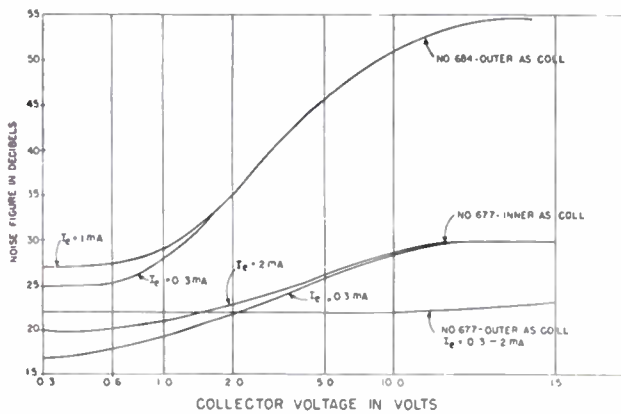
Fig. 8—Transistor noise versus operating point.

Fig. 8 shows the variation of the noise figures of 3 typical transistors with collector current and collector voltage. It was found that for most transistors the noise figure increases with the collector voltage in the manner shown by transistor #684 of Fig. 8. For some transistors, however, the noise figures seem to be almost independent of the collector voltage. Transistor #677 shows this behavior.

It is of interest to know how the other parameters of transistor (α, r_e, r_b, r_c, and C) vary with the dc operating point. Figs. 9 and 10 show these variations for the transistors for which the noise figures were plotted in Fig. 8. As may be seen, r_c increases with V_c, but decreases as I_e increases, while the parameter α remains constant for all V_c and I_e values from 1 to 5 volts and 0.5 to 1 ma, respectively. The parameter r_e is to a first

approximation independent of the collector voltage V_c, but decreases with increasing I_e, ranging from 110–130 ohms at 0.2 ma down to 10–20 ohms at 2 ma. These
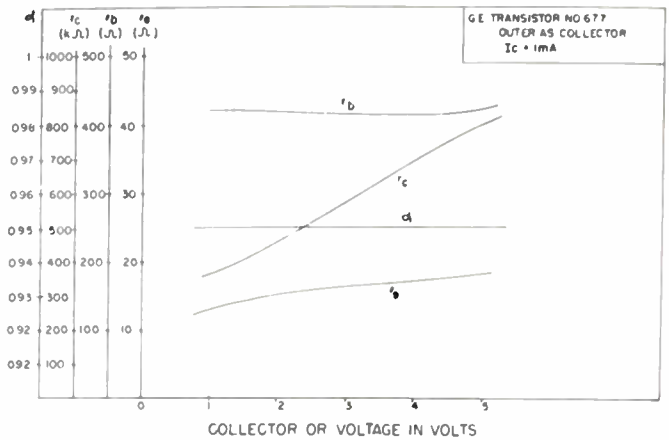
Fig. 9—Transistor parameters versus collector voltage.

curves show that the parameters of the equivalent circuit, as well as the noise figures, vary considerably with the dc operating point.
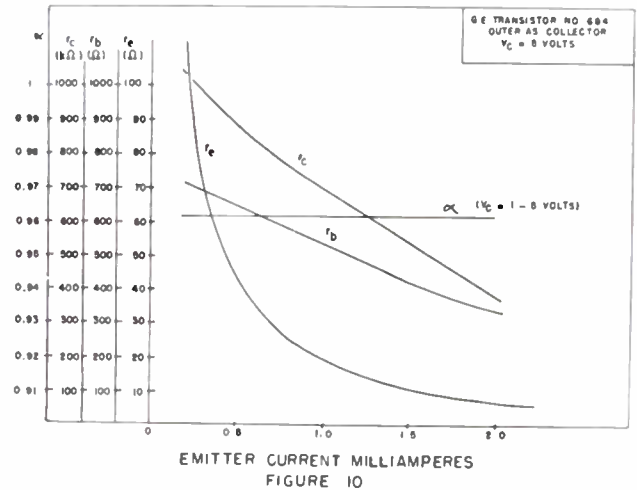
Fig. 10—Transistor parameters versus emitter current.

All these measurements were made on a limited number of experimental transistors of the p-n-p type developed at the Electronics Laboratories of the General Electric Company. Accordingly, they may not be representative of all types of junction transistors.

Transistor Noise in Circuit Applications*

H. C. MONTGOMERY†

Summary—Linear circuit problems involving multiple noise sources can be handled by familiar methods with the aid of certain noise spectrum functions, which are described. Several theorems of general interest dealing with noise spectra and noise correlation are derived. The noise behavior of transistors can be described by giving the spectrum functions for simple but arbitrary configurations of equivalent noise generators. From these, the noise figure can be calculated for any desired external circuit. Illustrative information is given for a number of *n-p-n* transistors.

INTRODUCTION

A SPECIFICATION of the properties of any transmission device to be complete must include information regarding the way in which it contributes background noise to a signal it transmits. In general, the noise behavior of such a device depends on certain properties of the device itself, and also on the circuit impedances with which it is terminated. In applications where noise is an important factor in performance, it is usually of interest to be able to determine the circuit conditions leading to the optimum signal-to-noise ratio.

Simple circuit problems involving sources of noise can often be handled on an intuitive basis by regarding the noise as a group of sinusoids closely spaced in frequency. When the circuit contains several independent sources of noise the contribution of each can be calculated independently and then combined on a mean-square basis, to give the total noise. However, in more complicated problems, especially those involving partly coherent noise sources, the intuitive method becomes unwieldy and a systematic approach is desirable.

The noise behavior of a transistor can be represented by a pair of fictitious noise generators associated with the input and output terminals, according to a theorem due to Peterson and described later in this paper. Since these noise generators need not have the actual internal configuration of the noise sources, they are in general partly coherent. The noise produced by the transistor in connected circuits may be calculated from the properties of the fictitious generators. This is the sort of problem which, in the general case, is not easy to work out by intuitive methods, but which is very straightforward with the systematic approach to be described.

The method of handling linear circuit problems in noise will be developed first. It consists essentially of defining certain functions for describing noise currents and voltages which can be manipulated in circuit equations in the same manner as the complex steady-state currents and voltages of familiar circuit theory. This is followed by a number of useful theorems and relations

based on this type of analysis, which are of a quite general nature. The final part deals with specific applications to transistor problems, including methods of representing noise behavior, conversion formulas, and noise data on a limited number of junction-type transistors.

Throughout the paper it will be assumed that we are dealing with stationary noise processes, by which we mean that the statistical properties which are used to describe them do not change with time, except for statistical fluctuations which tend to decrease as the time interval of averaging increases. The discussion will be limited to linear circuit problems. Except as occasionally noted, there is no implied restriction to Gaussian noise processes.

1. DEFINITION OF IMPORTANT NOISE FUNCTIONS

The functions required to describe the behavior of a system of noise currents and voltages are a *power spectrum* for each noise and a *cross spectrum* for each pair. The power-spectrum concept is quite familiar. The cross spectrum is less well known, although it is generally recognized that the result of adding two noises depends in an important way on the degree of coherence between them. The coherence properties are conveniently described by the cross spectrum. As a basis for establishing the important properties of these spectrum functions, we shall give an analytical derivation starting with the time function representing the noise disturbance.

Suppose $y(t)$ is the time function which represents a noise current or voltage. We may define a Fourier spectrum over any finite time interval by

$$S_1(f) = (2T)^{-1/2} \int_t^{t+T} y_1(t)e^{-i2\pi ft}dt. \qquad (1)$$

By a suitable smoothing along the frequency axis (discussed in the Appendix) we obtain a complex function of frequency the square of whose magnitude is the power spectrum

$$P_1(f) = \text{aver} \ | \ S_1(f) \ |^2 = \text{aver} \ S_1{}^*S_1, \qquad (2)$$

where the star denotes the complex conjugate. The angle of the S function is distributed at random, and does not constitute a significant description of the properties of a single noise.

From this definition it is seen that the term power spectrum is to some extent a misnomer, since it is a description of the statistical properties of a noise current (or voltage). However, it is the power which would result if the noise current (or voltage) were applied to a

resistance of one ohm. The appropriate units are amperes squared (or volts squared) per unit bandwidth.

The cross spectrum between two noise currents is defined as

$$P_{12}(f) = \text{aver } S_1{}^*S_2, \tag{3}$$

where suitable smoothing of both magnitude and angle along the frequency axis is assumed. This is a complex function of frequency, whose existence depends on there being a systematic phase relation between S_1 and S_2, which in turn implies an element of coherence between y_1 and y_2.

The cross spectrum can be measured by sending the two noise currents through identical narrow-band filters and applying the outputs to a device, such as a dynamometer, which indicates the product of the instantaneous values. The average reading of this device is the real part of the cross spectrum at the frequency passed by the filters. The imaginary part is found by shifting the phase of one of the noise currents by 90 degrees. Alternatively, the magnitude and phase can be determined directly by adjusting the phase of one current until a maximum reading is obtained from the product device. The equivalence of this procedure to the analytical definition may be seen by resolving corresponding long portions of each noise current into Fourier series, and considering the products of the terms in the two series (as discussed in the Appendix).

For many purposes it is convenient to use a normalized form of the cross spectrum, given by

$$P_{12}'(f) = P_{12}/(P_1 P_2)^{1/2}. \tag{4}$$

This function is closely related to the correlation between the noise currents in a narrow frequency band, which may be seen as follows:

The correlation between two noise currents $x_1(t)$ and $x_2(t)$ is defined as

$$r_{12} = \overline{x_1 x_2}/(\overline{x_1{}^2} \cdot \overline{x_2{}^2})^{1/2}.$$

Suppose that x_1 and x_2 are the currents which result when the noise currents y_1 and y_2 are passed through the pair of identical filters used to determine the cross spectrum. From the method of measuring the cross spectrum it is evident that the real part of the normalized cross spectrum is the correlation between the noise currents y_1 and y_2 in a narrow band at frequency f. Similarly, the magnitude of the cross spectrum is the maximum correlation which can be achieved. This will be referred to as the *intrinsic correlation* ρ_{12}, and is realized by shifting the phase of one noise current by the angle of the cross spectrum at the frequency in question. Thus we may write

$$r_{12} = \text{actual correlation} = \text{real part of } P_{12}', \tag{5}$$

$$\rho_{12} = \text{intrinsic correlation} = \text{magnitude of } P_{12}'. \tag{6}$$

An extensive discussion of the power spectrum and related matters, with many references, is given by Rice.[1]

The cross spectrum is discussed by Phillips,[2] and the mathematical background is treated in great detail by Wiener.[3]

2. Use of Noise Functions in Circuit Equations

Conventional steady-state circuit theory is based on linear equations giving the relations between complex quantities which describe the magnitude and phase of sinusoidal voltages and currents. These relations are stated in terms of complex impedances, admittances, or transfer functions. We will now show that the same circuit equations can be used to describe the noise behavior of a system merely by substituting the S functions of the preceding section for the steady-state complex currents or voltages.

We note first that for the system of noise currents

$$y_3(t) = y_1(t) + y_2(t),$$

it follows directly from the definition (1) of the S function that

$$S_3(f) = S_1(f) + S_2(f). \tag{7}$$

This is parallel to the relation for steady-state sinusoidal currents, where in complex notation

$$I_3 = I_1 + I_2.$$

Thus, the additive property of the complex S functions is established.

A second basic relation deals with the effect of complex impedance, admittance, or transfer operators on the S functions. If $y_B(t)$ is the noise current which results when $y_A(t)$ is passed through a linear network having a complex transfer constant $A(f)$, it follows that

$$S_B(f) = A(f)S_A(f). \tag{8}$$

This relation was pointed out by Fry[4], and is discussed by Phillips[2] and Guillemin.[5] This is parallel to the steady-state complex relation

$$I_B = A(f)I_A.$$

These two properties are sufficient for all the operations of linear network theory. Hence it is possible to use the S functions in circuit equations in place of steady-state voltages or currents.

In order to distinguish easily between noise currents and voltages, we shall introduce the notation

$$\begin{aligned} I &= S(f) \text{ for a noise current} \\ V &= S(f) \text{ for a noise voltage.} \end{aligned} \tag{9}$$

The bold-face notation will distinguish noise functions from sinusoidal or dc functions.

[1] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. Jour.*, vol. 23, pp. 282–332; July, 1944 and vol. 24, pp. 46–156; January, 1945.

[2] R. S. Phillips, "Theory of Servomechanisms," Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, N. Y., chap. 6; 1947.

[3] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, pp. 117–258; 1930; also N. Wiener, "Interpolation, Extrapolation and Smoothing of Time Series," John Wiley and Sons, Inc., New York, N. Y.; 1949.

[4] T. C. Fry, "The solution of circuit problems," *Phys. Rev.*, vol. 14, pp. 115–136; August, 1919.

[5] E. A. Guillemin, "Communication Networks," vol. II, chap. XI, John Wiley and Sons, Inc., New York, N. Y.; 1935.

The general procedure for linear circuit problems is to write the circuit equations in the same form as for a steady-state situation, using noise spectra I or V for the noise currents or voltages. The equations are manipulated in the customary manner, and usually as a last step desired power and cross spectra are obtained by applying definitions (2) and (3) of these functions.

3. THEOREMS OF GENERAL INTEREST

Several relations will now be derived which are generally useful in circuit work involving noise. These will also serve to illustrate the procedure described in the preceding section.

Addition of Noise Voltages or Currents

Suppose that two sources of noise voltage having power spectra P_1 and P_2 and a cross spectrum P_{12} are placed in series. We wish to find the power spectrum of the resulting voltage. The equation of instantaneous voltages is

$$v_3(t) = v_1(t) + v_2(t).$$

From relation (7), and using the notation (9), we see that the relation of voltage spectrum functions is

$$V_3 = V_1 + V_2.$$

Making a product with the complex conjugate, and using (2) and (3),

$$V_3^*V_3 = V_1^*V_1 + V_1^*V_2 + V_2^*V_1 + V_2^*V_2$$
$$P_3 = P_1 + P_{12} + P_{21} + P_2.$$

From (3), (4) and (5) it follows that

$$P_{21} = P_{12}^*$$
$$P_{12} + P_{21} = 2 \times \text{real part of } P_{12}$$
$$= 2(P_1 P_2)^{1/2} r_{12}. \qquad (10)$$

Using this relation, it is seen that

$$P_3 = P_1 + P_2 + 2(P_1 P_2)^{1/2} r_{12}, \qquad (11)$$

which is a simple and fundamental relation. A relation identical in form would result if we had started with three noise currents.

Theorem on Terminal Noise in a Three-Terminal Network

Relation (10) may be used to establish a useful theorem relating the terminal voltages or currents in a three-terminal network containing noise sources. Referring to the two networks of Fig. 1 we see that

$$v_1(t) + v_2(t) + v_3(t) = 0,$$
$$i_1(t) + i_2(t) + i_3(t) = 0.$$

Except for a change in one sign, which is of no consequence in this case, each of these equations is like the equation of the preceding section, and it is easily verified that the power-spectrum relation in either case is just (11). This closely resembles the formula for one side of a triangle in terms of the other two sides, and the included angle

$$X_3^2 = X_1^2 + X_2^2 - 2X_1 X_2 \cos \theta.$$

The power spectrum at frequency f is proportional to the mean-square voltage or current in a narrow band about f. Hence the root-mean-square noise voltages or noise currents may be represented by the sides of a triangle, where the correlation between any two is the negative of the cosine of the included angle, as shown in Fig. 1. The sign of the correlation will be as stated if one uses the sign conventions shown in Fig. 1 for the currents or voltages.
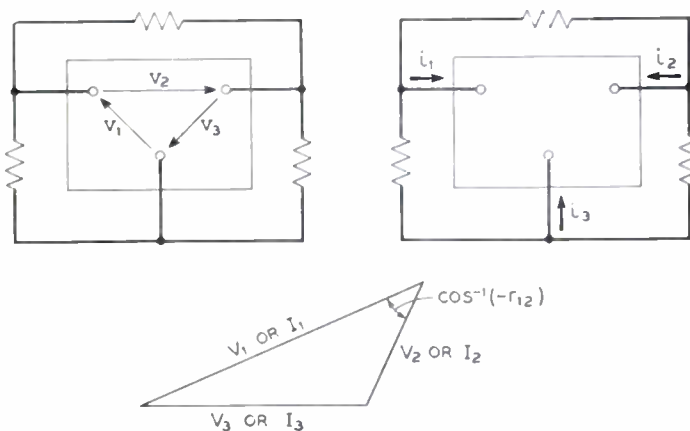
Fig. 1—Relations between terminal noise voltages or currents.

From this representation it can be seen that (a) if two voltages and one correlation coefficient are known the third voltage and the other two correlations can be calculated; (b) if one voltage and two correlations are known the other items can be calculated; (c) if all three voltages are known, all three correlations can be calculated; and (d) if one voltage is much smaller than the others, the others must be almost equal and almost completely (negatively) correlated. The last relation applies to transistors, where the open-circuit emitter-base noise voltage is usually less than one per cent of the collector-base voltage.

Simple Circuit Relations

Consider the case of a network of several meshes, containing independent noise sources I_1, $I_2 \cdots$ for which we know the power spectra P_1, $P_2 \cdots$. Since the sources are independent, the cross spectra P_{12}, $P_{23} \cdots$ are all zero. The resulting noises in two meshes, A and B, are given by

$$I_A = A_1 I_1 + A_2 I_2 + \cdots$$
$$I_B = B_1 I_1 + B_2 I_2 + \cdots,$$

where the A's and B's are complex transfer functions. We wish to find the power spectra of I_A and I_B and the cross spectrum between them.

For want of a general symbol, we have used I's in this example to represent either noise voltages or currents or some of each. Hence the A's may be impedances, admittances, or transfer ratios, as may be appropriate.

From (2) and (3) it is straightforward to show that

$$P_A = |A_1|^2 P_1 + |A_2|^2 P_2 + \cdots$$
$$P_B = |B_1|^2 P_1 + |B_2|^2 P_2 + \cdots$$
$$P_{AB} = A_1^* B_1 P_1 + A_2^* B_2 P_2 + \cdots$$
$$= |A_1 B_1| e^{i\theta_1} P_1 + |A_2 B_2| e^{i\theta_2} P_2 + \cdots,$$

where θ_j is the angle of B_j, minus the angle of A_j.

A special case of this relation may be used to establish the following theorem: If two noise currents (or voltages) are compounded partly of a common source, and partly from independent sources, and if α is the fraction of power in the first due to the common source and β the fraction of power in the second due to the common source, then the intrinsic correlation between the two currents is the geometric mean of α and β. This theorem gives some insight into the significance of correlation between noise currents. Let

$$I_A = A_1 I_1 + A_0 I_0$$
$$I_B = B_2 I_2 + B_0 I_0.$$

Then

$$P_A = |A_1|^2 P_1 + |A_0|^2 P_0$$
$$P_B = |B_2|^2 P_2 + |B_0|^2 P_0$$
$$P_{AB}' = A_0^* B_0 P_0/(P_A P_B)^{1/2} = (\alpha\beta)^{1/2} e^{i\gamma},$$

where

$$\alpha = |A_0|^2 P_0/P_A$$
$$\beta = |B_0|^2 P_0/P_B.$$

γ is the difference in angle of B_0 and A_0.

Since by (6) the magnitude of P_{AB}' is the intrinsic correlation between I_A and I_B, the theorem is established.

Another relation which will be used in Section 4 for interpreting the conversion formulas for equivalent-generator representations follows. Let

$$I_A = A_1 I_1 + A_2 I_2$$
$$I_B = B_1 I_1 + B_2 I_2,$$

where, in contrast to the earlier example, I_1 and I_2 are not necessarily independent. Then, by a similar procedure,

$$P_A = |A_1|^2 P_1 + |A_2|^2 P_2 + 2 \text{ real part } (A_1^* A_2 P_{12})$$
$$P_B = |B_1|^2 P_1 + |B_2|^2 P_2 + 2 \text{ real part } (B_1^* B_2 P_{12}) \quad (12)$$
$$P_{AB} = A_1^* B_1 P_1 + A_1^* B_2 P_{12} + A_2^* B_1 P_{21} + A_2^* B_2 P_2.$$

which gives all the spectra of I_A and I_B in terms of the spectra of the component sources I_1 and I_2 and the cross spectrum.

A second useful theorem may be established by considering a special case of the above relation. The intrinsic correlation between two noise currents (or voltages) is not changed by passing one or both currents through linear networks; the actual correlation is not changed by passing one or both currents through linear networks having real transfer functions. To show this,

let

$$I_A = A I_1$$
$$I_B = B I_2,$$

where I_1 and I_2 are not independent. Then from (12)

$$P_{AB}' = A^* B P_{12}/(P_A P_B)^{1/2} = |AB| P_{12} e^{i\gamma}/(P_A P_B)^{1/2}$$
$$= P_{12}' e^{i\gamma},$$

which shows that the magnitude of the two normalized cross spectra is the same, and the angle is changed by γ, the difference in angles of the two transfer functions. Hence the theorem is established.

Peterson's Equivalent Noise Generator Theorem

A theorem on the use of equivalent noise generators which is of great utility in circuit problems was apparently first stated by Peterson in an unpublished memorandum dated August, 1943, referred to briefly in a published article[6] and restated in an unpublished memorandum in 1949, from which we quote.

"It is well known that the signal performance of a linear active four-pole is completely determined by four parameters, which depend only on the internal structure of the four-pole. In order to describe the noise performance of the network two additional intrinsic parameters are needed, thus making six in all. These two additional parameters may take the form of two fictitious noise generators, one of which is placed in series with the input mesh and the other in series with the output mesh of the four-pole, which in this representation is itself entirely noiseless. This particular equivalent noise circuit is most suitable when used with the open-circuit signal parameter set. On the other hand, if the signal parameters are taken as the short-circuit admittance set, it is more convenient to select the noise parameters in the form of two noise currents impressed across the input and output terminals respectively."

It may be noted that the four transmission parameters are in general complex, so that eight real quantities are required to specify them. Likewise, the two noise parameters require two (real) power spectra and a (complex) cross spectrum to specify them, so that four more real quantities are involved, making a total of twelve.

Peterson's proof consisted of writing mesh equations for the noisy device and also for the quiet device and showing that there is a unique relation which insures equivalence. We shall give a somewhat different proof which has the advantage of making no assumptions regarding the internal structure of the networks (other than linearity of transmission properties) and of showing quite clearly what sort of statistical properties of the noise sources must be equated to make the networks equivalent.

Consider three linear networks N_1, N_2, N_3 which have identical transmission properties in both directions for

[6] L. C. Peterson, "Signal and noise in microwave tetrode," *Proc. I.R.E.*, vol. 35, pp. 1264–1272; November, 1947.

external signals. N_1 is the given network having internal sources of noise. N_2 is a noise-free network with a voltage generator in series with the input terminals and another in series with the output terminals, each generator producing a voltage instantaneously identical with the open-circuit voltage at the corresponding terminals of N_1. N_3 is a noise-free network with generators whose voltages may not be identical with, but have similar statistical properties to those of N_2. By a straightforward extension of Thevenin's theorem, networks N_1 and N_2 must supply identical currents and voltages to similar terminating impedances. Because of their identical transmission properties, N_2 and N_3 must supply currents and voltages with similar statistical properties to similar terminating impedances. Hence the same must be true for N_1 and N_3, which establishes the theorem. In particular, if the generators of N_3 have the same power and cross spectra as the open-circuit voltages of N_1, then the two networks will be equivalent in noise behavior for all linear circuit applications.

A similar theorem for short-circuit noise currents can be proved in an analogous manner.

The theorem is stated and proved for linear networks. It may often be usefully applied to small signals in nonlinear networks, a procedure which is well known and much used in transmission problems. In making such an application one must certainly satisfy the usual criteria for linear behavior with respect to externally applied small signals, but this is not sufficient. When the device contains internal sources of voltage or current, one must also be sure that the small-signal requirements are satisfied internally and that the external connections do not appreciably affect the generation mechanism.[7] Although no cases have been found experimentally where noise behavior of transistors was not correctly predicted by small-signal theory, this possibility should not be overlooked, particularly with small-bias values and circuit connections providing large amounts of feedback.

4. Noise Behavior of Transistors

Two different methods of describing the noise properties of transistors have been found useful. The first is in terms of equivalent noise generators, as described in the preceding section. This is the most compact way of giving a complete description of the noise properties. For linear circuit problems, a suitable description consists of the power spectra of two equivalent generators and a cross spectrum between them. These will be functions of two dc bias parameters, but independent of the external circuit impedances. The second method makes use of the noise figure, which is a simple and direct index of the noise behavior of the device as an amplifier. It depends on frequency and the dc bias parameters, and also on the generator impedance to which the device is connected, but is independent of the load impedance. The noise figure can be easily calculated from the equivalent-generator description. The

reverse process, while possible in principle, is not readily accomplished in practice.

Equivalent-Generator Representation

There are at least a dozen ways in which two voltage or current generators may be associated with two of the three pairs of terminals of a transistor to produce an equivalent noise network. These are equally general in their ability to represent the noise behavior. Thus it is clear that the ability of an equivalent network to completely represent the noise behavior of a transistor is no indication that it resembles the physical configuration responsible for the noise. It is possible to secure more elaborate representations by using generators at all three terminals, or by using both a voltage and a current generator at each pair of terminals. A choice between the various configurations depends on several considerations, such as (a) the simplicity of the representation with respect to the sort of spectra obtained, the correlation between the noise generators, and the variation of the noise parameters with dc bias values;
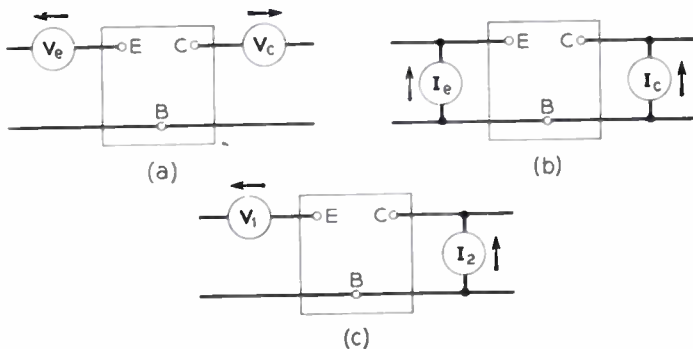
Fig. 2—Representations of noise behavior by equivalent-noise generators.

(b) the ease and accuracy with which the noise parameters can be measured; (c) convenience for calculating external-circuit behavior; and (d) degree of correspondence to the actual physical noise mechanism. Fortunately, it is not difficult to convert from one form of representation to another, and methods of doing this are described in the following.

Some representations used or seriously considered are shown in Fig. 2. The one which has been used most extensively to date is that of Fig. 2(a), which will be referred to as the $V-V$ representation, since it makes use of two low-impedance noise-voltage generators placed in series with the emitter and collector terminals. The generator V_e is equated to the noise voltage between emitter and base in the transistor, and V_c to that between collector and base, with all terminals open circuited.

The $I-I$ representation of Fig. 2(b) makes use of two high-impedance noise-current generators I_e and I_c equated to the emitter-base and the collector-base noise currents of the transistor, with both pairs of terminals short circuited.

A third representation, shown in Fig. 2(c), is a mixed system, designated $V-I$. It uses a voltage generator V_1

[7] These considerations were pointed out by W. Shockley.

in series with the emitter terminal and a current generator I_2 connected from collector to base. These are equated to the corresponding noise voltage and current in the transistor, with the emitter open circuited and the collector shorted to the base. It should be noted that in general V_e is not equal to V_1, nor is I_c equal to I_2, because the opposite-end terminating impedance is different in both cases.

Before writing the conversion formulas between the various representations it will be useful to summarize some of the methods of representing the transmission properties of a three-terminal network. In terms of open-circuit impedances

$$V_1 = Z_{11}I_1 + Z_{12}I_2,$$
$$V_2 = Z_{21}I_1 + Z_{22}I_2.$$

The Z's are simply related to the equivalent-T parameters often used to describe transistors.

$$Z_{11} = R_e + R_b, \qquad Z_{12} = R_b.$$
$$Z_{21} = R_m + R_b, \qquad Z_{22} = R_c + R_b.$$

In terms of short-circuit admittances,

$$I_1 = Y_{11}V_1 + Y_{12}V_2,$$
$$I_2 = Y_{21}V_1 + Y_{22}V_2.$$

A mixed system which is especially suited to the $V-I$ noise generator representation is

$$V_1 = H_{11}I_1 + H_{12}V,$$
$$I_2 = H_{21}I_1 + H_{22}V_2.$$

With the exception perhaps of the last, these systems of parameters are well known. Relations between them are given by Guillemin.[8]

The conversion formulas between the three equivalent noise-generator representations can be written down by inspection in terms of the transmission parameters. The relation between the $V-V$ and the $I-I$ systems is seen to be

$$V_e = Z_{11}I_e + Z_{12}I_c,$$
$$V_c = Z_{21}I_e + Z_{22}I_c. \tag{13}$$

The converse relation is

$$I_e = Y_{11}V_e + Y_{12}V_c,$$
$$I_c = Y_{21}V_e + Y_{22}V_c. \tag{14}$$

The relation between the $V-I$ and the $V-V$ systems is given both in terms of the Z's and the H's.

$$V_1 = V_e - (Z_{12}/Z_{22})V_c = V_e - H_{12}V_c,$$
$$I_2 = (1/Z_{22})V_c = H_{22}V_c. \tag{15}$$

The converse relation is

$$V_e = V_1 + Z_{12}I_2 = V_1 + (H_{12}/H_{22})I_2,$$
$$V_c = Z_{22}I_2 = (1/H_{22})I_2. \tag{16}$$

[8] E. A. Guillemin, "Communication Networks," vol. II, chap. IV, John Wiley and Sons, Inc., New York, N. Y.; 1935. Relations given by Guillemin must be modified somewhat to apply to active networks.

It will be seen that all of the conversion formulas are of the form of the circuit equations leading to the spectrum relations (12). When the power spectra and cross spectrum for one equivalent-generator representation is known, the corresponding spectra and cross spectrum for another representation can be calculated from relations (12), where the A's and B's are taken from the appropriate conversion formulas. When the transfer functions are real, the calculations are very simple. With complex transfer functions, the process is more involved, and some care is required to combine the angles in the proper sense.

Noise Figure—Grounded-Base Connection

In this paper we shall make use of the narrow-band noise figure F as an index of the noise behavior of a transistor when used as an amplifier. This may be defined as the ratio of the total noise power appearing in the load in a narrow band to that portion which is due to the amplified Johnson noise of R_G, the generator impedance to which the input is connected.[9] Its value depends on the noise properties of the transistor, the dc bias conditions, the frequency, and the generator impedance, but is independent of the load impedance.

The noise figure can be calculated from any of the equivalent-generator representations described in the preceding section by determining separately the contributions to the noise in the load from Johnson noise in R_G and from the two equivalent noise generators. Since noise figure is independent of load impedance, any convenient value of the latter may be selected. For example, consider the $V-V$ representation of Fig. 2(a), using Z parameters and terminating the input in Z_G and the output in open circuit. If the open-circuit noise voltages at the output are V_0 due to Johnson noise in R_G, V_A due to generator V_e, and V_B due to generator V_c, we find that

$$V_0 = (4kTR_G)^{1/2}Z_{21}/(Z_{11} + Z_G),$$
$$V_A = - V_e Z_{21}/(Z_{11} + Z_G),$$
$$V_B = V_c,$$

where kT is the Boltzmann constant times absolute temperature and Z_G is the generator impedance with real part R_G. The corresponding power and cross spectra are

$$P_0 = 4kTR_G \left| Z_{21}/(Z_{11} + Z_G) \right|^2,$$
$$P_A = P_e \left| Z_{21}/(Z_{11} + Z_G) \right|^2,$$
$$P_B = P_c,$$
$$P_{0A} = P_{0B} = 0,$$
$$P_{AB} = - P_{ec}[Z_{21}/(Z_{11} + Z_G)]^*.$$

From this, the noise figure is

$$F = (P_0 + P_A + P_B + 2 \times \text{real part } P_{AB})/P_0$$
$$= 1 + \frac{1}{4kTR_G}\left[P_e + P_c \left| \frac{Z_{11} + Z_G}{Z_{21}} \right|^2 \right]$$

[9] This is a narrow band form of the definition given by H. T. Friis, "Noise figures of radio receivers," Proc. I.R.E., vol. 32, pp. 419–422; July, 1944.

$$- 2 \times \text{real part} \left\{ P_{ec} \frac{Z_{11} + Z_G}{Z_{21}} \right\} \Bigg], \quad (17)$$

which is the expression given by Ryder and Kircher,[10] except for inclusion of the correlation term, and expression of noise in terms of unit bandwidth.

If the noise is expressed in volts-squared per cycle bandwidth, the constant $4kT$ is 1.64×10^{-20} at room temperature.

In terms of the I–I representation and the Y parameters, the noise figure is

$$F = 1 + \frac{1}{4kTG_G} \left[P_e + P_c \left| \frac{Y_{11} + Y_G}{Y_{21}} \right|^2 \right.$$
$$\left. - 2 \times \text{real part} \left\{ P_{ec} \frac{Y_{11} + Y_G}{Y_{21}} \right\} \right], \quad (18)$$

where Y_G is the admittance of the generator with real part G_G. It should be noted that the power spectra in this expression are not the same as those in (17), but are derived from I_e and I_c in an analogous way.

In terms of the V–I representation and the H parameters

$$F = 1 + \frac{1}{4kTR_G} \left[P_1 + P_2 \left| \frac{H_{11} + Z_G}{H_{21}} \right|^2 \right.$$
$$\left. + 2 \times \text{real part} \left\{ P_{12} \frac{H_{11} + Z_G}{H_{21}} \right\} \right], \quad (19)$$

where P_1, P_2, and P_{12} are the power spectra and cross spectrum of V_1 and I_2.

In all three formulas the last term is simplified if the frequency is low enough so that all the transmission parameters are real. In this case the cross spectrum is real and is equal to the geometric mean of the two preceding terms, multiplied by the correlation between the two noise generators.

The generator impedance R_G required for minimum noise figure can be determined from any of the expressions for noise figure. A particularly simple case occurs when, in terms of the V–V representation, the noise generator V_E is so small that its contribution to the load noise is negligible, and all the impedances are pure resistances. In this case, minimum noise figure is obtained when R_G equals Z_{11}. Experience has shown that this gives a fairly good approximation to the optimum R_G actually required by nearly all transistors over the usual required by most transistors at low frequencies in the usual range of dc bias values. This is usually a good deal higher than the generator impedance required for maximum power gain into practical working loads, so a compromise has to be made. Fortunately, the noise figure increases rather slowly when R_G departs from its optimum value as shown in the following table:[11]

R_G/optimum R_G	$\frac{1}{2}$ or 2	$\frac{1}{5}$ or 5	$\frac{1}{10}$ or 10
Increase in noise figure, db	0.5	2.6	4.8

[10] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, pp. 367–400; July, 1949.
[11] The table contains asymptotic values which are good approximations for noise figures greater than 10 db. For smaller noise figures the values are less than those shown in the table.

Noise Figure—Grounded Emitter or Collector

Noise-figure formulas were given by Ryder and Kircher[9] in terms of the V–V representation for all three grounding connections. Inspection of these formulas shows that for the impedance values as they exist in present transistors, the noise figures do not differ significantly among the various connections except for backward transmission in the grounded-collector case, where the noise figure may be quite large. The optimum value for R_G is substantially the same for grounded-emitter and for grounded-base operation. In the case of the grounded-emitter connection the optimum R_G for noise is usually a good deal smaller than the optimum R_G for maximum gain.

5. Noise Characteristics of Junction Transistors

The measurements now described are the result of study of about a dozen n-p-n type 1752 transistors taken at random from recent product.[12] Some of the noise properties were measured in only a few units, others in the whole group. The object was a general survey of noise properties, methods of measuring and representing them, and their significance in helping toward a physical understanding of the noise process. It is believed that the data presented are reasonably representative, but they are manifestly not extensive enough nor sufficiently systematic to serve as a complete specification of noise performance.

It is clear from the earlier part of this paper that a great variety of methods of representing noise behavior are available. Those used in this section were found quite serviceable, but no claim is made for having explored all the possibilities or foreseen all the requirements in various applications. It is hoped that the information here presented may serve as a useful start on the problem, and that the general methods described may be a helpful basis for future development.

Methods of Measurement

It does not seem worthwhile to give a detailed description of the measuring equipment, since high-gain low-level amplifier construction is a pretty well understood art. A few comments will suffice.

The measuring system consisted of six stages of wideband amplification, followed by a set of suitable filters and a vacuum-tube voltmeter. The input stage was a Western Electric 348A tube, pentode connected, for the range 20 to 15,000 cycles. A different preamplifier using a Western Electric 403B tube in a cascode connection[13] was used for the high-frequency range 1 kc to 1 mc. The amplifiers were operated from conventional power supplies using gas-tube regulated plate supplies and having a rectified heater supply for the three stages of the low-frequency preamplifier.

[12] Some early measurements are reported by R. L. Wallace and W. J. Pietenpol, "Some circuit properties of n-p-n transistors," *Bell Sys. Tech. Jour.*, vol. 30, pp. 530–563; July, 1951.
[13] R. Q. Twiss and Y. Beers, "Vacuum Tube Amplifiers," Rad. Lab. Series, McGraw Hill Book Co., Inc. New York, N. Y., chap. 13.10; 1948.

The filters were single-section constant-k structures having a ratio of upper to lower cutoff frequency of about 1.5. This has been found a convenient compromise between resolution and steadiness of output. The filter mid-frequencies are spaced at intervals of about an octave in most cases. For careful spectrum determinations at the lower frequencies, these filters were replaced by a General Radio Wave Analyzer, type 736A, whose chief advantage was its ability to get in between harmonics of 60 cycles when power-line pickup was troublesome.

The system was ordinarily used as a bridging amplifier, and was calibrated for gain and bandwidth with a low-level sinusoidal voltage. The effective bandwidth of the various filters was carefully determined by plotting the response in power units and integrating. For calibration it is found convenient to provide a sinusoidal signal in an impedance of 2 ohms which will give the same response on the measuring instrument as a noise voltage of 10^{-12} volts-squared per cycle bandwidth.

Bias was applied to the transistor under test through a terminating resistance of 20,000 ohms at the emitter and 5,000 ohms at the collector. Since the input impedance is usually less than 2,000 ohms, and the output impedance greater than 0.1 megohm, these terminations provided substantially an open circuit at the input and a short circuit at the output end. The noise-measuring amplifiers were bridged across the terminating resistances, thus measuring V_1 and $I_2 R_L$ directly.

Ideally, noise voltage squared should be determined by a square-law rectifier. In practice, it is much more convenient to use a vacuum-tube voltmeter such as the Ballantine or the Hewlett-Packard model 400C. These instruments are approximately linear full-wave rectifiers. Theory shows that such a rectifier calibrated to read the rms of a sine wave will read low by a factor of 0.886 in voltage (1.05 db) when used on Gaussian noise. Transistor noise is approximately Gaussian and experience shows that such a correction will give the true rms value to within a fraction of a db. Most measurements were made with the Hewlett-Packard instrument, with a 6,000 microfarad electrolytic condenser shunted across the meter to give a larger time constant, a correction of 1 db being added to the readings. The peak-reading type of vacuum-tube voltmeter is very undesirable for noise-power measurements, and the linear rectifier would have to be used with caution unless the Gaussian properties of the noise had been established.

The measuring amplifiers and filters are provided in duplicate as a means of determining correlation between two noise voltages. Correlation measurements were made by the following very convenient artifice, which avoids the use of a product-measuring device. The definition of the correlation between two noise voltages may be put in the following form:

$$r_{12} = \frac{\overline{v_1 v_2}}{[\overline{v_1^2 v_2^2}]^{1/2}} = \frac{\overline{(v_1 + v_2)^2} - \overline{(v_1 - v_2)^2}}{4[\overline{v_1^2 v_2^2}]}.$$

A simple switching arrangement between the outputs

of the two amplifier channels and the vacuum-tube voltmeter makes it possible to measure any of the following voltages, v_1, v_2, $v_1 + v_2$, $v_1 - v_2$. The squares of the readings differ from the mean square of the instantaneous voltages by the constant factor 0.886 mentioned previously which cancels out in the expression for r_{12}. As shown by the second theorem in Section 3, *Simple-Circuit Relations*, the correlation is independent of the relative gain of the two amplifier channels, and the best accuracy is attained when the gains are adjusted to make the meter readings equal for v_1 and v_2.

General Properties

Transistor noise is relatively steady in character, and aside from a characteristic difference in spectrum, is not
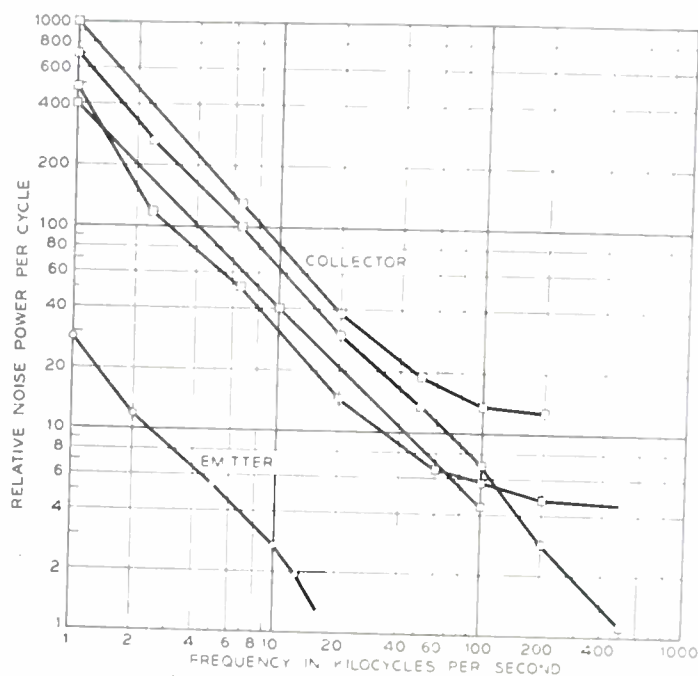
Fig. 3—Typical noise spectra for n-p-n transistors.

unlike Johnson noise in a resistor. Short-time fluctuations of a rectifier measuring device are somewhat greater than for Johnson noise, but of the same order. Drift is often observed for a few minutes after bias is applied, more often downward than upward, and amounting usually do not more than 2 or 3 db. Measurements at intervals over a period of several weeks generally agree to within 2 or 3 db. Application of large reverse bias to either the emitter or collector with the other terminal floating tends to raise the noise, sometimes substantially (10–15 db). This appears to be a temporary effect which disappears after minutes or hours. A small minority of units display bursts of noise of a very irregular character. Units which have been damaged by excessive biases tend to behave in this way. The properties of noise (except the magnitude) are so similar in junction and point-contact transistors as to suggest strongly that the basic noise mechanism is the same in both devices.

Spectra

At frequencies from 20 cycles up to around 50 kc the spectrum of the noise at emitter or collector terminals for a variety of bias conditions seems universally to be of the form

$$P(f) = K/f^n,$$

where the exponent is a little greater than unity, usually about 1.2. At higher frequencies, the spectrum of collector noise sometimes shows a rise of 5 or 10 db from the extrapolation of the low-frequency law; in other cases it follows the above law pretty closely to 500 kc, which is as high as our measurements have been carried. Several examples are shown in Fig. 3. It has not been possible to carry measurements of the emitter noise to the higher frequencies because the levels are so low. It is not clear whether the anomalies at high frequency are a part of the noise mechanism or are caused by changes in the transmission parameters, which begin to be important in this frequency range.
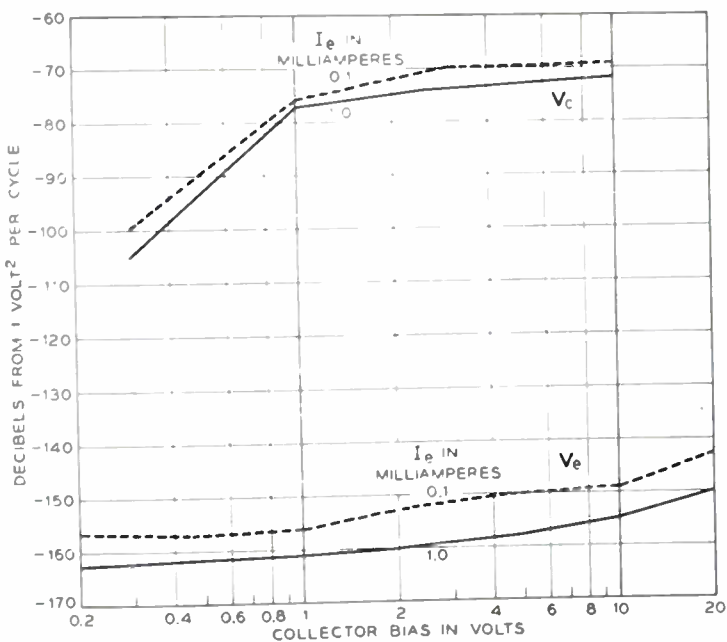
Fig. 4—Equivalent-noise generator voltages at 1 kc in the *V-V* system, for an *n-p-n* transistor. These values were calculated from the measurements in Fig. 5.

Since the spectra at frequencies below 50 kc are so similar in form, it is usually sufficient to give a value at a single frequency to represent data over this frequency range; 1 kc is often used as a reference frequency.

Equivalent-Generator Representation

It has been customary to represent the noise behavior of point-contact transistors by equivalent generators in the *V-V* system of Fig. 2(a). A good deal of information is available on the variation of these generator voltages with dc bias. A few measurements have been made of the I_2 generator in the *V-I* system of Fig. 2(c). In most cases it has been found that the I_2 generator behaves in a less simple way as a function of dc bias than the generators of the *V-V* system. Since the open-circuit noise and transmission properties are easily measured in the case of point-contact transistors, the

V-V representation seems to be generally satisfactory.

In the case of junction transistors, the situation seems to be different. Because of the high-collector impedance, it is not easy to make measurements directly in the *V-V* system. However, at low frequencies it is quite easy to
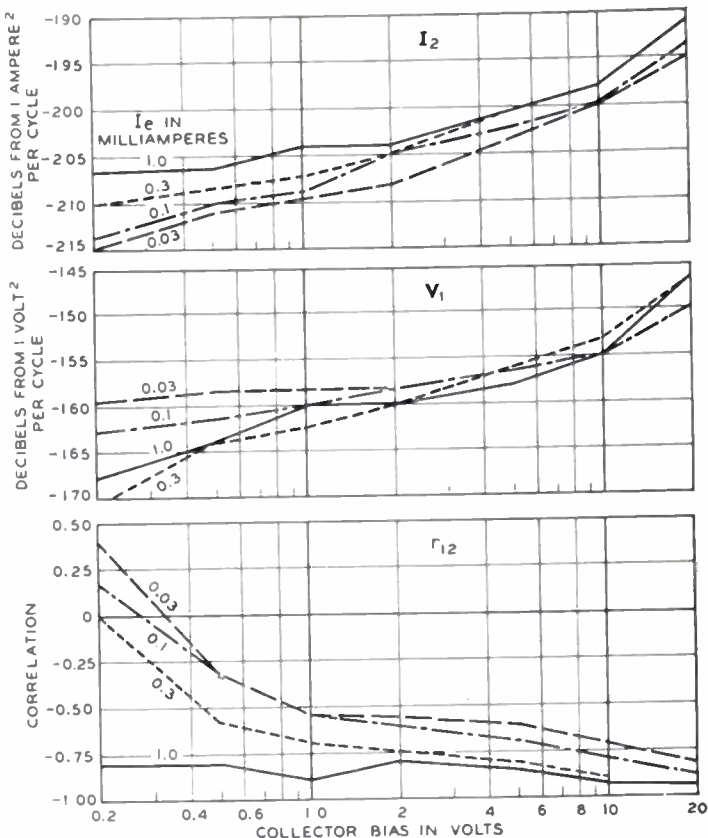
Fig. 5—Equivalent-noise generator voltage, current and correlation at 1 kc in the *V-I* system, for an *n-p-n* transistor.

calculate values from data in the *V-I* system, and this has been done in Fig. 4, which shows the equivalent-generator values in the *V-V* system for the same data which is shown in Fig. 5 for the *V-I* system. A few
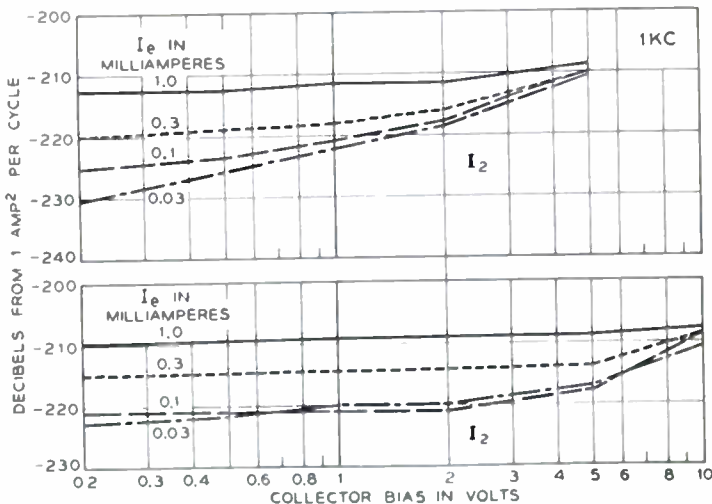
Fig. 6—Equivalent-noise generator current I_2 at 1 kc in the *V-I* system for two other *n-p-n* transistors.

points on the V_e curves of Fig. 4 have been checked by direct measurement by means of a comparison voltage introduced in series with the collector circuit. Fig. 6 shows partial data for several other transistors in the

V–I system, to give some idea of the variability among different units. For junction transistors, it appears that the V–I system represents the noise behavior in a somewhat simpler manner as a function of bias than does the V–V system. Although the theory of noise production is in a rather rudimentary state, there is some reason to believe that the noise should be proportional to the bias current rather than the voltage, and the V–I representation fits in with this picture better than the V–V representation. On the whole, it seems that the V–I system of equivalent-noise generators has several advantages in the case of junction transistors.

Noise Figure

It was pointed out above that, with certain simplifying assumptions, the noise figure should vary rather slowly with the generator impedance R_G to which a transistor is connected and should have its minimum when $R_G = Z_{11}$. Since the simplifying assumptions are not always met in practice, Fig. 7 gives the extreme
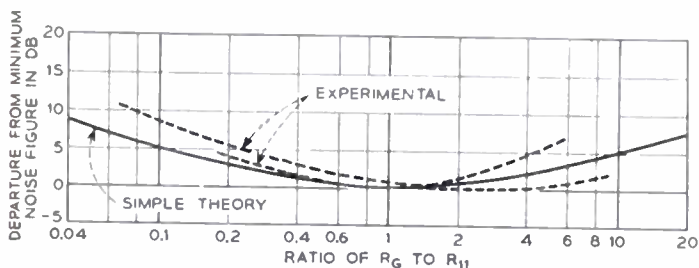
Fig. 7—Variation of noise figure with generator impedance R_G. Solid curve shows simplified theory. Dashed curves are extreme limits of experimental data for several *n-p-n* transistors under various bias conditions and at 1 kc and 200 kc.

range of experimental data on several junction-type transistors under various bias conditions and at frequencies of 1 and 200 kc. It is clear that the behavior is in reasonably good agreement with that calculated for the simplified case.
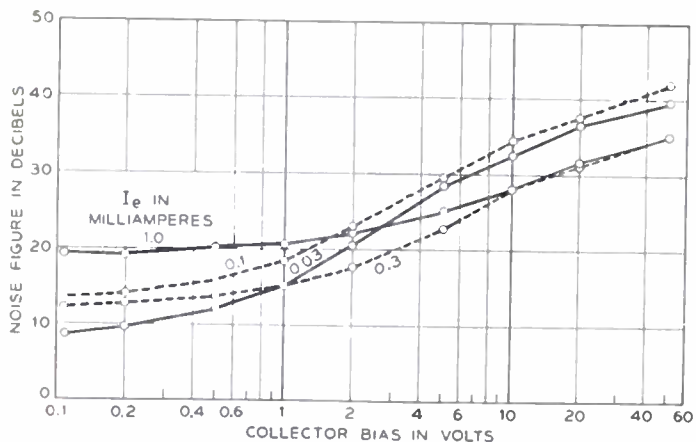
Fig. 8—Noise figure of an *n-p-n* transistor at 1 kc for various bias conditions.

In Fig. 8 the noise figure at 1 kc for a transistor with the grounded-base connection is plotted as a function of the bias parameters I_e and V_c, with R_G adjusted for minimum-noise figure at each point. Fig. 9 shows similar data for another transistor, and includes one curve for the frequency 200 kc. While not representing extreme cases, these two sets of data are representative of

the variation among units. Similar information for 13 transistors at three selected bias conditions and at 1 and 200 kc is given in Table I. The average difference
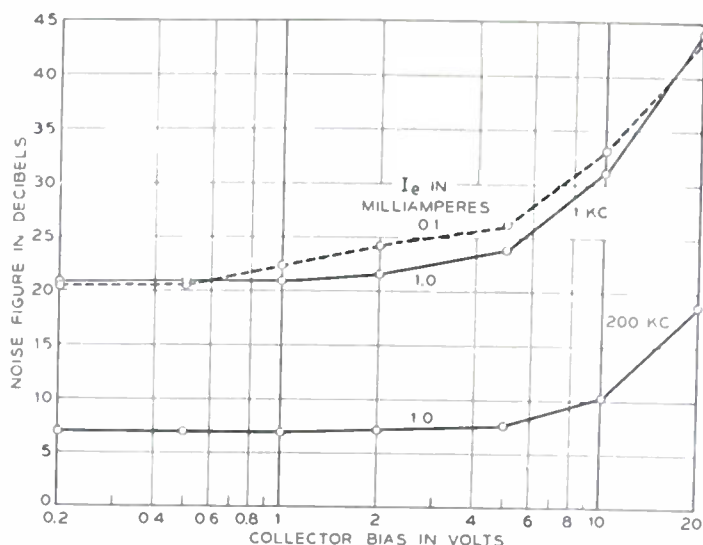
Fig. 9—Noise figure of another *n-p-n* transistor at 1 and 200 kc for various bias conditions.

in noise figure at the two frequencies is 14 db, whereas we would expect a difference of 27.5 db from a simple extrapolation of the low-frequency behavior according to the spectral law. The discrepancy is partly due to the anomalies in the spectrum shown in Fig. 3 and partly to reductions in gain at the higher frequency. It will be noted that operation at low bias values usually makes an appreciable improvement in the noise figure. The optimum value for I_e is probably between the two values used in the table, probably round 0.1 ma for an average unit.

TABLE I

Noise Figures of *n-p-n* Transistors at 1 and 200 kc

Noise figure in db for three selected bias conditions. Generator impedance picked approximately to give lowest noise figure; for $I_E = 1.0$ ma, $R_G = 500$ ohms; for $I_E = 0.030$ ma, $R_G = 1,000$ ohms.

	Frequency					
	1 kc			200 kc		
I_e V_e	1.0 4.5	1.0 0.5	0.03 0.5	1.0 4.5	1.0 0.5	0.03 0.5
1	20.1	18.1	19.1	7.2	5.5	5.0
2	20.2	19.7	24.1	5.8	5.2	5.4
3	19.1	16.9	21.3	5.5	5.5	8.0
4	23.3	19.9	17.0	5.2	4.2	3.9
5	22.3	22.3	14.0	5.9	6.7	5.4
6	25.6	24.7	13.0	6.6	6.5	8.7
7	21.7	22.2	11.6	6.2	6.7	4.4
8	24.4	21.6	21.1	7.0	6.2	5.1
9	20.8	21.8	13.5			
10	45.7	24.7	23.3	22.3	8.1	8.1
11	26.9	17.2	26.8	10.3	5.8	10.3
12	22.6	20.3	16.2	6.5	6.8	10.1
13	27.4	16.2	20.0	13.6	6.2	11.8
Average	24.6	20.4	18.5	8.5	6.1	7.2

Conclusions

The systematic method of dealing with linear circuit problems in noise set forth early in this paper is quite general, and should prove useful in many types of

problems. The discussion of applications of the method to the noise behavior of transistors makes it apparent that many forms of description are possible, among which a choice may be made based on convenience. We have indicated certain choices which have suited our applications, and have endeavored to present the methods in sufficient generality to enable the reader to make similar choices. The specific information on noise behavior of n-p-n transistors, while based on a small number of units, is believed to be reasonably indicative of the behavior of the present product, and should serve as at least a rough guide in circuit-design work.

Nothing has been said regarding the mechanism of noise production in transistors. This is a subject which is not at all completely understood at present, and discussion of it was felt to be beyond the scope of this paper. The interested reader is referred to a forthcoming paper.[14]

Acknowledgment

To many of my associates I am indebted for helpful discussion of matters presented in this paper, and for provision of the transistors whose noise behavior is reported.

Appendix

The function $S(f)$, defined formally by (1), has detailed properties which depend on the nature of the noise function $y(t)$. Generally speaking, both the magnitude and angle of $S(f)$ fluctuate along the frequency axis in such a way that there is little correlation in the values over frequency intervals of the order of $1/T$. To obtain a well-behaved function it is desirable to smooth this function over frequency intervals much larger than $1/T$, limited of course by the desired frequency resolution in the spectrum. We shall give a definition of the smoothed functions required for (2) and (3) which depends on a Fourier series expansion of the noise current over a finite time interval. With this approach it is not hard to show that the definition is in accord with the quantities measured by the procedures outlined previously.

Let $y_1(t)$ and $y_2(t)$ be two noise currents for which power and cross spectra are to be defined. Let y_1' and y_2' be the result of cyclic repetition of that portion of y_1 and y_2 lying in a time interval T. If T is much longer than the reciprocal of the bandwidth of the measuring filters, transients due to discontinuities at the boundaries of the time intervals will have a negligible effect on the average measurements. If the noise currents are stationary, the response of the measuring system to y_1' and y_2' will not differ significantly from the response to y_1 and y_2. The primed currents are described in complete detail by their Fourier series expansions

$$y_1' = \sum a_k \cos (2\pi kt/T - \theta_k),$$
$$y_2' = \sum b_k \cos (2\pi kt/T - \phi_k).$$

[14] H. C. Montgomery, "Electrical noise in semiconductors," *Bell Sys. Tech. Jour.*, vol. 31, pp. 950–975; September, 1952.

If the magnitude of $S_1(f)$ is obtained by rms smoothing of the a_k's over the frequency interval of the measuring filter (multiplied by $T^{-1/2}$, which is the number of components per unit frequency interval), it is evident that the power spectrum defined by (2) is just the power contained in those Fourier components passed by the filter.

$$(1/2T) \sum a_k{}^2 \quad \text{or} \quad (1/2T) \sum_k{}^2.$$

The cross spectrum defined by (3) is a smoothed version of

$$(1/2T) a_k b_k e^{i(\phi_k - \theta_k)}.$$

The measuring process described for obtaining the cross spectrum involves filtering y_1 and y_2, taking the product, and dividing by the bandwidth. This is equivalent to

$$(1/\delta f) \sum a_k \cos (2\pi kt/T - \theta_k) \times \sum b_k \cos (2\pi kt/T - \phi_k),$$

where each sum is carried over those terms passed by the filters. By a well-known property of trigonometric series, products of terms of unlike frequency average to zero, so the above is equivalent to

$$(1/\delta f) \sum a_k b_k \cos (2\pi kt/T - \theta_k) \cos (2\pi kt/T - \phi_k)$$
$$= (1/2\delta f) \sum a_k b_k [\cos (4\pi kt/T - \theta_k - \phi_k) + \cos (\phi_k - \theta_k)]$$
$$= (1/2\delta f) \sum a_k b_k \cos (\phi_k - \theta_k),$$

since the first term in the bracket averages to zero. The summation involves $\delta f/T$ terms, so the average value is

$$(1/2T) \sum a_k b_k \cos (\phi_k - \theta_k).$$

This is the real part of the cross spectrum, as determined experimentally, and this relation may be taken to define the smoothing process required in (3). The imaginary part of the cross spectrum is determined by a quadrature measurement, and is evidently

$$(1/2T) \sum a_k b_k \sin (\phi_k - \theta_k).$$

For the special case of two identical noise currents $y_1' = y_2'$, it is seen that $a_k = b_k$ and $\theta_k = \phi_k$; hence the cross spectrum is real and equal to the geometric mean of the two power spectra. If y_1' and y_2' are derived from a common source through linear networks whose transfer functions vary slowly over the interval δf, we have approximately $a_k = \text{const.} \times b_k$ and $\theta_k - \phi_k = \text{const.}$ In this instance the cross spectrum is complex, but its magnitude is again the geometric mean of the power spectra. Either of these cases represents completely coherent noise currents. In the case of completely incoherent or independent noise currents the angles of the individual terms are completely random, and the average is negligibly small compared to the power spectra, and no fixed phase shifts in the system can increase it. It may be noted that apparent incoherence (with respect to the measuring system as described) can be produced by subjecting one noise current to phase shifts which vary substantially over the frequency interval δf. However, this effect can be removed by a complementary phase shift.

Variation of Transistor Parameters with Temperature*

ABRAHAM COBLENZ† AND HARRY L. OWENS†, MEMBER, IRE

Summary—In this paper results of a study to determine the temperature dependence of the electrical characteristics of Type 1698 and 1768 transistors are given. Variations of transistor parameters as a function of temperature are given for each type of unit. Cumulative distribution curves are also given for the transistors studied. The implications of the results obtained with regard to transistor operation at elevated temperatures are discussed briefly.

INTRODUCTION

OF THE MORE than 60 experimental types of transistors known, only two, Western Electric Types 1698 and 1768, have been made available to the military services in appreciable quantities. The extensive use of these types in experimental circuit studies and exploratory development models of equipment has been handicapped by the lack of published data regarding temperature effects. To provide these data, parameter measurements have been made on 20 Type 1698 and 30 Type 1768 transistors over the temperature range 25° to 85°C. The current amplification factor (alpha) has been measured as a function of frequency on an additional group of 50 Type 1698 transistors. Subject to the limitations noted, it is believed that results which have statistical value have been obtained.

The reader is cautioned that 85°C operation of these transistors is far in excess of the maximum temperature ratings recommended by the manufacturer. In the investigations reported herein destructive testing was considered desirable in order to obtain information regarding the absolute limits of operation for these type transistors.

In considering the significance of the results obtained, it is quite apparent that all engineers would not agree as to the temperature limitations of transistor operation implied by these data. Perhaps the difficulties in formulating specific recommendations are responsible for the present lack of information on the subject. In this paper these difficulties are reviewed and tentative conclusions regarding higher-temperature operation of transistors are discussed.

METHODS OF TEST

All results reported herein were obtained with laboratory-constructed test equipment of conventional design. Although standards regarding methods of testing transistors have not been adopted, it is believed that the voltmeter-ammeter techniques employed are quite similar to those in general use. A schematic diagram of the equipment used to measure the equivalent circuit parameters is shown in Fig. 1. Direct-current bias for the transistor under test was provided by two high-voltage power supplies. Large series resistances intro-

duced in both the emitter and the collector leads made the bias sources essentially constant current and assisted in eliminating unwanted oscillations in transistors not unconditionally stable.

The beat frequency-type signal generator provided a constant current at two thousand cycles which was applied to the emitter or the collector as appropriate for the particular parameter being measured. In no case did the magnitude of this alternating current exceed 15 μa. As this current represents approximately one-thousandth of the maximum ordinate of the respective voltage-current characteristics, it was considered that this value of current might properly be considered to fall within the small-signal class. Due to the low-signal current levels used, some shielding was found necessary to avoid stray pickup.
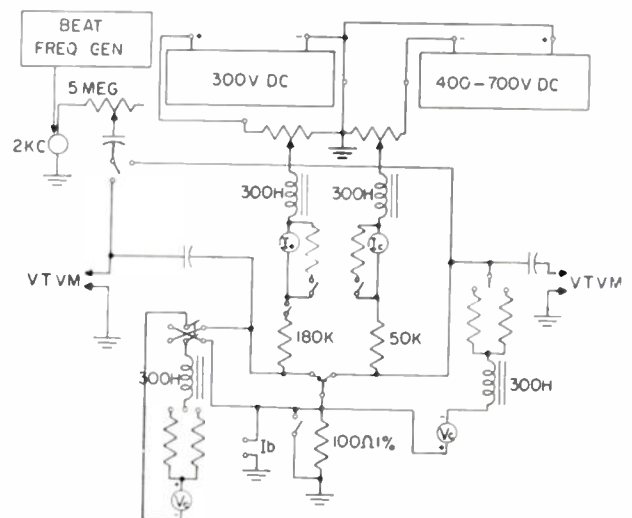
Fig. 1—Schematic diagram for parameter measurements.

In taking the parameter measurements, one first establishes the desired direct-current bias and positions the alternating current signal to the emitter or collector as appropriate for the parameter under test. The signal produces a voltage drop proportional to the resistance of the parameter under test. The vacuum-tube voltmeters for these measurements are shown on either side of the schematic in Fig. 1. From the known value of signal current and its voltage drop across the unknown parameter, one may readily calculate the latter. An obvious expedient is to calibrate the voltmeter directly in ohms. The following transformation equations were used to transform from the equivalent 4-pole parameters actually measured to the equivalent active T-network parameters.

Alpha measurements were made as a function of frequency, with the circuit arrangement shown in the schematic diagram of Fig. 2. Analysis of the circuits will reveal precautions taken to maintain adequate

TABLE I

Generator Positioned To	Voltmeter Positioned To	Parameter Measured
Collector	Emitter	$r_{12} = r_b$
Emitter	Emitter	$r_{11} = r_e + r_b$
Emitter	Collector	$r_{21} = r_m + r_b$
Collector	Collector	$r_{22} = r_c + r_b$

bandwidth in the test set itself. In addition, experimental evidence was obtained indicating that the response of the system was well beyond that of the transistors investigated.
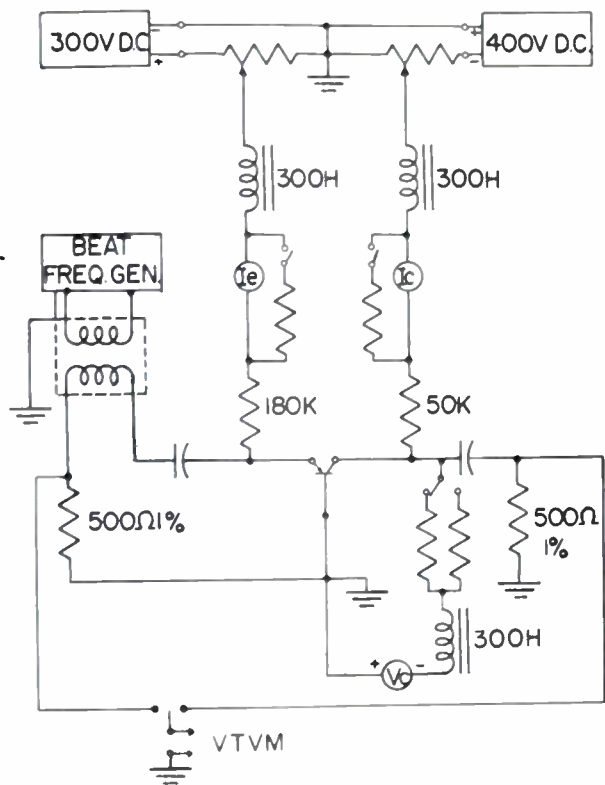
Fig. 2—Schematic diagram for frequency-response tests.

Experimental Results

The Type 1698 is a point-contact transistor triode in cartridge form. It is designed for use in switching circuits where the large-signal parameters of the active device are of primary interest. The maximum-rated ambient temperature for this type of transistor is 55°C. The manufacturer recommends the operating precaution of connecting the transistor to its circuit under zero-bias conditions. Where this is not possible, the base connection should always be made not later than second. As previous investigations in which these precautions were violated resulted in permanent damage to some transistors, the precautions were strictly observed during the tests reported herein.

The average variation of the equivalent circuit parameters as a function of temperature for 20 Type 1698 transistors is shown in Fig. 3. Every effort was made to avoid oscillations, to observe the maximum-rated collector dissipation of 120 mw, and to measure the parameters under the normal test conditions of bias:

$V_c = -30$ volts and $I_e = 1$ ma. The abscissa of the curves showing temperature variations are intended to indicate a temperature cycle beginning at room temperature, rising to 85°C, and then decreasing to room temperature, from left to right. Although the units were main-
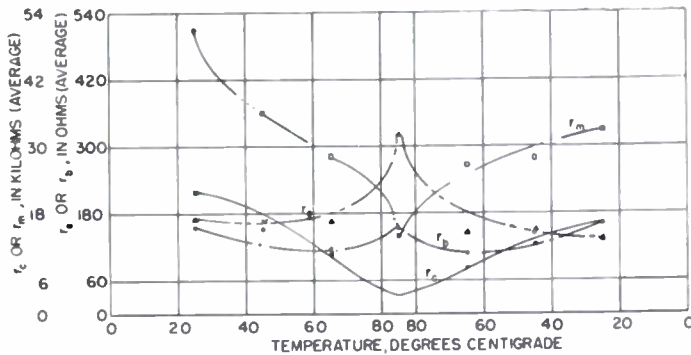
Fig. 3—Variation of equivalent, active-T 4-pole parameters with temperature, type 1698.

tained at the elevated temperatures only for the time necessary to achieve thermal equilibrium and make the electrical measurements, it is quite evident from the appreciable permanent changes encountered both for r_m and r_c that the units were subjected to excessive temperatures. As further evidence of permanent damage, some of the transistors began to oscillate at the elevated temperatures and were not stable at the operating point defined above, even after being returned to room temperature. Such units were sometimes found capable of stable operation at reduced bias. Additional investigations are required to adequately determine optimum bias conditions for a given range of required operating temperatures.
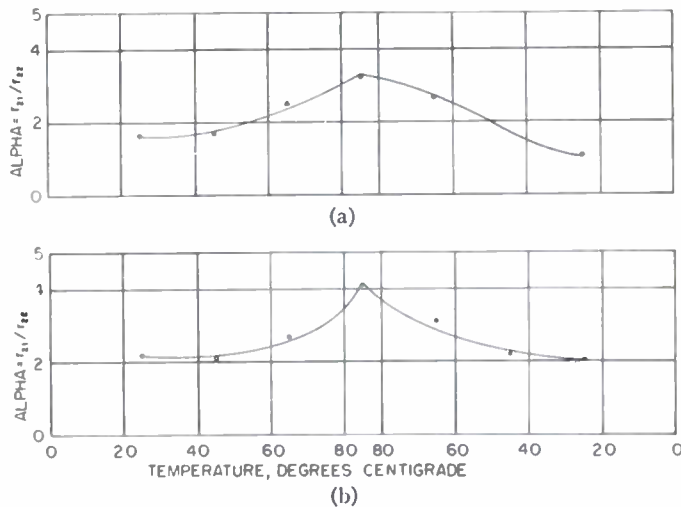
Fig. 4—Variation of current gain, alpha, with temperature. (a) Upper curve, Type 1768. (b) Lower curve, Type 1698.

Variation of alpha as a function of temperature for the Type 1698 units is shown in the lower curve of Fig. 4. Values for alpha indicated were calculated from data presented in Fig. 3, using the following equation:

$$\text{alpha} = \frac{r_{21}}{r_{22}} = \frac{r_m + r_b}{r_c + r_b}. \tag{1}$$

For the same group of 1698 transistors, the cumulative distribution curves of Fig. 5 show unit-to-unit variations at room temperature for three parameters of particular interest. For a chosen value of the ordinate, the curves show per cent of units with a parameter value equal to or less than the corresponding abscissa reading. It is an integral of the usual distribution curve.
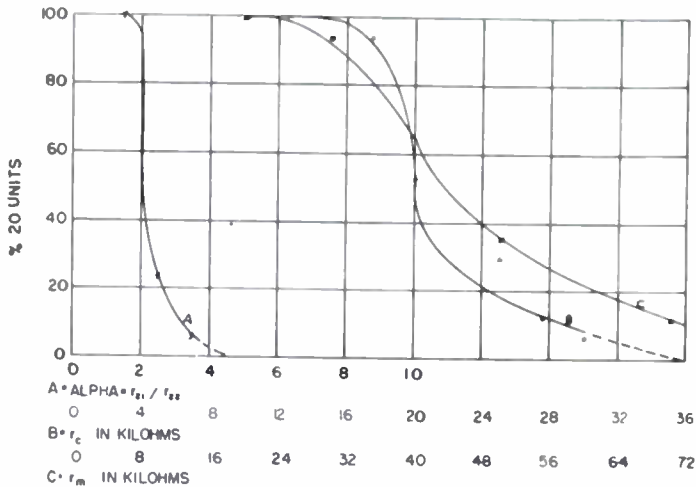
Fig. 5—Cumulative distribution curves of alpha, r_c and r_m for Type 1698 at 25°C. (a) Curve A, alpha. (b) Curve B, r_c. (c) Curve C, r_m.

In switching applications, it is necessary to insure adequate collector resistance when the transistor is in the cutoff or high resistance condition. Type 1698 units should have 20,000 ohms or greater resistance at room temperature as determined by measurement of I_{c0}, the
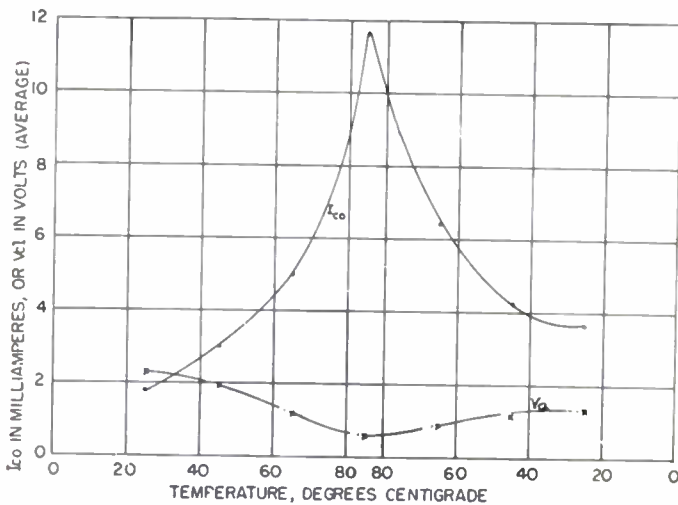
Fig. 6—Variation of I_{c0} and V_{c1} with temperature, Type 1698.

collector current under conditions of zero-emitter current and 40-volts collector potential. Temperature dependence of I_{c0} for the 1698 transistors tested is given in Fig. 6. Another parameter also shown in Fig. 6 is the V_{c1} or collector voltage for the transistor when in the low resistance or saturation condition. V_{c1} is measured with 1-ma emitter current and 2-ma collector current.

The Type 1768 transistor was developed for use in low-voltage, low-power drain, high-efficiency oscillator and amplifier applications. These properties have been

emphasized by the optimization of the characteristic of previous point-contact transistor types. Additiona information regarding the Type 1768 units is reportec elsewhere in this issue.[1] The transistor-design objectives were established to satisfy the requirements of a particular application. Performance data on production models are considered highly satisfactory. By optimizing parameters to satisfy the development objectives, the field of application for units of this type has been narrowed due to impairment of high-temperature and
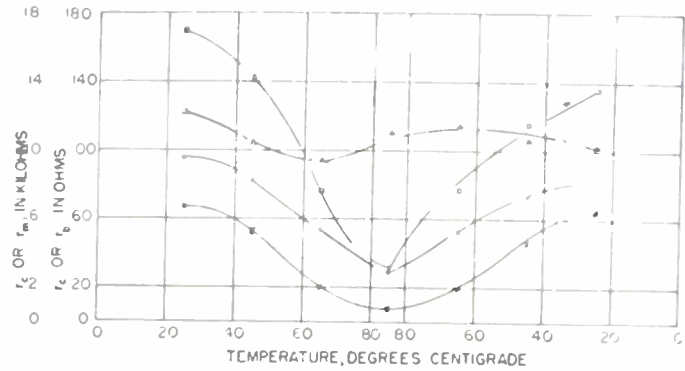
Fig. 7—Variation of equivalent, active-T, 4-pole parameters with temperature, Type 1768.

high-frequency operation. The maximum-rated ambient temperature for this type of transistor is 40°C and its maximum-rated collector dissipation is 120 mw.

The average variation of the equivalent circuit parameters as a function of temperature for 30 Type 1768 transistors is shown in Fig. 7. Alpha as a function
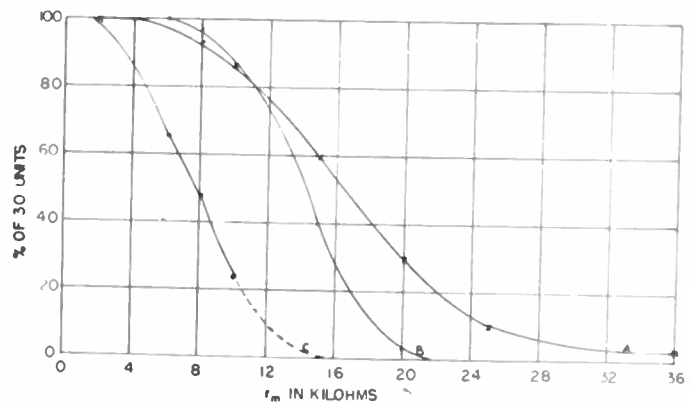
Fig. 8—Cumulative distribution curves of r_m, Type 1768. (a) Curve A, 25°C. (b) Curve B, 45°C. (c) Curve C, 65°C.

of temperature for these units is shown in the upper curve of Fig. 4. Cumulative distribution curves at three operating temperatures for r_m, r_c, and alpha are shown, respectively, in Figs. 8, 9, and 10.

Alpha has been measured as a function of frequency on both types of transistors previously discussed, utilizing the equipment shown in Fig. 2. Results obtained on 20 Type 1768 units are given in the upper curve of Fig. 11.

[1] D. E. Thomas, "Low-drain transistor audio oscillator," vol. 40, pp. 1385–1395.

It may be readily shown that the condition for maximum current gain requires the transistors' output load to be small compared to r_{22}. The 500-ohm load resistor used in the test equipment does not fully satisfy this condition for 1768 Type units. Consequently, the
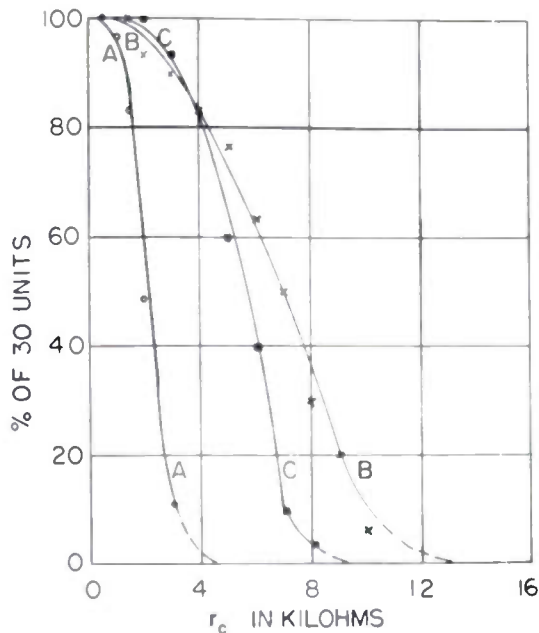
Fig. 9—Cumulative distribution curves of r_c, Type 1768. (a) Curve A, 65°C. (b) Curve B, 45°C. (c) Curve C, 25°C.
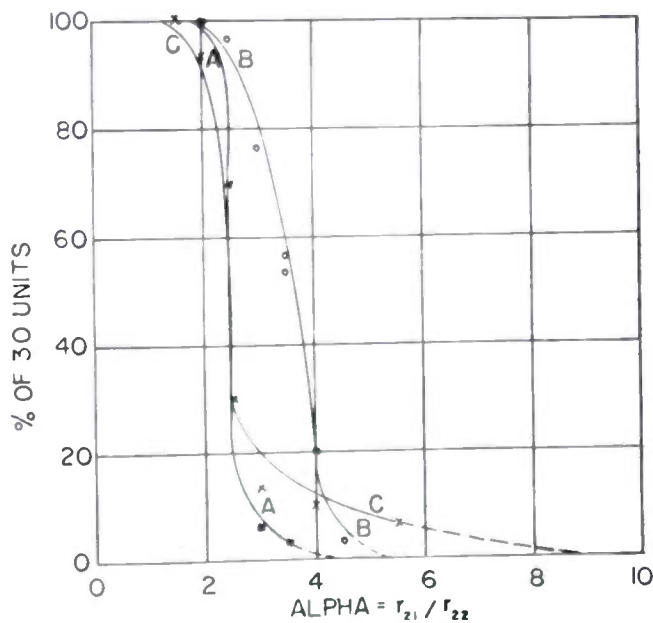
Fig. 10—Cumulative distribution curves of alpha, Type 1768. (a) Curve A, 25°C. (b) Curve B, 45°C. (c) Curve C, 65°C.

maximum current gain may be up to 10 per cent higher than reported. In the lower curve of Fig. 11 are shown the results obtained on 50 Type 1698 transistors. Frequency cutoff is usually defined as the frequency at which alpha is 3 db below its low-frequency value. The dashed lines shown in Fig. 11 establish the average frequency cutoff of Type 1768 and 1698 transistors as 450 kc and 2.5 mc, respectively.
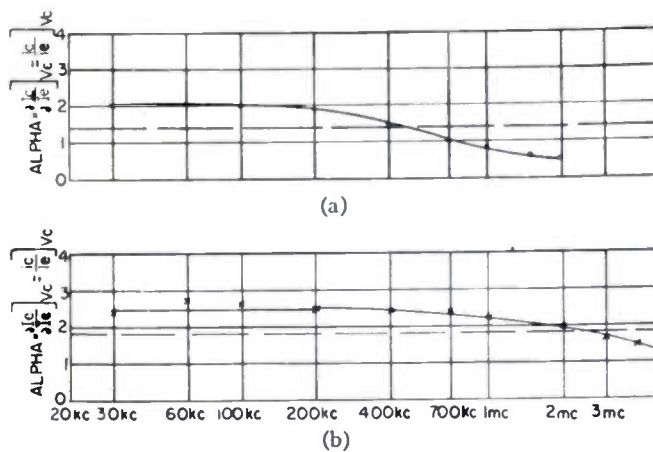
Fig. 11—Variation of alpha with frequency. (a) Upper curve, Type 1768. (b) Lower curve, Type 1698.

CONCLUSIONS

It is quite apparent from the results of these and other investigations which have been conducted that the data now available are insufficient to justify anything more than tentative conclusions regarding temperature limits for transistors. Transistor circuit requirements and the properties of the devices themselves require further study before temperature limits can be defined with confidence. Additional data are being obtained on both operational- and storage-temperature limitations for transistors.

In spite of very substantial changes in the small signal characteristics at the elevated temperatures, it is interesting to note that the transistors are capable of appreciable gain even at 85°C. Calculated gain data are given in Table II, based on linear class A operation of average 1698 units in the grounded-base connection. The presence of positive feedback due to r_b is neglected in the calculation of maximum available gain. Operating power gain is given for $R_G = 300$ ohms and $R_L = 15,000$ ohms.

TABLE II

Temperature	25°C	85°C
Maximum Available Gain $= \dfrac{r_m{}^2}{4r_c r_e}$	22.4 db	16.2 db
Operating Power Gain $= 4R_G R_L \left(\dfrac{-(r_b + r_m)}{\Delta} \right)^2$	21.5 db	14.3 db

$$\Delta = (R_G + r_e + r_b)(R_L + r_e + r_b) - r_b(r_b + r_m)$$

Although useful gain would be delivered by both 1689 and 1768 units at 85°C, reliable operation is not now feasible due to other problems, such as thermal and electrical instability, which result in permanent as well as temporary impairment of characteristics.

From general practical considerations it has been concluded that Type 1698 and Type 1768 units are satisfactory up to approximately 60°C for many small signal applications. Gain, for example, is decreased by only about 2 db within this temperature range. Rela-

tively constant gain is achieved because the decrease of r_c with temperature is largely compensated by the increase in alpha. Approximate per cent variations of these two parameters are given in Table III for both types of transistors investigated.

TABLE III
PER CENT VARIATION OF TRANSISTOR PARAMETERS

Temperature °C	Type 1698		Type 1768	
	α	r_e	α	r_e
25°C				
30		− 2.3		− 1.5
40		−11.3	1.9	−11.8
50	5	−27.3	11.5	−33.8
60	11.4	−43.2	23.0	−58.8
70	27.3	−63.7	38.5	−78.0
80	59.0	−79.5	56.0	−88.3
85	86.5	−84.0	61.7	−88.3

The temperature dependence of the electrical characteristics of Type 1698 transistors impose serious limitations on their application to switching circuits required to operate reliably at high-ambient temperatures. Many of the direct-coupled switching circuits which have been developed will not operate satisfactorily with Type 1698 transistors beyond the temperature range of 40° to 50°C. This range may be increased somewhat if the transistor's direct-current operating point is stabilized by suitable auxiliary circuit features. The use of alternating-current coupling, where permissible, should also prove advantageous in extending the useful temperature range. Even though direct coupling is avoided, the inescapable fact is that the temperature dependence of I_{e0} as shown in Fig. 6 will result in sufficient self-engendered heat to induce thermal instability as the maximum dissipation ratings are approached at ele-

vated temperatures.

The 1689 transistor now under development performs functions similar to the 1698 unit, and has improved temperature characteristics.[2,3] The average temperature coefficient for r_c of about −1 per cent per degree reported for the 1689 transistor compares with about −5 per cent per degree for 1698 units as determined from Fig. 6. Reliable operation of 1689 transistors at 70°C is understood to be achieved in most switching applications investigated. For many straight transmission applications superior performance would be predicted for the Type 1729 transistor on the basis of reported higher cutoff frequency and improved temperature characteristics.[2,3] The development of the 1689 and the 1729 units clearly demonstrates the advances being made through an improved understanding of design considerations in transistors and better control of materials and production processes.

Frequency Variations of Current-Amplification Factor for Junction Transistors*

R. L. PRITCHARD†, ASSOCIATE MEMBER, IRE

Summary—In a grounded-emitter or grounded-collector connection of a junction transistor, the effective current amplification is proportional to the factor $(1-\alpha)^{-1}$, where α is the current-amplification factor of the transistor. As a result of the phase shift associated with α, the magnitude of $(1-\alpha)^{-1}$ decreases rapidly with increasing frequency. It is shown that the magnitude of $(1-\alpha)^{-1}$ is 3 db below its low-frequency value at a frequency of the order of magnitude of $(1-\alpha_0)$ times the frequency at which the magnitude of α alone has decreased 3 db below its low-frequency value of α_0. In addition, experimental results of measurements of phase and amplitude characteristics of α as a function of frequency for several fused-impurity p-n-p junction transistors are presented and are compared with the variation of α with frequency calculated in accordance with the theory of Shockley, et al.

INTRODUCTION

WHEN the current-amplification factor α of a transistor is close to unity, the electric-circuit properties of the transistor resemble fairly closely those of the vacuum tube.[1] This is true of the junction transistor for which practical values of α range from approximately 0.85 or 0.90 to 0.99+. Thus, there are three possible methods of connection of the transistor, namely, grounded-base, grounded-emitter, and grounded-collector that correspond fairly close to the grounded-grid, grounded-cathode, and grounded-plate

[1] This similarity was pointed out for point-contact transistors, which may have a value of α between zero and three or more; R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, no. 3, pp. 374–376; July, 1949.

connections of the vacuum tube.

In two of these methods of connection, grounded-emitter and grounded-collector, the current gain of the stage is proportional to the factor $1/(1-\alpha)$, which may be quite large in magnitude. For cascaded amplifier stages, the voltage gain per stage is proportional to $1/(1-\alpha)$, while the power gain per stage is proportional to the square of this factor.

At high frequencies the magnitude of α decreases, and consequently the magnitude of $1/(1-\alpha)$ may decrease quite rapidly.[2] Of more importance, however, is the fact that the phase shift associated with α may cause a substantial decrease in the magnitude of $1/(1-\alpha)$ at frequencies for which the magnitude of α has not decreased appreciably from its low-frequency value.

The fact that there is a phase shift associated with α is not a new discovery. For the point-contact transistor this was mentioned by Ryder and Kircher,[3] and results of measurements of the phase and amplitude of α have been presented by Becker and Shive.[4] On the other hand, for the junction transistor, except for a brief section in the paper by Shockley, *et al*,[5] describing the theory of the junction transistor, little has been published concerning the frequency variation of α. Consequently, in the investigation of grounded-emitter or grounded-collector transistor amplifier circuits, there is the possibility that phase effects due to α may be neglected.[6]

The purpose of this paper is to point out in a quantitative manner that the phase shift associated with α may limit the frequency response of a junction transistor in the grounded-emitter or grounded-collector connections to a considerably lower frequency than has been expected previously. (Only the grounded-emitter connection will be considered explicitly; the case of the grounded-collector connection is quite similar.) In addition, results of measurements of the frequency variation of α are presented for several fused-impurity *p-n-p* junction transistors developed by Saby.[7]

EFFECT OF FREQUENCY VARIATION OF ALPHA UPON GROUNDED-EMITTER CURRENT GAIN

The current gain K_e for the grounded-emitter connection of the transistor is

$$K_e \equiv z_{21}{}^{(e)}/z_{22}{}^{(e)},\tag{1}$$

where $z_{ij}{}^{(e)}$ is the $i-j$ impedance parameter for the grounded-emitter connection. These two parameters are related to the parameters $z_{ij}{}^{(b)}$ for the grounded-base connection as follows:[8]

$$z_{21}{}^{(e)} = z_{11}{}^{(b)} - \alpha z_{22}{}^{(b)};$$
$$z_{22}{}^{(e)} = z_{11}{}^{(b)} - z_{12}{}^{(b)} + z_{22}{}^{(b)} \cdot (1 - \alpha),$$

where $\alpha \equiv z_{21}{}^{(b)}/z_{22}{}^{(b)}$ is the complex current-amplification factor. For most junction transistors, and for the frequency range of interest, $z_{11}{}^{(b)}$ and $z_{12}{}^{(b)}$ are considerably smaller in magnitude than $z_{22}{}^{(b)}$. Consequently, K_e may be written in the form

$$K_e = -\alpha/(1-\alpha).\tag{1a}$$

In general the current-amplification factor α at a circular frequency ω may be written in complex form as

$$\alpha \equiv A(\omega) \cdot \epsilon^{j\phi(\omega)} \equiv \alpha_r + j\alpha_i.\tag{2}$$

Substitution of (2) in (1a) yields

$$K_e = \frac{-A\epsilon^{j\phi}}{(1 - A\cos\phi) - jA\sin\phi} = \frac{-(\alpha_r + j\alpha_i)}{(1 - \alpha_r) - j\alpha_i}.\tag{3}$$

At very low frequencies, $\alpha = \alpha_0$, for example, $A = \alpha_0$, $\phi = 0$, and the current gain is

$$K_{e0} = -\alpha_0/(1-\alpha_0).\tag{4}$$

As the frequency is increased, the amplitude $A(\omega)$ decreases, rather slowly at first, while the phase angle ϕ decreases, linearly with frequency at first. It is convenient to define a cutoff frequency ω_c as the frequency at which $|\alpha| = A$ is 3 db below its low-frequency value α_0. Then for frequencies substantially less than ω_c, the behavior of α may be represented in terms of dimensionless parameters by an approximate expression of the form

$$[\alpha(\omega)/\alpha_0] \approx \{1 - b^2(\omega/\omega_c)^2\}\epsilon^{-ja(\omega/\omega_c)},$$
$$(\omega/\omega_c)^2 \ll 1,\tag{5}$$

where a and b are numerical constants of the order of magnitude of unity.[9]

[2] R. L. Wallace, Jr. and W. J. Pietenpol, "Some circuit properties and applications of *n-p-n* transistors," *Bell Sys. Tech. Jour.*, vol. 30, no. 3, p. 560; July, 1951. Also, PROC. I.R.E., vol. 39, p. 766; July, 1951.

[3] R. M. Ryder and R. J. Kircher, *op. cit.*, pp. 381–382.

[4] J. A. Becker and J. N. Shive, "The transistor—a new semiconductor amplifier," *Elec. Eng.*, vol. 68, p. 220, fig. 12; March, 1949.

[5] W. Shockley, M. Sparks, and G. K. Teal, "*p-n* junction transistors," *Phys. Rev.*, vol. 83, no. 1, section IX, p. 161; July 1, 1951.

[6] R. L. Wallace, Jr. and W. J. Pietenpol, *op. cit.*, point out the effect of the decrease in the magnitude of α, but they do not mention the effect of the phase shift.

[7] J. S. Saby, "Recent developments in transistors and related devices," *Tele-Tech.*, vol. 10, p. 34; December, 1951. Also, J. S. Saby, "Fused-impurity *p-n-p* junction transistors," PROC. I.R.E. pp. 1358–1360 of this issue.

[8] At low frequencies, $z_{21}{}^{(e)}$ and $z_{22}{}^{(e)}$ are usually written as $(r_e - r_m)$ and $(r_e + r_c - r_m)$, respectively. See R. L. Wallace, Jr. and W. J. Pietenpol, *op. cit.*, p. 547. Also, PROC. I.R.E., vol. 39, p. 760; July, 1951.

[9] Measurements on several transistors of the phase variation of α with frequency, which is essentially linear for $(\omega/\omega_c) < 0.3$, indicate values of a of 1.0 to 1.35. Similarly, measurements of the amplitude variation of α indicate values of b^2 of 0.35 to 0.65.

The theoretical values of a and b for an idealized one-dimensional transistor depend to a slight extent upon the low-frequency value α_0. If it is assumed that the frequency-dependent factor comprising α (the fraction β of the injected carriers that reaches the collector) is essentially unity at low frequencies, then $a = 1.22$ and $b^2 = 0.49$, as is shown in Appendix I.

It might be noted briefly that the variation of α at low frequencies given by (5) resembles the frequency variation of the response of a simple low-pass RC network. The response of such a network is described by the equation

$$1/[1 + j(\omega/\omega_c)], \text{ with } \omega_c = 1/RC,$$

which, at sufficiently low frequencies, may be approximated by an expression of the form of the right-hand member of (5) with $b^2 = 1/2$ and $a = 1$.

After substituting (5) in (1a), expanding the sine and cosine in power series, and neglecting terms of higher order than $(\omega/\omega_c)^2$, the expression for the magnitude of K_e may be written in the form

$$|K_e| = \frac{\alpha_0/(1-\alpha_0)}{\left|1 + \frac{(\omega/\omega_c)^2}{(1-\alpha_0)}\left[b^2 + \frac{\alpha_0 a^2}{2}\right] + j\,\frac{a\alpha_0}{(1-\alpha_0)}(\omega/\omega_c)\right|} . \quad (6)$$

The right-hand member of (6) will be 3 db below its low-frequency value $|K_{e0}|$ of (4) at a frequency ω_0 given by

$$\left(\frac{\omega_0}{\omega_c}\right) \doteq \left\{\frac{a^2\alpha_0^2}{(1-\alpha_0)^2} + \frac{(2b^2 + \alpha_0 a^2)}{(1-\alpha_0)}\right\}^{-1/2}, \quad (7)$$

subject to the restriction $(\omega_0/\omega_c)^2 \ll 1$. Since $\alpha_0/(1-\alpha_0)$ is assumed to be fairly large, the second term within the braces of (7) will be small compared to the first term. Hence, the right-hand member of (7) may be expanded in a series:

$$\left(\frac{\omega_0}{\omega_c}\right) \doteq \frac{(1-\alpha_0)}{a\alpha_0}\left[1 - \frac{(1-\alpha_0)}{2\alpha_0}\left(\frac{2b^2}{\alpha_0 a^2}+1\right)\cdots\right],$$
$$(\omega_0/\omega_c)^2 \ll 1. \quad (7a)$$

This result shows that the frequency at which the magnitude of the grounded-emitter current gain is down 3 db from its low-frequency value is of the order of $(1-\alpha_0)$ times the frequency at which the magnitude of the current-amplification factor is 3 db below its low-frequency value! Hence, when α_0 is very close to unity, the grounded-emitter amplifier may cut off at a fairly low frequency, even though the α-cutoff frequency may be relatively high. For example, in the illustration used by Wallace and Pietenpol,[10] $\alpha_0 = 0.9785$, $(\omega_c/2\pi) \sim 5$ mc, the grounded-emitter cutoff frequency is $\sim(0.0215) \times 5 \times 10^6 = 108$ kc. In the case of the present fused-impurity p-n-p junction transistors, with $(\omega_c/2\pi) \sim 50 - 200$ kc, when $\alpha_0 \geq 0.9$, grounded-emitter cutoff may occur in the audio-frequency range.

Since the dominant term in (7) does not contain the amplitude constant b, it is obvious that the 3-db decrease in $|K_e|$ relative to $|K_{e0}|$ that occurs at the frequency ω_0 is caused predominantly by phase shift. Thus, if it were assumed that the amplitude of α decreased with no accompanying phase shift, then in (5) a would be set equal to zero, and the cutoff frequency according to (7) would be

$$(\omega_0/\omega_c) \doteq (1/b)\sqrt{(1-\alpha_0)/2}.$$

When α_0 is close to unity, this frequency would be considerably higher than that given by (7a) when phase shift is *not* neglected, for example, when $a \approx 1$ rather than $= 0$.

The variation of the magnitude of K_e over a wider range of frequency has been calculated from curves

[10] R. L. Wallace, Jr. and W. J. Pietenpol, *op. cit.*, p. 758. Also, *Bell Sys. Tech. Jour.*, vol. 30, p. 543; July, 1951.

based on the experimental values for $|\alpha|$ and ϕ as functions of frequency to be described below.[11] Results of such calculations are given in Fig. 1, which shows the magnitude of K_e in db as a function of frequency relative to the α-cutoff frequency for $\alpha_0 = 0.90$, 0.95, 0.98, and for the limiting case of $\alpha_0 = 1$. The K_e-cutoff frequency ω_0 in each case (except for $\alpha_0 = 1$) is indicated by a circle. The rapid decrease in $|K_e|$ which occurs at low frequencies and the decrease of the K_e-cutoff frequency as α_0 is increased toward unity are quite apparent. It might be noted that the limiting curve is useful for observing graphically what the grounded-emitter cutoff frequency will be for a given low-frequency current gain. The K_e-cutoff frequency is approximately equal to the abscissa value at the intersection of the limiting curve with the ordinate value corresponding to the low-frequency current gain.
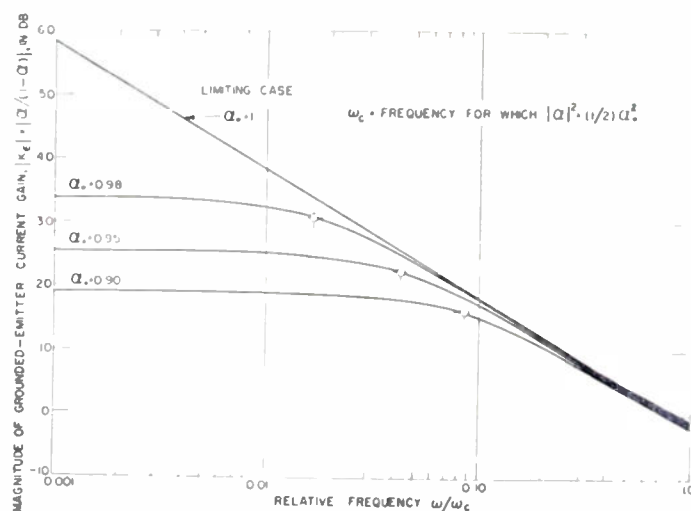
Fig. 1—Approximate behavior of the magnitude of grounded-emitter current gain as a function of frequency relative to α-cutoff frequency.

MEASUREMENT OF PHASE AND AMPLITUDE CHARACTERISTICS OF CURRENT-AMPLIFICATION FACTOR

Measurements of the complex current-amplification factor α were made over a frequency range of 500 cps to 200 kc for several of the early fused-impurity p-n-p junction transistors developed by Saby.[7] AC collector and emitter currents were determined by measuring the voltage drop across resistances in the collector and emitter circuits, respectively.

Amplitude variation of α was measured directly with a vacuum-tube voltmeter, while the phase was determined by either of two methods. For some measurements, relative phase was determined directly by observing Lissajous figures on the screen of a cathode-ray

[11] For computing $|K_e|$ as a function of frequency, the values selected for $|\alpha|/\alpha_0$ were taken arbitrarily to lie on a fictitious curve passing between the two curves drawn through the experimental points shown in Fig. 2. Similarly, the values of ϕ were taken from a fictitious curve passing amidst the three curves of Fig. 3. At low frequencies, values of $a = 1.2$, $b^2 = 0.4$ were assumed. Hence, the curves for $|K_e|$ versus frequency do not apply exactly to any given transistor but are indicative of the behavior to be expected.

oscilloscope. Precision of the results based on such measurements is estimated to be approximately $\pm 5°$. Alternatively, the phase angle of α was calculated from the values of the amplitude only of α and of the current gain K_e as follows: From (2),

$$\alpha_i{}^2 = A^2 - \alpha_r{}^2. \tag{8}$$

This may be substituted in the expression for the square of the magnitude of K_e obtained from (3) to yield

$$|K_e|^2 = A^2/[(1 - \alpha_r)^2 + A^2 - \alpha_r{}^2]. $$

Solving for α_r,

$$\alpha_r = \tfrac{1}{2}[(1 + A^2) - A^2/|K_e|^2]. \tag{9}$$

Hence, the phase ϕ of α may be calculated from the equation

$$\phi = \tan^{-1}(\alpha_i/\alpha_r), \tag{10}$$

with α_i and α_r obtained from (8) and (9), respectively, or from

$$\phi = \cos^{-1}(\alpha_r/A). \tag{10a}$$

This method of determining ϕ yields satisfactory results for frequencies up to approximately ω_e. Beyond this frequency, $|K_e|$ is less than unity, (Fig. 1), and it becomes difficult to calculate α_r accurately from moderately accurate measurements of A and $|K_e|$.

For several transistors, the phase and amplitude both of α and of K_e were measured, and calculations were made to check the consistency of the results. In general, the consistency among the results of the four measurements at any one frequency was good.
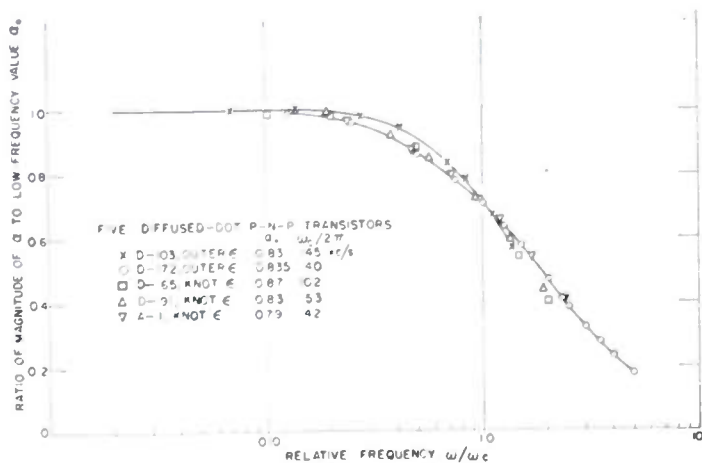
Fig. 2—Variation of magnitude of current-amplification factor α with relative frequency.

Typical results of the measurements are shown in terms of dimensionless variables in Figs. 2 and 3. In Fig. 2, the ratio of the amplitude of α at the frequency ω to the low-frequency value α_0 is shown as a function of the dimensionless frequency ratio ω/ω_e (where ω_e as defined above is the frequency at which $|\alpha| = 0.707\alpha_0$). Although the experimental points do not lie on a single universal curve, the divergence of the results from a median curve is not especially great, at least over the

frequency range considered. The two curves that have been drawn in Fig. 2 represent fairly well the extremes of values obtained from measurements of approximately ten transistors. In Fig. 3, the phase angle (lag) of α is shown as a function of ω/ω_e. Here also there is no universal curve, and the results shown are typical for the transistors measured.

Most of the measurements were made at a dc operating point of collector voltage $E_c = -4.5v$, emitter current $I_e = 1.0$ ma. Measurements on one transistor (D-103) at various operating points indicated that the relative variation of α with frequency is essentially independent of dc operating point, although both α_0 and ω_e may depend partly on choice of operating point.

DISCUSSION OF RESULTS

The observed variation of α with frequency is in fair agreement with the theoretical variation for the idealized one-dimensional transistor discussed in the Appendix. However, the construction of the Saby-type transistors is sufficiently different from the theoretical model that extremely good agreement should not be expected.[12] A comparison of the experimental values shown in Figs. 2 and 3 with the theoretical curves shown in Fig. 4 in the Appendix for $\beta_0 = 1$ indicates the following: The curve of $\phi(\omega)$ drawn through the crosses (x) in Fig. 3 is almost exactly coincident with the theoretical curve for $\phi(\omega)$. Thus, the curve drawn through the circles ($\bigcirc$) in Fig. 3 lies approximately 15 degrees below the theoretical curve in the region of $\omega/\omega_e \approx 3$. On the
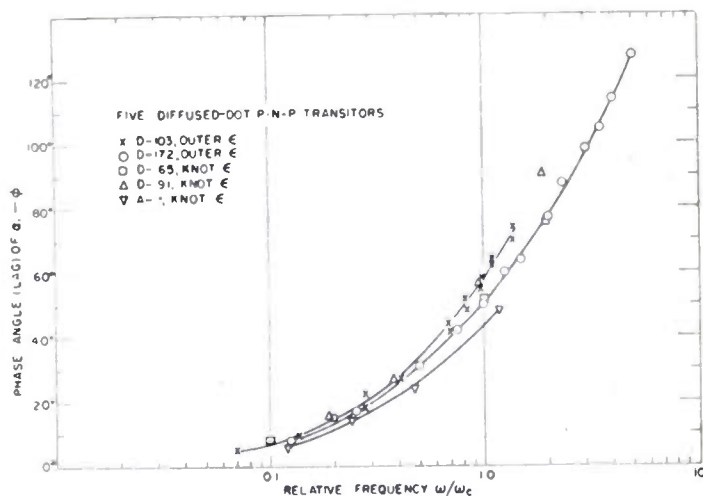
Fig. 3—Variation of phase angle (lag) of current-amplification factor α with relative frequency.

other hand, the theoretical curve of $A(\omega)$ passes between the two experimental curves shown in Fig. 2. From the number of transistors that was measured, it was not possible to establish any definite correlation between the nature of the observed variation of $|\alpha|$ or ϕ for a given transistor and its value of α_0 and/or ω_e.

[12] On the other hand, recent theoretical work by E. L. Steele (pp. 1424–1429 of this issue) indicates that the behavior of α versus frequency for the Saby transistors can be described by essentially the same expression given by Shockley, *et al.* for the one-dimensional model.

CONCLUSION

It has been shown that a decrease in the magnitude of the effective current gain in grounded-emitter and grounded-collector connections of a junction transistor may occur at a fairly low frequency because of the phase shift associated with the current-amplification factor α. The frequency at which the effective current gain is 3 db below its low-frequency value has been calculated and has been found to be of the order of $(1 - \alpha_0)$ times the cutoff frequency of α (where α_0 is the low-frequency value of α). In addition, results of measurements of the frequency variation of α for several fused-impurity p-n-p junction transistors have been presented.
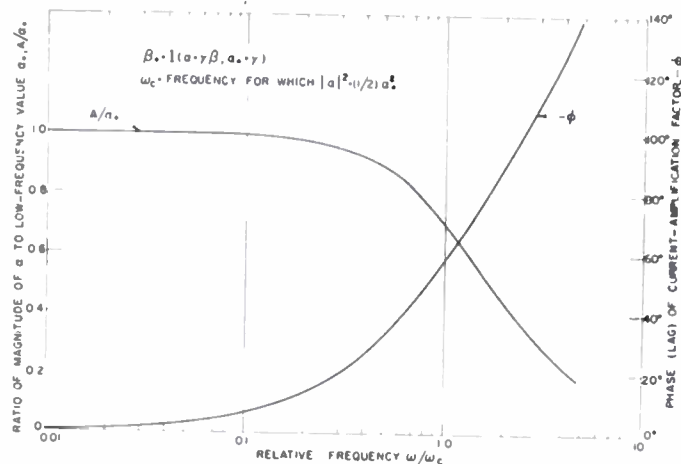
Fig. 4—Theoretical behavior of current-amplification factor $\alpha = A\,e^{j\phi}$ as a function of relative frequency.

It should be noted that the product of the low-frequency grounded-emitter current gain and the grounded-emitter cutoff frequency is essentially constant and equal to the α-cutoff frequency, independent of α_0. For a flat frequency response in the grounded-emitter or grounded-collector connections, a value of α_0 very close to unity may not be desirable. In conclusion, the writer wishes to point out that with given transistors it is possible to extend the grounded-emitter response somewhat beyond the cutoff frequency as calculated above, by means of negative feedback.

APPENDIX

Theoretical Derivation of the Frequency Variation of the Current-Amplification Factor

For a transistor with a simple p-n junction collector, the current-amplification factor α may be written as the product of two factors:[13]

$$\alpha = \gamma\beta, \qquad (I-1)$$

where γ is the fraction of the current at the emitter junction produced by emitter voltage that is carried by minority carriers in the base, and β is the fraction of these injected carriers that reaches the collector. The fraction γ is assumed to be independent of frequency,

[13] W. Shockley, *et al.*, *op. cit.*, p. 156.

but the frequency variation of β may be described by the expression[6]

$$\beta = \text{sech}\left\{(1 + j\omega\tau_m)^{1/2}(w/L_m)\right\}, \qquad (I-2)$$

where τ_m is the lifetime of the minority carriers ($m = n$ for electrons, as in the n-p-n transistor; $m = p$ for holes, as in the p-n-p transistor), w is the thickness of the center, or base, region (the p region for the n-p-n transistor), and L_m is the diffusion distance for the minority carriers in the base region.

At sufficiently low frequencies that $(\omega\tau_m)^2 \ll 1$,

$$\beta = \beta_0 \equiv \text{sech}(w/L_m).$$

For a good transistor,

$$(w/L_m)^2 \ll 1, \text{ so that}$$
$$\beta_0 \doteq 1 - \tfrac{1}{2}(w/L_m)^2 \approx 1.$$

Under the assumption that $\beta_0 = 1$, the frequency variation of $\beta = \alpha/\alpha_0$ from (I-2) may be expressed by the relation

$$\alpha/\alpha_0 = \beta \equiv \text{sech}\,(j\omega\tau_D)^{1/2}, \qquad (I-3)$$

where

$$\tau_D \equiv (\tau_m w^2/L_m^2) = (w^2/D_m)$$

is the diffusion time of the minority carriers through the base region. ($D_m = L_m^2/\tau_m$ is the diffusion constant.) However, for the purpose of this discussion, τ_D may be considered simply as an inherent time constant of the transistor.

At a frequency ω_c such that $\omega_c\tau_D = 2.43$, $|\beta|^2 = 1/2$ and $|\alpha|$ will be down 3 db below its low-frequency value $\alpha_0 = \gamma$. For frequencies small compared to $\omega_c = 2.43/\tau_D$, the hyperbolic secant in (I-3) may be approximated by the first few terms of its power series expansion.

$$\alpha/\alpha_0 \doteq 1 - j\tfrac{1}{2}(2.43\omega/\omega_c) - \frac{5}{4!}(2.43\omega/\omega_c)^2,$$
$$(\omega/\omega_c)^2 \ll 1. \qquad (I-4)$$

The amplitude of (α/α_0) is approximately

$$|\alpha/\alpha_0| \doteq \left\{1 + (2.43\omega/\omega_c)^2\left[\frac{1}{2^2} - \frac{2\cdot5}{4!}\right]\right\}^{1/2}$$

or

$$|\alpha/\alpha_0| \approx 1 - \frac{1}{12}(2.43\omega/\omega_c)^2, \qquad (I-5)$$

whereas the phase ϕ of (α/α_0) is approximately

$$\phi \doteq -\tfrac{1}{2}(2.43\omega/\omega_c). \qquad (I-6)$$

Comparison of (I-5) and (I-6) with (5) shows that for $(\omega/\omega_c)^2 \ll 1$,

$$b^2 = (2.43)^2/12 = 0.49,$$
$$a = (2.43)/2 = 1.22.$$

The theoretical variation of $|\alpha|/\alpha_0$ and of ϕ as a function of relative frequency ω/ω_c over a wider range of frequencies has been calculated according to (I-3), and is shown by the curves in Fig. 4.[14]

In conclusion it might be noted briefly that the theoretical cutoff frequency ω_c calculated in terms of the thickness w of the base region is

$$\frac{\omega_c}{2\pi} = \frac{2.43}{2\pi\tau_D} = \frac{2.43 D_m}{2\pi\omega^2}. \qquad \text{(I-7)}$$

If w is expressed in mils (10^{-3} inch), then (using $D_p = 44 \times 10^{-4} m^2/\text{sec}$, $D_n = 94 \times 10^{-4} m^2/\text{sec}$)

$$\frac{\omega_c}{2\pi} = \begin{cases} (5.6 \times 10^6/w^2)\text{cps} & \text{for an } n\text{-}p\text{-}n \text{ transistor,} \\ (2.6 \times 10^6/w^2)\text{cps} & \text{for a } p\text{-}n\text{-}p \text{ transistor.} \end{cases}$$

Hence, a cutoff frequency of 100 kc for a p-n-p transistor corresponds to a base thickness of approximately 5 mils, which is the order of magnitude achieved in the fabrication process.

---

[14] Similar calculations have been made for the case of $\beta_0 = 0.91$. When the results are expressed in terms of ω/ω_c, it is found that they differ very little from the results shown in Fig. 4 for $\beta_0 = 1.00$. The cutoff frequency for $\beta_0 = 0.91$ is $\omega_c = 2.60/\tau_D$ and $b^2 = 0.47$, $a = 1.21$.

Transistor Amplifier—Cutoff Frequency[*]

D. E. THOMAS†, SENIOR MEMBER, IRE

Summary—The effect of positive feedback through the internal base resistance of a transistor on circuit cutoff frequency is considered.

AS WOULD BE EXPECTED from standard theory, the cutoff frequency of transistor amplifiers is greatly affected by feedback. In particular, the positive feedback associated with the internal base resistance r_b normally lowers the cutoff frequency. Corresponding changes in low-frequency gain are, of course, also present, so that the product of gain and bandwidth is approximately conserved. This paper gives several examples of transistor amplifiers in which the amplifier cutoff frequency and the transistor α cutoff frequency may differ considerably as a result of this feedback even though linear, nonreactive circuit elements only are used.

Consider first the simple grounded-base transistor amplifier driven from a pure resistance linear generator and driving a linear resistance load. The schematic circuit of such an amplifier and its equivalent circuit are given in Figs. 1 and 2, respectively. In the interest of

Fig. 1—Grounded-base transistor amplifier.
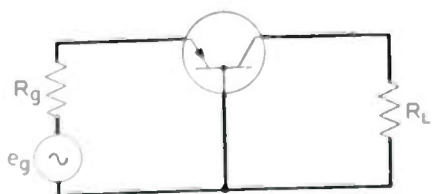
maximum simplicity, let us further assume that the load resistance is negligibly small compared to the transistor

collector resistance. The two linear equations relating the voltages and currents of this amplifier are

$$R_t i_1 + r_b i_2 = e_g$$
$$(r_m + r_b)i_1 + (r_c + r_b + R_L)i_2 = 0, \qquad (1)$$

where

$$R_t = R_g + r_t + r_b.$$

When $R_L \ll r_c$ and $r_b \ll r_m$, the load current i_2 is given by

$$i_2 = -\frac{\alpha e_g}{R_t\left(1 - \frac{\alpha r_b}{R_t}\right)}. \qquad (2)$$

Let us assume the frequency characteristic of alpha is given by

$$\alpha = \frac{\alpha_0}{1 + j\left(\dfrac{f}{f_0}\right)}, \qquad (3)$$

Fig. 2—Equivalent circuit of Fig. 1.

where α_0 is the low-frequency value of alpha and f_0 is the frequency at which the magnitude of α is down by 3 db from its low-frequency value. Many transistors show this type of cutoff. Substituting the expression for alpha in the expression for the load current i_2, we get

$$i_2 = -\frac{e_g \alpha_0}{R_e\left(1 - \frac{\alpha_0 r_b}{R_e}\right)\left[1 + j\left[\frac{f}{f_0\left(1 - \frac{\alpha_0 r_b}{R_e}\right)}\right]\right]}, \quad (4)$$

and the load current i_2 is seen to have a cutoff characteristic similar to that of alpha but in which the cutoff frequency is now $[1 - (\alpha_0 r_b / R_e)]$ times the alpha cutoff frequency.

If we consider the very low-frequency return difference[1] F of the feedback through the base resistance as determined from (1), we see that

$$F = 1 - \mu\beta = \frac{\Delta}{\Delta_0} = 1 - \frac{\alpha_0 r_b}{R_e}, \quad (5)$$

where Δ_0 is equal to Δ when r_m is made equal to zero.

For the assumed α cutoff of 6 db per octave, the circuit cutoff frequency is changed in the same ratio as the gain by the feedback through the base. Therefore, to the extent to which the low-frequency gain is enhanced by positive feedback, the circuit cutoff frequency is reduced with respect to the alpha cutoff frequency by the same ratio. In the asymptotic region well above the alpha cutoff frequency where the feedback loop gain becomes small compared to 1 as a result of α cutoff, the circuit gain is the same as it would be if r_b were equal to 0. This indicates that the reduction in circuit cutoff frequency is caused by increasing the gain by positive feedback at the expense of narrowing the band, a result which is inevitable in the light of well-known circuit and feedback theory.
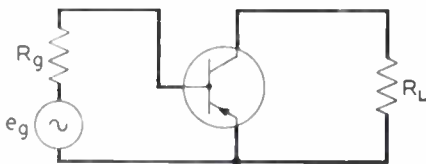
Fig. 3—Grounded-emitter transistor amplifier.
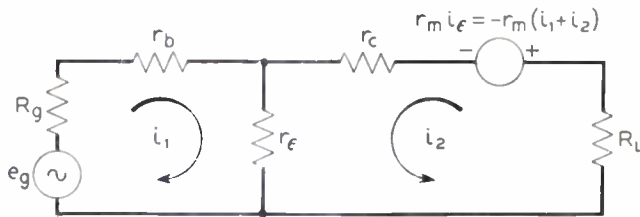
Fig. 4—Equivalent circuit of Fig. 3.

Next consider a grounded-emitter transistor amplifier driven from a pure resistance linear generator and driving a linear resistance load. The schematic circuit of such an amplifier and its equivalent circuit are given in Figs. 3 and 4, respectively. The two linear equations relating the voltages and currents of this amplifier are

[1] See H. W. Bode "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co., Inc., New York, N. Y., pp. 47–48.

$$R_e i_1 + r_e i_2 = e_g, \quad (6)$$

$$(r_e - r_m)i_1 + (r_e + r_c + R_L - r_m)i_2 = 0,$$

where

$$R_e = R_g + r_e + r_b.$$

When $R_L \ll r_c$ and $r_b \ll r_m$, the load current i_2 is given by

$$i_2 = \frac{\alpha e_g}{R_e\left(1 - \frac{\alpha(r_b + R_g)}{R_e}\right)}. \quad (7)$$

If the very low-frequency return difference for the grounded-emitter amplifier is determined from (6), we get

$$F = 1 - \mu\beta = \frac{\Delta}{\Delta_0} = 1 - \frac{\alpha_0(r_b + R_g)}{R_e}. \quad (8)$$

From the similarity of (7) and (8) with (2) and (5), respectively, we can expect the same correlation between gain enhancement and circuit cutoff frequency reduction due to positive feedback through the base circuit as was observed in the grounded-base amplifier. Actually, the effect is increased in the grounded-emitter amplifier by the introduction of the generator resistance in the base circuit, thereby increasing the positive-feedback loop gain.

Next consider the current gain of a grounded-emitter n-p-n amplifier in which the load resistance is negligibly small compared to the collector resistance; thus

$$i_2 = \frac{\alpha}{1 - \alpha} i_1 \quad (9)$$

or for the example given in the article where α equals 0.9785,

$$i_2 = 45.5 i_1.[2]$$

If we put the expression of (3) for alpha in (9), we get

$$i_2 = \frac{\alpha_0}{(1 - \alpha_0)\left(1 + j\left[\frac{f}{(1 - \alpha_0)f_0}\right]\right)} i_1. \quad (10)$$

From (10) we see that to the extent that the current gain is increased at low frequencies by $(1/1-\alpha_0)$ the cutoff frequency is reduced by $(1/1-\alpha_0)$. Thus in the circuit in question, if the transistor has a low frequency α of 0.9785 and an α cutoff given by (3) above with f_0 equal to 2 mc, the cutoff of the current gain given by (10) above would be at $2 \times 10^6(1 - 0.9785)$ or 43 kc.

It is interesting to examine the mechanism which causes the current gain cutoff to be so far below the α cutoff frequency. In the assumed circuit the current

[2] R. L. Wallace and W. J. Pietenpol, "Some circuits properties and application of n-p-n transistors," PROC. I.R.E., vol. 39, pp. 753–767; July, 1951.

gain depends upon $1/1-\alpha$, and is large because α is very close to unity. If the expression for α as a function of frequency given in (3) above is expanded in a Taylor series, we get

$$\alpha = \alpha_0 \left[1 - j\left(\frac{f}{f_0}\right) - \left(\frac{f}{f_0}\right)^2 + j\left(\frac{f}{f_0}\right)^3 + \left(\frac{f}{f_0}\right)^4 - \cdots \right] \quad \text{for} \quad \left(\frac{f}{f_0}\right)^2 < 1. \quad (11)$$

The first quadrature term of α will become important in $1-\alpha$ and, therefore, in changing current gain when $\alpha_0 f / f_0$ becomes appreciable in magnitude compared to $1-\alpha_0$. This will occur at frequencies of the order of $(1-\alpha_0)f_0$. However, since $1-\alpha_0 \ll 1$, this means that all higher-order terms of α than the first quadrature term are negligible in this frequency region. It follows that only the phase shift in α is important in reducing the current gain cutoff frequency in question; the magnitude change in α is second order. The same thing is true of more complex α-frequency cutoffs, and also would hold for phase shifts of the nonminimum type if they should be found to exist.

The above examples demonstrate the necessity of careful consideration of the relationship between circuit and α cutoff frequency in those transistor circuits in which maximum frequency response is a design objective.

On Some Transients in the Pulse Response of Point-Contact Germanium Diodes*

M. C. WALTZ†

Summary—Several recovery time effects in germanium point contact diodes are described and a hypothesis is proposed to explain the observed effects. This hypothesis postulates the presence of traps in the germanium p layer near the point electrode. Measurements and computations based on this hypothesis give trap densities of the order of $10^{16}/\text{cm}^3$, trap depths in the energy band of the order of 0.3 electron volts, and capture cross-section diameters for the traps of the order of 0.3 Å.

INTRODUCTION

THE "HOLE-STORAGE EFFECT" observed and named by Michaels and Meacham[1] has for some time now been a deterrent to the use of germanium diodes in certain pulse applications. The effect has shown up as a form of crosstalk in certain PCM systems, and more recently has been particularly troublesome in high-speed computer systems.

The hole-storage effect is observed whenever a voltage is applied to a germanium diode, suddenly changing it from the conducting to the nonconducting state. The effect may be thought of as a slowness in the ability of the diode to become nonconducting after the nonconducting voltage has been applied. In other words, the high reverse resistance commonly associated with the diode does not immediately appear if there has been a forward current flowing just previous to the application of reverse voltage. Thus, if a diode, for instance, is used as a rectifier, as shown in Fig. 1, it should have an output as

shown at (a) if it exhibits no hole-storage effect, but in a diode showing the hole-storage effect, the output is as shown at (b).

A closer look at certain aspects of the hole-storage effect has brought to light several difficulties with the original concept that the effect was one of hole storage in the simplest sense.
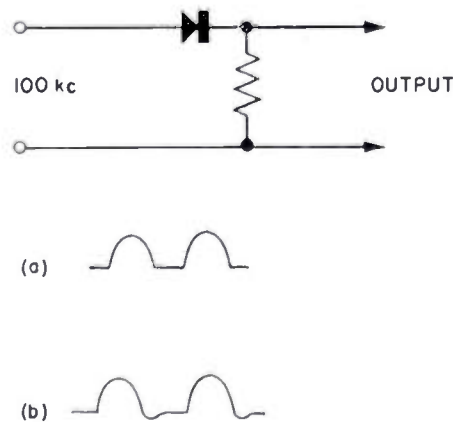
Fig. 1—Output wave forms of germanium point contact diode as a rectifier; (a) at low frequencies, (b) at high frequencies.

PRESENT PICTURE OF THE POINT CONTACT DIODE

The point contact germanium diode which we are going to discuss is one made by the Western Electric Company. It is fabricated in the following fashion: A wafer approximately 0.050 inch $\times$ 0.050 inch $\times$ 0.025 inch is cut from n-type germanium having a purity such that the resistivity is the order of the one-to-ten ohm-cm. The

[1] S. E. Michaels and L. A. Meacham, Letter to the Editor, *Phys. Rev.*, vol. 78, pp. 175–176; April 15, 1950.

material may be either a single crystal or polycrystalline. The wafer is attached to one electrode of the diode by a copper plating and soft soldering technique in such a way that an ohmic or nonrectifying contact is established between the electrode and the germanium wafer. The exposed surface of the germanium wafer is cleaned by a chemical or an electrochemical etching treatment. A wolfram wire, sharpened to a point, is connected to the other electrode of the diode and brought into contact with the exposed, cleaned surface of the wafer. An electrical current of the order of an ampere is forced through the device for about one second to "form" the device. Diodes made by this means are generally referred to as "400 type varistors." A section of such a diode is shown in Fig. 2.
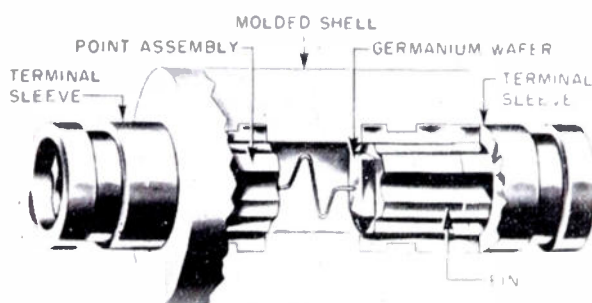
Fig. 2—Sectional view of point contact germanium diode.

The seat of the rectification of this device is in the vicinity of the point contact on the germanium surface. The effect of the forming is to cause a conversion of the n-type germanium near the point to p-type. Fig. 3 shows
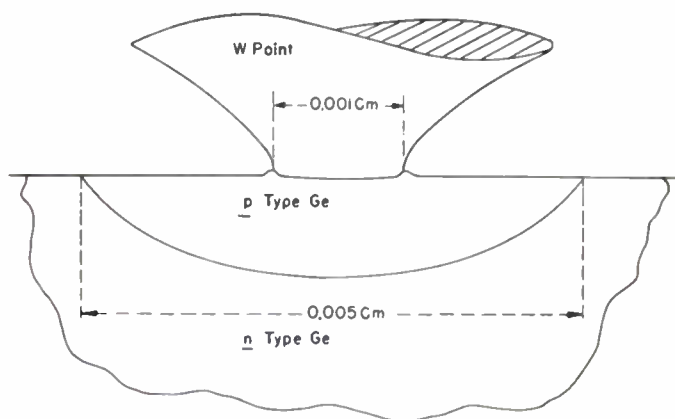
Fig. 3—Section of region near the point in a point contact germanium diode.

a drawing of the section through the point contact. Probe type measurements have been made which permit us to give the approximate dimensions of the formed region, as shown. The resistivity of the p-type germanium is quite high, of the order of ten ohm-cm. The wolfram point is assumed to make a nonrectifying con-

tact to the p-type germanium, all the rectifying behavior being due to the n-p junction created by this means. Computations[2,3] based on the geometry shown in Fig. 3 and using the above assumptions check the order of magnitude of the reverse saturation current observed in this type of diode. This simple picture allows us to explain the dc forward characteristic and the low-level reverse characteristics, i.e., in the saturation region.

MEASUREMENTS OF THE EFFECT

The hole-storage effect has been measured by a variety of means of which the method illustrated in Fig. 1 is the simplest. A method due to R. R. Blair is shown in Fig. 4. In this circuit a current in the forward direction
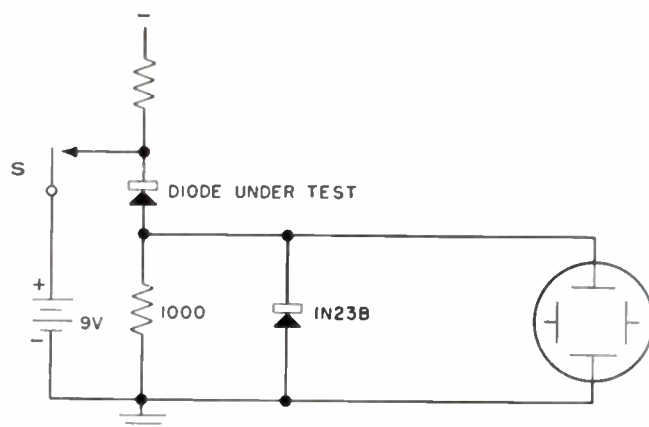
Fig. 4—Schematic diagram of measuring circuit.

is forced through the diode under test. When the contact s closes, a reverse voltage is applied and the peak reverse current measured.

Although the measurements made on this test set using a peak reading voltmeter were consistent and reproducible, they were not always a sufficient criterion to determine the usefulness of a diode in a particular circuit. There seemed to be some aspect of the effect that was not being checked by this circuit. Replacing the peak reading meter by an oscilloscope having a bandwidth of the order of 18 mc allowed us to observe the actual shape of the reverse current transient. Fig. 5(a) shows the approximate shape of the transient as observed with this oscilloscope. Using a viewing oscilloscope with a wider bandwidth (70 mc) the transient was as shown in Fig. 5(b). The difference in shape between 5(a) and 5(b) indicated that the true shape of the initial rise of the transient was not being observed. Accordingly, a special oscilloscope assembled by L. H. Germer was used to examine the leading edge of the transient. This instrument had a rise time of 1.5 mμs. The shape

[2] H. C. Torrey and C. A. Whitmer, "Crystal Rectifier," McGraw-Hill Publishing Co., New York, N. Y.; 1948.
[3] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., New York, N. Y.; 1950.

of the leading edge is shown in Fig. 5(c). Due to the type of sweep used with this instrument, it was difficult to locate the base line for longer times than shown in Fig. 5(c).
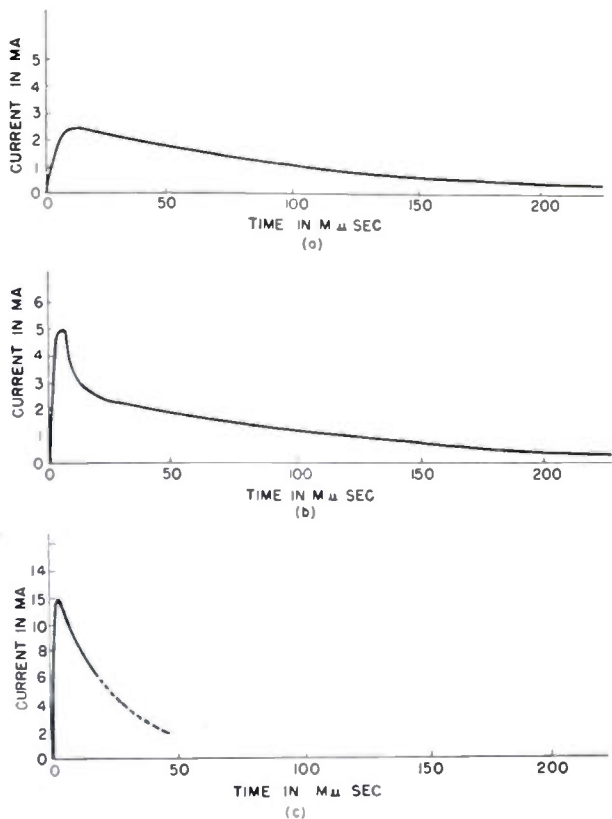
Fig. 5—Shape of the reverse current transient as viewed with varying bandwidth oscilloscope. (a) 18 mc; (b) 70 mc; (c) 200 mc.

From the above measurements the transient is seen to consist of at least two parts. The first is a very large current lasting for a time of the order of 10 mμs, followed by a trailing off part having a time constant of the order of 50 mμs. Fig. 6 shows a typical idealized transient.
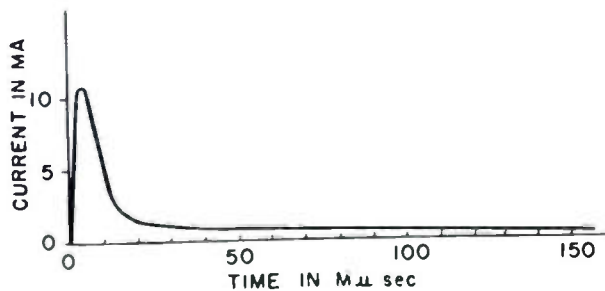
Fig. 6—Idealized reverse current transient.

Some Computations Concerning the Effect

For simplicity in making the computations, we will assume a linear geometry rather than a radial one. This will introduce an error, but will be sufficiently accurate at this stage to check the order of magnitude of some of the parameters.

First let us compute the fraction of current carried by holes in this particular n-p junction during forward current flow.

$$\gamma_p = \frac{I_p}{I} = \frac{1}{1 + \sigma_n L_p / \sigma_p W} \quad (1)$$

where

I_p = current carried by holes,
I = total current,
σ_n and σ_p = electrical conductivity of n and p regions,
L_p = diffusion length for holes = $\sqrt{D\tau_p}$,

where

D = diffusion constant for holes,
τ_p = average lifetime of holes,
W = thickness of p layer.

Now assuming reasonable values for these parameters, namely,

$\sigma_n = \frac{1}{5}$ mhos/cm,
$\sigma_p = \frac{1}{10}$ mhos/cm,
$L_p = \sqrt{42 \times 50 \times 10^{-6}} = 45 \times 10^{-3}$ cm,
$W = 2 \times 10^{-3}$ cm,

we get

$$\gamma_p = \frac{1}{1 + \dfrac{45 \times 10^{-3}}{5} \Big/ \dfrac{2 \times 10^{-3}}{10}} = \frac{1}{1 + 45}$$

$$\approx 2 \text{ per cent.}$$

(The error due to the assumption of a linear geometry will make the above computed value too low.)

This tells us that when forward current is flowing through the diode, only about 2 per cent of the current is carried by holes moving into the n material and therefore about 98 per cent of the current is carried by electrons flowing into the p region. We can therefore almost disregard the current flow into the n region, and deal only with the p region. As supporting evidence, very low values of γ have been observed when using a formed point of this type as an emitter in a transistor.

Now let us compute the transit time of an electron across the p layer. This time should be a measure of the length of the transient when the voltage on the diode is changed from forward to reverse.

The transit time, t_t, neglecting diffusion, is

$$t_t = \frac{W}{\mu_n E} \quad (2)$$

where μ_n = mobility of electrons,
E = field,

assuming a drop of 0.1 volt in the p layer, a drop in agreement with spreading resistance calculations, we get

$$t_t = \frac{2 \times 10^{-3}}{3,500 \times \dfrac{0.1}{2 \times 10^{-3}}} = \frac{4}{3.5} \times 10^{-8} \approx 10 \text{ mμs.}$$

This time is very much shorter than the time associated with the tail of the transient, but does check fairly

closely with the initial part of the transient. Thus, we are led to believe that the first part of the transient may be an actual transit storage effect, but it is an "electron storage" instead of a "hole storage."

We are now confronted with the problem of explaining the second part of the transient. It could be due to the collection of the holes that have been injected into the n material. Even though the fraction of current carried by the holes is very small, the volume for the storage of holes is large enough to permit a storage of charge as large as we have shown in Fig. 6. However, if the tail were due to this cause, the decay time of the tail would then be related to the lifetime of holes in the n material. Our observations show that the decay time seldom exceeds 50 mμs, and is not often much less than 30 mμs. These time constants are two to three orders of magnitude smaller than the lifetime of the germanium.

Since the simple picture does not seem to explain the observations, we are led to postulate a more complex structure. Suppose, during the forming process, that not only are acceptor levels generated in the region under the point, but that trapping levels are also introduced. Thus, we are to imagine that the region under the point has a certain density of traps. Now when forward current is flowing in the diode, some of the electrons in flowing through the p region are caught in traps. The number caught in the traps would be proportional to the density of traps, the density of electrons, and the capture cross section of the trap. The time an electron would stay in the trap before it was released due to thermal agitation would be related to the depth of the trap in the energy gap and to the absolute temperature. Now, when the applied voltage is reversed, we would first collect the electrons in transit to give us the spike of the transient curve, then we would collect the electrons as they were released by the traps to give us the tail of the curve.

To check the validity of this hypothesis, it is instructive to make a few computations to check that unreasonable values of trap density and depth of trap in the energy band are not obtained.

First, if the current in the tail is due to trapping, then the area under the tail should be a measure of the total electrons trapped. From Fig. 6 it is seen that the extrapolated current at zero time is approximate 2 ma, and the time constant is about 50 mμs. Thus the area under the curve is

$$Q = 2 \times 10^{-3} \times 50 \times 10^{-9} = 10^{-10} \text{ coulombs}$$

$$n = \frac{10^{-10}}{1.6 \times 10^{-19}} = 6 \times 10^8 \text{ electrons}$$

giving a density of trapped electrons, assuming geometry of Fig. 4, of

$$n_t = \frac{6 \times 10^8}{3 \times 10^{-8}} = 2 \times 10^{16} \text{ electrons/cm}^3.$$

If all the traps are filled, this would also correspond to the trap density, so

$$N_t = 2 \times 10^{16} \text{ traps/cm}^3.$$

This is a reasonable trap density, corresponding to one trap for each million germanium atoms.

To find the depth of the trap in the energy level, we can make use of the approximate equation[4]

$$u = \frac{kT}{q} \ln (\nu\tau) \text{ electron volts} \qquad (3)$$

where τ is mean time in trap, ν is frequency of atomic vibration $\sim 10^{13}$/sec for most materials, u is depth of trap. Making $\tau = 50$ mμs $= 50 \times 10^{-9}$ sec., we get $u = 1/40 \ln (10^{13} \times 50 \times 10^{-9}) = 0.32$ electron volt.

This value for the depth of the traps appears to be reasonable, whereas a depth of 0.4 or more ev in germanium would not be reasonable.

Since temperature enters into the trap depth in (3), it is possible to make measurements of the recovery characteristic as a function of temperature, and check theory and experiment. Also, since it is possible to fill the traps by shining light on the germanium, the effect can be measured as a function of the light intensity, to provide a further check.

The presence of the electrons in the traps, if the traps were originally electrically neutral, requires a hole to be drawn from the metal electrode to establish neutrality in the p-region. The presence of these extra holes should contribute to the conductivity of the p material and thereby lower the spreading resistance when forward current is flowing. An effect of this type is observed. Fig. 7 shows the forward current that flows when a for-
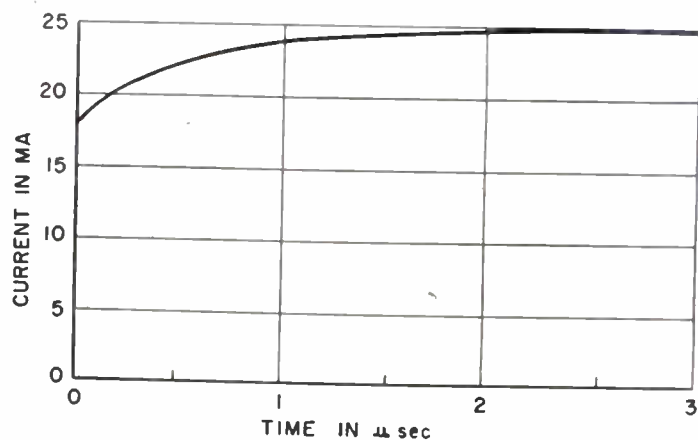
Fig. 7—Typical forward current transient.

ward voltage is suddenly applied. The forward current rises rapidly to some value, then, much more slowly, to some final value. The time constant of the slow rise is the order of 1μ second, with the amount of the slow rise amounting to from zero to 50 per cent of the final value.

[4] N. F. Mott and R. W. Gurney, "Electronic Processes in Ionic Crystals," Clarendon Press, Oxford; 1940.

The presence of reverse current just preceding the forward has no additional effect.

If this slow rise is due to the trapping of electrons in the p layer with subsequent increase in the number of hole carriers, then the time constant of the slow rise should be the average time an electron can stay in the p region before becoming trapped. The capture time, along with the trap density, allow us to compute a capture cross section for the traps. Thus,

$$\sigma = \frac{1}{\tau_f v N_T} \tag{4}$$

where τ_f = capture time, v = thermal velocity of electron $(1 \times 10^7$ cm/sec$)$, N_T = number traps.

Substituting $\tau_f = 10^{-6}$ sec, $v = 10^7$ cm/sec, $N_T = 2 \times 10^{16}/\text{cm}^3$, we get

$$\sigma = \frac{1}{10^{-6} \times 10^7 \times 2 \times 10^{16}} = \frac{1}{2 \times 10^{17}} = 5 \times 10^{18} \text{ cm}^2,$$

or a diameter of approximately 0.3 Å. Here, again, the answer is reasonable.

In view of the above checks between observations and computations, it would seem reasonable to suppose that the role of trapping levels in explaining some of the transient phenomena in point contact germanium diodes has been established. More experimental work will be required to establish definitely the place of traps in these phenomena.

Pulse Duration and Repetition Rate of a Transistor Multivibrator*

G. E. McDUFFIE, Jr.†

Summary—Expressions are obtained for the pulse duration and repetition rate of an astable multivibrator circuit employing contact-type transistors. These expressions are confirmed by experiment for repetition rates of from 200 to 10,000 cycles per second.

Introduction

UNDER certain circuit conditions, the emitter, base, and collector of contact-type transistors can exhibit negative-resistance characteristics.[1] It is the purpose of this paper to present expressions for an astable multivibrator circuit, relating pulse length and repetition rate with the emitter negative-resistance characteristics of contact-type transistors and the external RC termination.

Relaxation oscillation will take place if a load consisting of a parallel resistance and capacitance is placed across the emitter terminals of a contact-type transistor employing base feedback.[2] Such a circuit and its associated curve are shown in Fig. 1. For astable operation, the value of emitter load resistance R_E is such that the load line intersects the negative-resistance portion of the curve. The operating point would normally be stable. However, the condenser produces a virtual short-circuit in parallel with R_E and the operating point moves

around paths 1, 2, and 3. Thus the condenser is alternately charging and discharging between voltages V_{EP} and V_{EV}. When the operating point is in the nega-
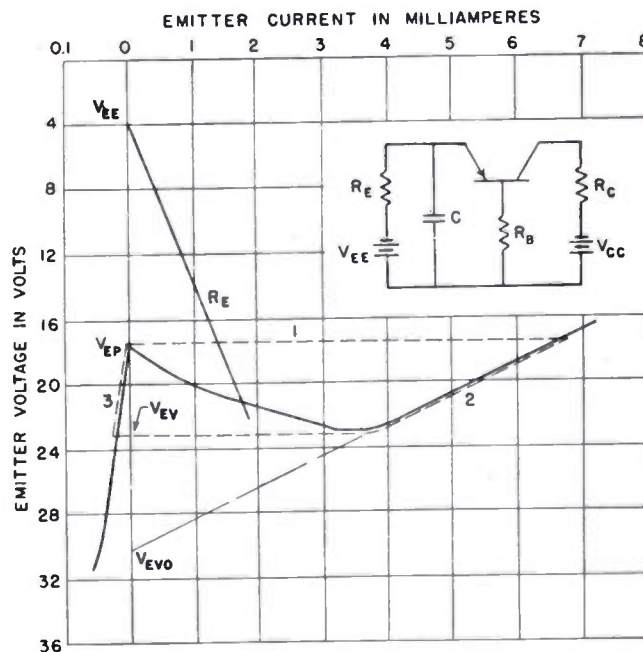
Fig. 1—Operating characteristics of emitter negative-resistance astable multivibrator.

tive low-current region, the condenser is discharging through R_E in parallel with a relatively high emitter-input resistance. When operation is in the positive high-current region, the condenser is charging through a relatively low emitter-input resistance paralleled by

[1] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.* vol. 28, pp. 367–401; July, 1949.

[2] A. E. Anderson, "Transistors in switching circuits," Proc. I.R.E., vol. 40, pp. 1541–1559; this issue.

R_E. Signal output of rectangular pulses is obtained from the collector. Low emitter-current operation corresponds to the "dead time" between collector pulses and high-current emitter current to collector-pulse duration.

Fig. 2 shows oscillograms of emitter and collector voltage waveforms as measured between emitter and ground and collector and ground for astable operation of a typical transistor multivibrator employing a development model of the Western Electric A-1698 contact-type transistor.
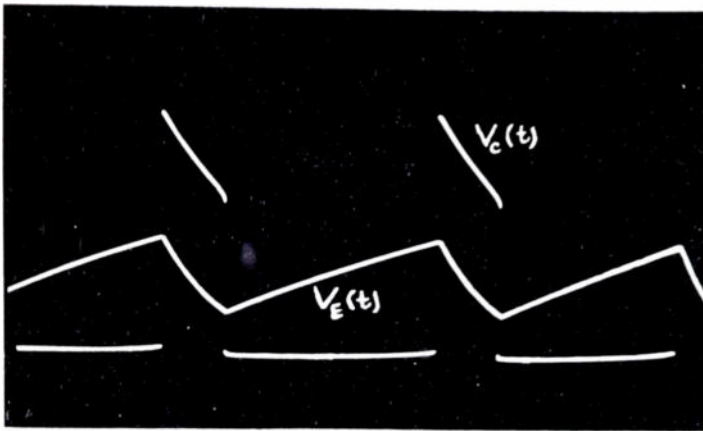
Fig. 2—Emitter voltage, $V_E(t)$, superimposed on collector voltage $V_C(t)$. Total time base is approximately 12 msec.

COLLECTOR PULSE LENGTH AND REPETITION RATE

The equivalent circuit of an astable emitter negative-resistance transistor multivibrator can be reduced to that shown in Fig. 3 with the following assumptions:
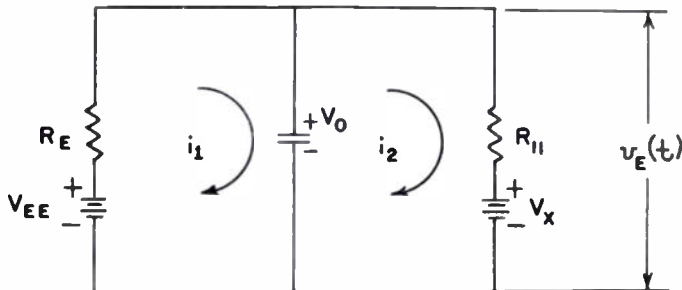
Fig. 3—Equivalent circuit of emitter negative-resistance astable multivibrator.

The equivalent circuit of an astable emitter negative-resistance transistor multivibrator can be reduced to that shown in Fig. 3 with the following assumptions: (a) transition from the negative, low-current region of the emitter negative-resistance characteristic curve to the positive current-positive-resistance region is instantaneous, resulting in zero rise and fall time of the collector pulse (i.e., only frequencies well below cutoff will be considered); (b) the emitter negative-resistance characteristic curve is linear in the negative, low-current

region and the positive current-positive-resistance region.[3]

In the equivalent circuit, R_{IN} is the input resistance as seen between emitter and ground and corresponds to the reciprocal of the slope of the assumed linear region under consideration. V_x is the dc component of the assumed linear regions, and corresponds to the intersections of the linear approximations and the voltage axis. The intercepts for the two assumed linear regions are shown in Fig. 1; the low-current intercept is designated V_{EP} and the positive current-positive-resistance intercept V_{EVo}. V_{EE} is the terminal voltage of the emitter supply.

To find the general expression for the time required for the voltage across the emitter terminals to change from one limiting value to another, the differential loop equations for the equivalent circuit are solved for $i_2(t)$ and the following expression obtained:[4]

$$i_2(t) = \frac{V_0 - V_x}{R_{IN}} e^{-\beta t} + \frac{V_{EE} - V_x}{R_E + R_{IN}} (1 - e^{-\beta t}). \quad (1)$$

V_0 is the initial charge on the condenser and

$$\beta = \frac{R_E + R_{IN}}{R_E R_{IN} C}.$$

The instantaneous emitter voltage, $v_E(t)$, is found by the summation of the voltage drop across R_{IN} and V_x.

$$v_E(t) = V_x + (V_0 - V_x)e^{-\beta t}$$
$$+ \frac{(V_{EE} - V_x)R_{IN}}{R_E + R_{IN}} (1 - e^{-\beta t}). \quad (2)$$

Solving for t gives

$$t = \frac{1}{\beta} \ln \frac{V_0(R_E + R_{IN}) - (V_x R_E + V_{EE}R_{IN})}{v_E(t)[R_E + R_{IN}] - (V_x R_E + V_{EE}R_{IN})}. \quad (3)$$

Equation (3) defines the time required for the voltage across the condenser to change from one limiting value to another.

Consider astable operation as shown in Fig. 1. In the low-current region of the curve $v_E(t)$ will change from V_{EV} to V_{EP}. Thus

$$V_0 = V_{EV}$$
$$v_E(t) = V_{EP}$$
$$V_x = V_{EP},$$

and (3) becomes

$$t_L = \frac{1}{\beta_L} \ln \frac{V_{EV}(R_E + R_{INL}) - (V_{EP}R_E + V_{EE}R_{INL})}{V_{EP}(R_E + R_{INL}) - (V_{EP}R_E + V_{EE}R_{INL})}. \quad (4)$$

Equation (4) defines the time during which operation is in the low-current region; this corresponds to the quiescent time between collector pulses.

---

[3] G. F. Farley, "Dynamics of transistor negative-resistance circuits," PROC. I.R.E., pp. 1497–1508; this issue.
[4] See Appendix.

In the high-current region of the curve, $v_E(t)$ varies between V_{EP} and V_{EV} and

$$V_0 = V_{EP}$$
$$v_E(t) = V_{EV}$$
$$V_X = V_{EVO}.$$

Equation (3) now becomes

$$t_H = \frac{1}{\beta_H} \ln \frac{V_{EP}(R_E + R_{INH}) - (V_{EVO}R_E + V_{EE}R_{INH})}{V_{EV}(R_E + R_{INH}) - (V_{EVO}R_E + V_{EE}R_{INH})} \cdot \quad (5)$$

Equation (5) defines the time during which operation is in the high-current region and corresponds to collector pulse duration. The reciprocal of the repetition rate is the sum of t_L and t_H.

EXPERIMENTAL RESULTS

To verify the developed expressions, the static V_E versus I_E curve was plotted. Typical values of circuit parameters giving negative-resistance are shown in Table I. From this curve the necessary emitter negative-

TABLE I
CIRCUIT PARAMETERS

$$R_C = 2,200 \text{ ohms}$$
$$R_B = 6,800 \text{ ohms}$$
$$V_{CC} = -45 \text{ volts}$$
$$V_{EE} = 0 \text{ volts}$$

resistance parameters were obtained; values are shown in Table II. Collector pulse lengths and repetition rates were then measured for an emitter load resistance of 15,600 ohms and various values of capacity.

TABLE II
NEGATIVE-RESISTANCE PARAMETERS

$$R_{IN} \text{ (high-current region)} = 1820 \text{ ohms}$$
$$R_{IN} \text{ (low-current range)} = 240,000 \text{ ohms}$$
$$V_{EV} = -23.0 \text{ volts}$$
$$V_{EVO} = -30.1 \text{ volts}$$
$$V_{EP} = -16.8 \text{ volts}$$

Fig. 4 is a plot of calculated and measured values of collector pulse lengths and the reciprocals of the repetition rates versus capacity. The curves were plotted using the calculated values of t_H and $t_L + t_H$. The measured values were then spotted on the co-ordinates and enclosed by circles representing approximately 5-per cent deviation. From these curves it can be seen that the developed expressions are valid over a wide range of pulse duration and repetition rates.
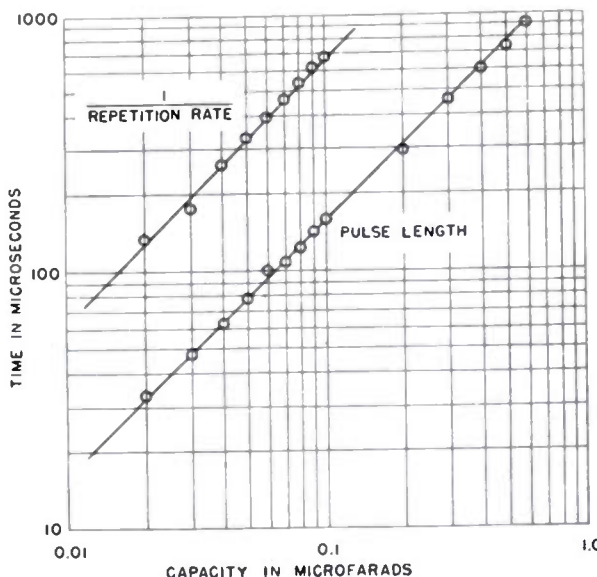
Fig. 4—Pulse duration and repetition rate versus capacity.

ACKNOWLEDGMENTS

APPENDIX

The differential loop equations are as follows:

$$V_{EE} = i_1 R_E + \frac{1}{C} \int i_1 dt - \frac{1}{C} \int i_2 dt,$$

$$-V_X = -\frac{1}{C} \int i_1 dt + i_2 R_{11} + \frac{1}{C} \int i_2 dt.$$

Introducing Laplacian transforms gives

$$\frac{V_{EE}}{S} = I_1(S) R_E + \frac{1}{CS} I_1(S) + \frac{V_0}{S} - \frac{1}{CS} I_2(S),$$

$$-\frac{V_X}{S} = -\frac{1}{CS} I_1(S) + R_{11} I_2(S) + \frac{1}{CS} I_2(S) - \frac{V_0}{S} \cdot$$

Solving for $I_2(s)$,

$$I_2(S) = \frac{V_0 - V_X}{R_{11}(S + \beta)} + \frac{V_{EE} - V_X}{S(S + \beta) R_E R_{11} C} \cdot$$

The current as a function of time is then

$$i_2(t) = \frac{V_0 - V_X}{R_{11}} e^{-\beta t} + \frac{V_{EE} - V_X}{R_E + R_{11}} (1 - e^{-\beta t}).$$

Junction Transistor Equivalent Circuits and Vacuum-Tube Analogy*

L. J. GIACOLETTO†, SENIOR MEMBER, IRE

Summary—The junction transistor possesses operating characteristics that are closely comparable to a modified triode vacuum tube. A direct comparison between the two devices is particularly interesting if a π equivalent circuit is used for the transistor. The vacuum-tube analogy and transistor equivalent circuits are considered in some detail in this paper. For purposes of comparison, a tabulation of operating characteristics of a transistor and a vacuum tube has been prepared.

INTRODUCTION

IN RECENT PUBLICATIONS,[1,2] the point-contact transistor has been considered as a device that is a dual of a vacuum tube. The junction transistor,[3] on the other hand, can be related directly to the vacuum tube.[4] This direct relationship is emphasized by comparison with a slightly unusual vacuum tube.
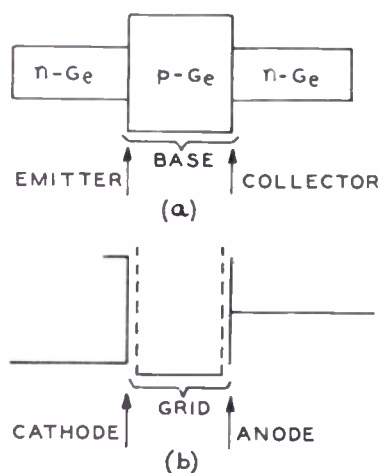
Fig. 1—*n-p-n* transistor and corresponding vacuum tube.

[1] R. L. Wallace, Jr. and G. Raisbeck, "Duality as a guide in transistor circuit design," *Bell Sys. Tech. Jour.*, vol. 30, pp. 331–417; April, 1951.

[2] R. L. Wallace, Jr., "Duality, a new approach to transistor circuit design," PROC. I.R.E., vol 39, p. 702: June, 1951.

[3] R. L. Wallace, Jr. and W. J. Pietenpol, "Some circuit properties and applications of *n-p-n* transistors," PROC. I.R.E., vol. 39, pp. 753–767; July, 1951.

[4] A direct relationship between the junction transistor and vacuum tube was indicated by W. Shockley, U. S. Patent 2,569,347. See, also, W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Sys. Tech. Jour.*, vol. 28, pp. 435–489; July, 1949 and W. Shockley, M. Sparks, and G. K. Teal, "*p-n* junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; July, 1951.

VACUUM-TUBE ANALOGY

Consider an *n-p-n* junction transistor in Fig. 1(a) consisting of an emitter plane, base, and collector plane. The comparable vacuum tube is shown immediately below in Fig. 1(b). The cathode is directly analogous to the emitter, and the anode to the collector. In the transistor the active region is the *p*-type germanium base. The corresponding active region in the vacuum tube is the space included between the cathode and anode. In order that the vacuum-tube active space be field free, as is approximately the case in the transistor, the vacuum-tube active space is enclosed between two grids connected together to form a grid structure which is now comparable with the transistor base element.

The junction transistor collector characteristics are of a pentode character, *i.e.*, the collector current saturates as a function of the collector voltage. This is equivalent to saying that the collector current is independent of the collector voltage, or more accurately, that the emitter is not influenced to any appreciable degree by the collector voltage. The same operation is characteristic of the structure of Fig. 1(b) since the cathode field will be essentially independent of the anode voltage. For the structure of Fig. 1(b) the cathode will have to be biased negatively with respect to the grid in order to get any significant anode current. This again is entirely analogous to similar bias arrangements in a transistor. The ensuing grid current is comparable to the base current. A current amplification factor, $\alpha = -I_a/I_k$, can be defined for the vacuum tube similar to that defined for a transistor. Here, I_a is the short-circuited ac anode current and I_k is the ac cathode current.

Consider the grid-to-anode region. The electrons leaving the grid structure are drawn to the anode because of the positive voltage thereupon. If the grid-to-anode voltage is small, the electron flow will be of a space-charge character. For a somewhat larger grid-to-anode voltage the anode current reaches a value which remains the same for any larger grid-to-anode voltage, *i.e.*, the anode current has saturated at a value corresponding approximately with the current entering the grid structure. Now, if the cathode is made more negative with respect to the grid, more electrons are injected into the grid structure so that the anode current saturation value is correspondingly increased. A closely analogous operation takes place in the *n-p-n* transistor. The base-to-emitter voltage determines the

number of electrons entering the base wafer. These electrons flow to the collector and become the corresponding saturation collector current. The correspondence between collector characteristics and anode characteristics is immediately apparent.

The discussion above has centered about an *n-p-n* junction transistor. A *p-n-p* junction transistor might be considered analogous to a vacuum tube of the type shown in Fig. 1(b) in which the cathode emits positrons. The charge carrier for either device is now positive so that all voltage polarities must be reversed after which the statements made above can be applied.

In the past, the collector characteristics have usually been plotted with the emitter current as the running parameter. This method of presentation using data of an early experimental RCA *p-n-p* junction transistor is shown in Fig. 2(a). In order to emphasize the tube analogy it is desirable to use the base-to-emitter voltage as the running parameter as shown in Fig. 2(b). In accordance with vacuum-tube practice, the voltages in Fig. 2(b) are measured with the emitter as the reference point so that V_{CE} will differ slightly from V_{CB}.[5]
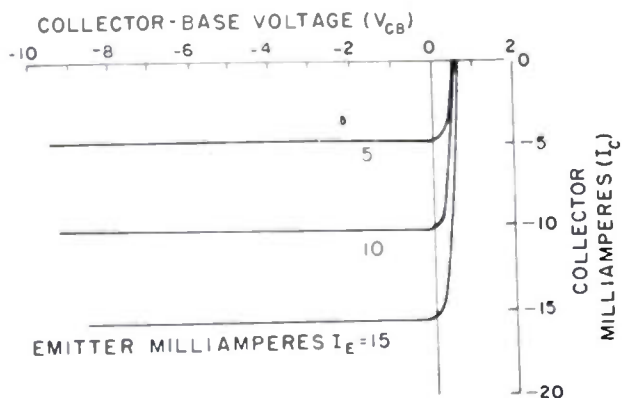
Fig. 2(a)—Collector characteristics of *p-n-p* junction transistor.
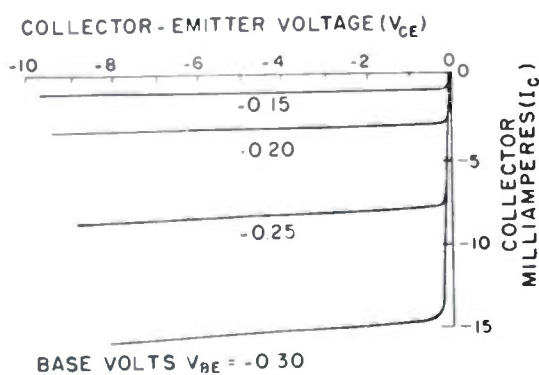
Fig. 2(b)—Collector characteristics of *p-n-p* junction transistor.

Fig. 3(a) shows the input characteristics of the transistor to complement the output characteristics of Fig. 2(b). The associated transfer characteristics are shown in Fig. 3(b) and are similar to transfer characteristics for the vacuum tube of Fig. 1(b) if due allowance is

made for the difference in carrier polarity. It is apparent that the transistor exhibits nonlinear transfer characteristics. The linear operation of a transistor is therefore not appreciably different from a vacuum tube, with degeneration being required for either device if linear operation is desired.
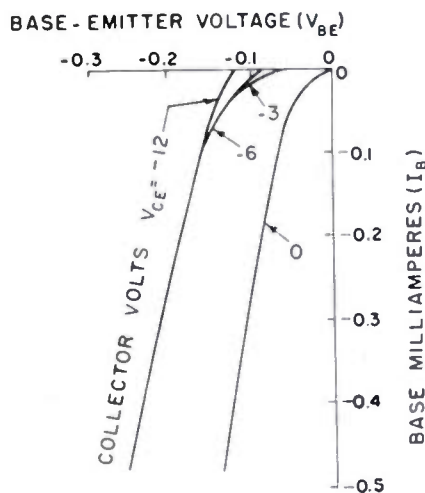
Fig. 3(a)—Base characteristics of *p-n-p* junction transistor.
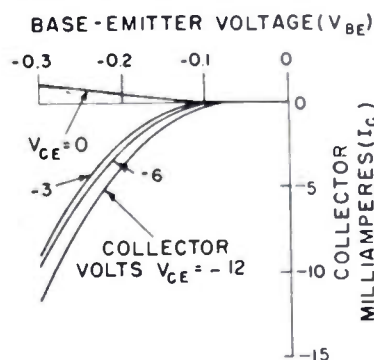
Fig. 3(b)—Transistor characteristics of *p-n-p* junction transistor.

TRANSISTOR TRANSCONDUCTANCE

The transistor transconductance is shown in Fig. 4. The g_m curve is roughly exponential and is somewhat similar to vacuum tube g_m curves except for magnitude.
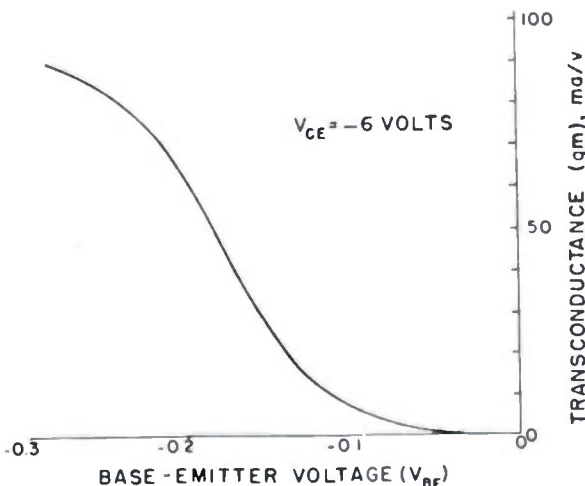
Fig. 4—Transconductance characteristics of *p-n-p* junction transistor.

[5] Throughout this paper, upper-case letter subscripts will be used to denote dc values, and lower-case letter subscripts will be used to denote root-mean-square ac values.

The nature of the transistor g_m curve can be understood in an approximate way from the transistor operation described above. Since the collector current is approximately the same as the emitter current, the transconductance is about equal to the emitter conductance, $1/r_e$, which Shockley[4] has shown to be eI_E/kT. Consequently, the transconductance to emitter current ratio, $g_m/I_E = e/kT = 38.6$. Here, e is the absolute charge of an electron, k is Boltzmann's constant, and T is the absolute operating temperature of the transistor. This ratio of 38.6 is a theoretical value, but does not differ appreciably from measured values when the emitter current is small. For vacuum tubes the same transconductance to current ratio has a theoretical value of 11.6 although measured values are hardly ever larger than 2. Thus the transistor has at least a 20:1 larger value than vacuum tubes for this important constant.

π Equivalent Circuits

The T equivalent circuit has been used most frequently in connection with transistors.[6,7,8] In contrast, the π equivalent circuit has usually been used in connection with grounded-cathode vacuum tubes. If Fig. 5(a) there is shown the common-emitter T equivalent circuit for the p-n-p transistor being studied. The major
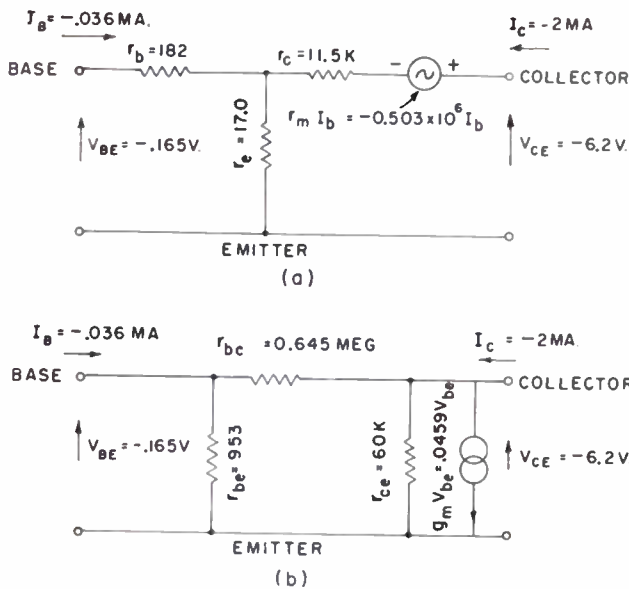
Fig. 5—T and π equivalent circuits of p-n-p junction transistor.

difference[8] between this equivalent circuit and the more familiar common-base circuit is that the resistor in series

with the collector terminal for the latter is $r_c - r_m = 0.514$ MΩ. Fig. 5(b) shows the equivalent π circuit of the same p-n-p junction transistor shown in Fig. 5(a). The T equivalent circuit is transformed to a π circuit in accordance with the theory developed by Peterson.[9]

From the component values given in Fig. 5(b) it is seen that, although the resistance between the input and output is not infinite, as is the case for a vacuum tube, it is large so that, to an approximation it is permissible to consider the input and output circuits independently. For comparison purposes, the π equivalent circuits for the common-base and the common-collector circuits are shown in Fig. 6(a) and (b), respectively.
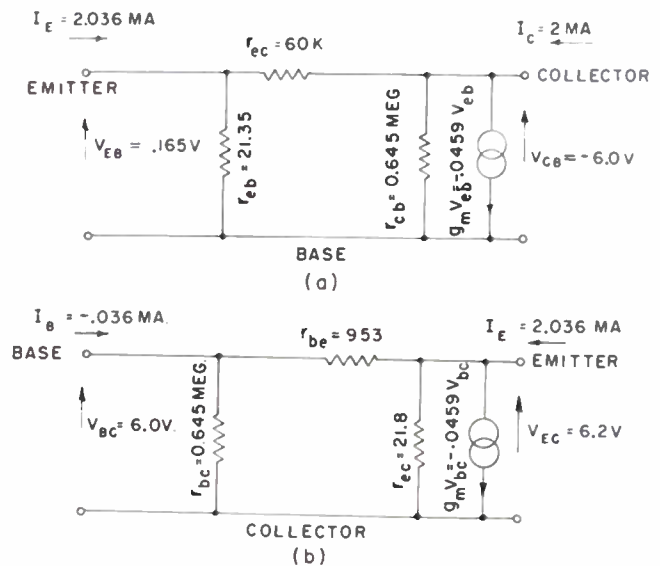
Fig. 6—π equivalent circuits of p-n-p junction transistor.
(a) Common base. (b) Common collector.

It is, of course, immaterial whether a direct or a dual comparison is drawn between a vacuum tube and a transistor or whether a T or π equivalent circuit is employed. It is believed that for many engineers who have developed a background of vacuum-tube experience, the direct comparison and the π equivalent circuit will be easier to use. The comparison is sufficiently complete that most of the vacuum-tube technology may be transferred directly to the transistor. An important difference between the two devices is the presence in the transistor of a finite resistance between base and collector.

Comparative Operating Data

Since a direct comparison has been drawn between the vacuum tube and the transistor, it is interesting to prepare a comparison between the two. For this comparison, shown in Table I, a 6AG5 pentode and the previously described RCA p-n-p junction transistor have been selected. A few comments concerning this tabula-

[6] W. M. Webster, E. Eberhard, and L. E. Barton, "Some novel circuits for the three-terminal semiconductor amplifier," *RCA Rev.* vol. 10, pp. 5–16; March, 1949.

[7] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, pp. 367–400; July, 1949.

[8] M. J. E. Golay, "The equivalent circuit of the transistor," Proc. I.R.E., vol. 40, p. 360; March, 1952.

[9] L. C. Peterson, "Equivalent circuits of linear active four-terminal networks," *Bell Sys. Tech. Jour.*, vol. 27, pp. 593–622; October, 1948.

ion are appropriate. There is no special reason for comparing the 6AG5 with the junction transistor other than that the 6AG5 is a well known vacuum tube. It would be more appropriate to compare the transistor with the unconventional vacuum tube of the type shown in Fig. l(b). The large g_m and g_m/I_E ratio of the transistor have already been mentioned. The input impedance of a negative grid vacuum tube is very large as compared with 438 ohms for the transistor. The large input impedance is an important advantage for the vacuum tube and accounts in part for the large power gain. In order to facilitate comparison, a grid resistor of 10 megohms has been associated as an integral part of the vacuum tube. Large grid resistors of this magnitude are sometimes employed in low-frequency, low-level vacuum tube circuits. The current and noise associated with the grid resistor are charged to the vacuum tube and account for the grid current, the finite current amplification factor and the 3-db noise factor.

In studying the data of Table I it is important to remember that the transistor selected is an early developmental unit and that transistors are in their early stages of development as compared with vacuum tubes.

TABLE I
COMPARISON BETWEEN VACUUM TUBE AND TRANSISTOR CHARACTERISTICS AND PERFORMANCE

6AG5 pentode	RCA p-n-p junction transistor
Heater power $= 1.9$ w	Heater power $= 0$ w
Grid-cathode voltage $= V_{GK}$ $= -0.9$ v	Base-emitter voltage $= V_{BE}$ $= -0.165$ v
Grid current $= I_G = 0.09$ μa	Base current $= I_B = -0.036$ ma
Screen-cathode voltage $= V_{SK}$ $= 125$ v	Collector-emitter voltage $= V_{CE}$ $= -6.2$ v
Anode-cathode voltage $= V_{AK}$ $= 125$ v	
Screen current $= I_S = 2.1$ ma	Collector current $= I_c = -2.0$ ma
Anode current $= I_A = 7.2$ ma	
Current amplification factor (grounded grid) $= -(I_a/I_k) \sim 0.775$	Current amplification factor (common base) $= -(I_c/I_e) = 0.977$
Current amplification factor (grounded cathode) $= -(I_a/I_g) \sim -80 \times 10^3$	Current amplification factor (common emitter) $= -(I_c/I_b) = -43.6$
$r_{ak} = 10$ MΩ	$r_{be} = 953\Omega$
$r_{ga} = \infty$	$r_{bc} = 0.645$ MΩ
$r_{ak} = 0.5$ MΩ	$r_{ce} = 60$ KΩ
$g_m = 5.1$ ma/v	$g_m = 45.9$ ma/v
$g_m/I_K = 0.55$	$g_m/I_E = 22.6$
$R_{\text{input match.}} = 10$ MΩ	$R_{\text{input match.}} = 438$ Ω
$R_{\text{output match.}} = 0.5$ MΩ	$R_{\text{output match.}} = 25.3$ KΩ
Maximum power gain $= 75$ db	Maximum power gain $= 40$ db
Noise factor at 1 kc $= 3$ db	Noise factor at 1 kc $= 22$ db

Matrix Representation of Transistor Circuits[*]

JACOB SHEKEL†, ASSOCIATE, IRE

Summary—The paper outlines a method to treat transistors as 3-terminal "black boxes," represented by their admittance matrices. Instead of using an equivalent circuit with internal sources, the treatment is direct and uses only measurable quantities. Once the admittance matrix of the grounded-base transistor is measured, it is easy to derive the matrices for grounded-emitter and grounded-collector stages. Matrix representation also presents a direct method for analyzing cascaded stages.

The method is illustrated by an analysis of circuits with a transistor that has been described elsewhere.[1]

INTRODUCTION

THE APPLICATION of transistors as amplifiers or oscillators has recently received a considerable amount of attention. Apart from the description of the physical principles underlying transistor operation, some recent papers[2,3,4] have treated the transistor empirically, describing it—for engineering purposes—

as a "black box" with certain relations between the currents and voltages at its terminals. For greater ease in visualizing its operations, certain equivalent circuits have been proposed, all of them based on the loop-equation analysis of networks.

Now, in a "black-box" treatment, by definition, the only measurable currents are those entering any terminal, and the only measurable voltages are those between pairs of terminals, more often the voltages between any terminal and a reference "ground" terminal. Any internal generators introduced into hypothetical loops inside the black box are needless complexities that have, no doubt, some value as aids to visualize relation between terminal phenomena, but are nevertheless not essential to the representation and computation of the circuits.

In this paper an attempt is made to discuss transistor circuits, specifying a transistor only by its measurable terminal magnitudes: input and transfer impedance (or admittance). The mathematical method for this treatment was developed in a recent paper,[5] and its main points are briefly reviewed here.

[1] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, p. 373; July, 1949.

[2] R. M. Ryder and R. J. Kircher, *ibid.*, p. 367.

[3] R. L. Wallace, Jr. and G. Raisbeck, "Duality as a guide in transistor circuit design," *Bell Sys. Tech. Jour.*, vol. 30, p. 381; April, 1951.

[4] R. L. Wallace, Jr. and W. J. Pietenpol, "Some circuit properties and applications of n-p-n transistors," *Bell Sys. Tech. Jour.*, vol. 30, p. 530; July, 1951. Also PROC. I.R.E., vol. 39, p. 753; July, 1951.

[5] J. Shekel, "Application of matrices to the analysis of linear networks," *Bull. Scien. Dep't.*, Ministry of Defense, No. 25–26; Tel Aviv, Israel; July–August, 1951. (In Hebrew.)

A *terminal* of a network is a point at which external sources and loads may be connected. The only definable and measurable currents are those entering a terminal; the only definable and measurable voltages are those between pairs of terminals.

The *current* entering a network is the one-column matrix **I** whose components I_i are the currents into the i-th terminal.

The *voltage* of the network is a one-column matrix **V** whose components V_i are the voltages between the i-th terminal and an arbitrary reference terminal.

The *admittance* of the network is the square matrix **Y**, defined by

$$\mathbf{I} = \mathbf{YV}. \tag{1}$$

Owing to the linear nature of (1), this treatment may be applied to linear networks only. In nonlinear networks, the treatment will give only the "small-signal" approximation, where I_i and V_i are only changes of current and voltage about a fixed operating point.

In an n-terminal network, only $n-1$ currents are independent, and the voltage of one terminal may be fixed arbitrarily. It was shown that those two facts put certain restrictions on **Y**, namely, that the sum of any row or any column in it is zero. This singular matrix was termed *the indefinite admittance matrix*, due to the fact of the voltage reference terminal being undefined.

The following operations on networks are represented by corresponding changes in the admittance:

1. Specifying a terminal as "ground" leads to a reduced $(n-1) \times (n-1)$ matrix, which is obtained from **Y** by crossing out the corresponding row and column.

2. Shorting two terminals together is represented by adding the corresponding rows and columns to form one row and column, respectively.

3. When two or more networks are connected in parallel, their corresponding admittance matrices are added together. In this way, the admittance of the composite network may be computed by adding together the matrices of the composing elements, which may be two-terminal admittors or "black boxes" of higher complexity.

Main advantages of above-mentioned method are:

1. In full analogy to the treatment of simple circuits, the only quantities that are used in representation or computation are admittances and impedances. No engineer would dream of computing the resistance of a combination of parallel resistors by introducing hypothetical currents and voltages which later cancel out; nor would he compute the voltage transfer ratio of a simple voltage divider except by writing down the ratio of two resistances. This same treatment is extended to composite networks, giving all results without introducing loop currents which are to be cancelled out later.

2. The treatment allows complete freedom in changing the grounded terminal. Once the indefinite admittance is known, it is immaterial which terminal, if any,

is to be grounded. In transistor circuits, this allows equally easy calculation of grounded base, emitter or collector stages, or even stages where all three terminals are ungrounded.

THE GENERAL "BLACK-BOX" TREATMENT

As an introduction to matrix representation of transistors, the general treatment of an n-terminal "black box" will be outlined. For the sake of simplicity in writing the matrix expressions, the treatment of a 4-terminal box will be shown. The method is easily extended to any other number of terminals.

When attempting to measure the input and transfer impedances, the ground terminal has to be specified. Let this be terminal 4.

Equation (1), written out in full, is

$$
\begin{Vmatrix} I_1 \\ I_2 \\ I_3 \end{Vmatrix} = \begin{Vmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \end{Vmatrix} \begin{Vmatrix} V_1 \\ V_2 \\ V_3 \end{Vmatrix}, \tag{2}
$$

giving, for example,

$$I_2 = Y_{21}V_1 + Y_{22}V_2 + Y_{23}V_3.$$

If all voltages are zero, except, say, V_j, then

$$I_i = Y_{ij}V_j. \tag{3}$$

The Y_{ij} are, therefore, the short-circuit input and transfer admittances. They may be determined experimentally in two ways:

1. In linear circuits, a voltage V_j is applied between the j-th terminal and ground, and all other terminals are grounded. The currents into all terminals are then measured, giving the Y_{ij}.

2. In nonlinear "black boxes," e.g., vacuum tubes, the various current-voltage characteristics are plotted, and, for any operating point,

$$Y_{ij} = \frac{\partial I_i}{\partial V_j}.$$

The "black box" under consideration may be unstable under short-circuit conditions, but then an alternative method for determining the network matrix is possible by inverting (2) to give

$$
\begin{Vmatrix} V_1 \\ V_2 \\ V_3 \end{Vmatrix} = \begin{Vmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{Vmatrix} \begin{Vmatrix} I_1 \\ I_2 \\ I_3 \end{Vmatrix}. \tag{4}
$$

If all currents be zero, except, say, I_j,

$$V_i = Z_{ij}I_j. \tag{5}$$

The Z_{ij} are, then, the open-circuit input and transfer impedances, and may be measured directly or computed from suitable characteristic curves.

Comparing (2) and (4). it is evident that the **Z** matrix is the inverse of the **Y** matrix.

Once the **Y** matrix is found, directly or by inverting the **Z** matrix, it is brought to the indefinite form

$$\begin{Vmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} \\ Y_{31} & Y_{32} & Y_{33} & Y_{34} \\ Y_{41} & Y_{42} & Y_{43} & Y_{44} \end{Vmatrix}, \qquad (6)$$

where

$$Y_{14} = -Y_{11} - Y_{12} - Y_{13}, \quad Y_{41} = -Y_{11} - Y_{21} - Y_{31},$$

and so on,

$$Y_{44} = Y_{11} + Y_{12} + Y_{13} + Y_{21} + Y_{22} + Y_{23}$$
$$+ Y_{31} + Y_{32} + Y_{33},$$

so that the sum of every row and column is zero.

If the "black box" is used in a circuit with terminal 2 grounded, the admittance will be

$$\begin{Vmatrix} Y_{11} & Y_{13} & Y_{14} \\ Y_{31} & Y_{33} & Y_{34} \\ Y_{41} & Y_{43} & Y_{44} \end{Vmatrix}. \qquad (7)$$

Supposing that a conductance G is connected between terminals 1 and 3, the admittance of the combination will be

$$\begin{Vmatrix} Y_{11} + G & Y_{13} - G & Y_{14} \\ Y_{31} - G & Y_{33} + G & Y_{34} \\ Y_{41} & Y_{43} & Y_{44} \end{Vmatrix}. \qquad (8)$$

An element connected between any terminal and ground will appear only once, on the main diagonal of the matrix.

Let **Y** be the matrix of the composite network "black box" and added elements and let **Z** be the matrix inverse to **Y**. If i be the input terminal and k the output terminal, the following relations may be computed:

$$\text{input impedance} = Z_{ii}$$
$$\text{output impedance} = Z_{kk}$$
$$\text{open-circuit voltage amplification} = Z_{ki} : Z_{ii}$$
$$\text{short-circuit current amplification} = Y_{ki} : Y_{ii}.$$

A passive bilateral "black box" is characterized by an admittance matrix which is symmetric, and has nonnegative elements on its main diagonal and nonpositive elements elsewhere. This is not the case with vacuum tubes, neither with transistors, which are active and unilateral elements.

ADMITTANCE MATRIX OF A TYPICAL TRANSISTOR

To illustrate the application of the above methods to transistors, a typical specimen is chosen.[1] In a type-A transistor, the following open-circuit impedances were measured at a certain operating point:

$$Z_{11} = \quad 530 \text{ ohms} \qquad Z_{12} = \quad 290 \text{ ohms}$$
$$Z_{21} = 34,000 \text{ ohms} \qquad Z_{22} = 19,000 \text{ ohms},$$

where 1 denotes the emitter terminal and 2 the collector terminal, and the base terminal is grounded.

The impedance matrix is, then,

$$\mathbf{Z} = \begin{Vmatrix} 530 & 290 \\ 34,000 & 19,000 \end{Vmatrix}. \qquad (9)$$

Inverting it, the admittance matrix for grounded-base operation is obtained.

$$\mathbf{Y} = \begin{Vmatrix} +90.48 & -1.38 \\ -161.90 & +2.52 \end{Vmatrix} \times 10^{-3}. \qquad (10)$$

A third row and column are added, so as to make the sum of every row and column equal to zero, thus giving the indefinite admittance matrix of the transistor

$$\mathbf{Y} = \begin{Vmatrix} +90.48 & -1.38 & -89.10 \\ -161.90 & +2.52 & +159.38 \\ +71.42 & -1.14 & -70.28 \end{Vmatrix} \times 10^{-3}. \qquad (11)$$

The matrix is not symmetric, and a negative value appears on the main diagonal. The last fact is indicative of short-circuit instability. Suppose that the emitter and collector should be shorted together and grounded, the first and second rows and columns are to be crossed out, and the matrix degenerates into a single negative conductance of -70.28 millimhos, which is the input admittance at the base terminal.

GROUNDED-BASE AMPLIFIER

A grounded-base amplifier is shown schematically in Fig. 1, which is its equivalent circuit for small signals (all dc potential differences disregarded). G_g is the internal
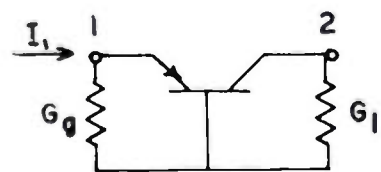
Fig. 1—Grounded-base amplifier.

conductance of the input signal source and G_l the conductance of the load. Typical values are

$$G_g = 2 \quad \text{millimhos}$$
$$G_l = 0.05 \text{ millimhos}.$$

The admittance of the circuit is constructed as follows:

1. The third row and column are omitted from (11) (grounded base).
2. G_g and G_l are added to Y_{11} and Y_{22}, respectively.

$$\mathbf{Y} = \begin{Vmatrix} +92.48 & -1.38 \\ -161.90 & +2.57 \end{Vmatrix} \times 10^{-3}. \qquad (12)$$

The voltage gain in this circuit is[6]

$$\text{forward voltage gain} = \frac{Z_{21}}{Z_{11}} = \frac{-Y_{21}}{Y_{22}} = \frac{+161.9}{+2.57} = 63$$

$$\text{backward voltage gain} = \frac{Z_{12}}{Z_{22}} = \frac{-Y_{12}}{Y_{11}} = \frac{1.38}{92.48} = 0.015.$$

The positive sign of both ratios shows that there is no phase reversal.

The power dissipated in the load is $V_2^2 G_L$, the available power from the source is $I_1^2 : (4 G_g)$, and the power gain is then

$$\text{forward power gain} = 4 G_g G_l (V_2 : I_1)^2 = 4 G_g G_l Z_{21}^2$$

$$= 4 \times 2 \times 0.05 \times 10^{-6}$$

$$\times \left(\frac{161.9 \times 10^{-3} \cdot}{92.48 \times 2.57 - 161.9 \times 1.38} \right)$$

$$= 51.$$

The backward power gain is, similarly, $4 G_g G_l Z_{12}^2$, and the ratio

$$\frac{\text{forward power gain}}{\text{backward power gain}} = \left(\frac{Z_{21}}{Z_{12}} \right)^2 = \left(\frac{Y_{21}}{Y_{12}} \right)^2$$

is independent on the source and load, being a characteristic of the transistor itself. In this case, the ratio is

$$\frac{\text{forward power gain}}{\text{backward power gain}} = \left(\frac{161.9}{1.38} \right)^2 = 1.37 \times 10^4.$$

GROUNDED-COLLECTOR AMPLIFER

The admittance matrix of a grounded-collector stage (Fig. 2) is constructed in the following steps:

1. Second row and column crossed out of (11) (grounded collector).
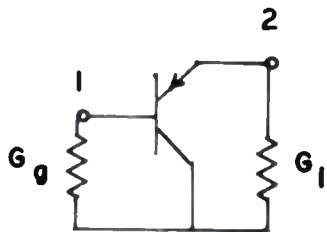
Fig. 2—Grounded-collector amplifier.

2. The remaining rows and columns are interchanged, so that the indices 1 and 2 denote base and emitter, respectively, conforming with Fig. 2.

3. G_g and G_l are added to Y_{11} and Y_{22}, respectively. To illustrate an actual computation, let

$$G_g = 0.05 \text{ millimhos}$$

$$G_l = 0.1 \text{ millimhos.}$$

$$\mathbf{Y} = \begin{Vmatrix} -70.38 & +71.42 \\ -89.10 & +90.48 \end{Vmatrix} \times 10^{-3}. \qquad (13)$$

[6] Z_{ik} is the co-factor of Y_{ki}, divided by the determinant of Y. The ratio of two Z's is therefore the ratio of the corresponding co-factors, which are, in a 2×2 matrix, just elements of Y.

By the same general expressions used in the preceding sections,

$$\text{forward voltage gain} = \frac{89.10}{90.48} = +0.985$$

$$\text{backward voltage gain} = \frac{-71.42}{70.38} = -1.01,$$

the negative sign being indicative of a phase reversal in the backward direction.

$$\frac{\text{forward power gain}}{\text{backward power gain}} = \left(\frac{89.10}{71.42} \right)^2 = 1.55.$$

CASCADE AMPLIFER

The admittance of a network of cascaded transistors (Fig. 3) is found as a sum of three admittance matrices, corresponding to the component networks (Fig. 4).
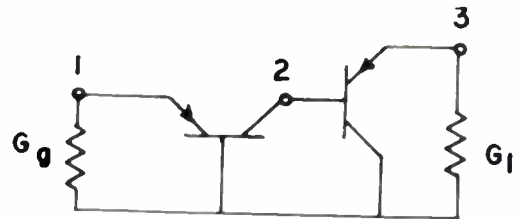
Fig. 3—Cascaded amplifier stages.
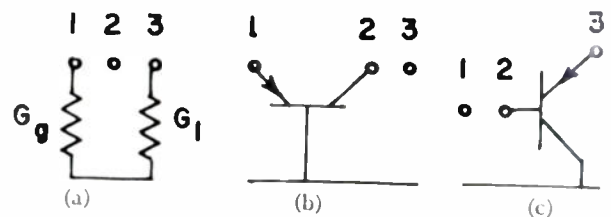
Fig. 4—Component networks of the amplifier of Fig. 3.

1. The passive network.

$$\mathbf{Y}_a = \begin{Vmatrix} G_g & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & G_l \end{Vmatrix}. \qquad (14)$$

2. The first transistor, the usual grounded-base matrix, with a third row and column of zeroes corresponding to the isolated third terminal.

$$\mathbf{Y}_b = \begin{Vmatrix} +90.48 & -1.38 & 0 \\ -161.90 & +2.52 & 0 \\ 0 & 0 & 0 \end{Vmatrix} \times 10^{-3}. \qquad (15)$$

3. The second row and column are omitted from the general matrix (11), and the remaining rows and columns interchanged to conform with the terminal numbering in Fig. 4 (c).

$$\mathbf{Y}_c = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & -70.28 & +71.42 \\ 0 & -89.10 & +90.48 \end{Vmatrix} \times 10^{-3}. \qquad (16)$$

he three matrices are then added together. Putting

$$G_g = 2 \quad \text{millimhos}$$

$$G_l = 0.1 \text{ millimhos,}$$

his gives

$$Y = \left\|\begin{array}{ccc} +92.48 & -1.38 & 0 \\ -161.90 & -67.76 & +71.42 \\ 0 & -89.10 & +90.58 \end{array}\right\| \times 10^{-3}. \quad (17)$$

The forward voltage gain is

$$\frac{Z_{31}}{Z_{11}} = \frac{\left|\begin{array}{cc} -161.90 & -67.76 \\ 0 & -89.10 \end{array}\right|}{\left|\begin{array}{cc} -67.76 & +71.42 \\ -89.10 & +90.58 \end{array}\right|} = +64.$$

Similarly, all other results may be computed by the same general expressions.

Conclusion

A general method of treating n-terminal "black boxes" has been outlined, which may be applied to any linear "black box," whether it be active or passive, bilateral or unilateral. The method was then applied to a certain type of transistor, and some circuit arrangements were analyzed.

The paper was not intended to be a treatise on transistors but a discussion of methods of circuit analysis; therefore, no general conclusions were derived about transistor circuits and their properties.

Dynamics of Transistor Negative-Resistance Circuits*

B. G. FARLEY†, ASSOCIATE, IRE

Summary—A general method is presented for calculating approximately the behavior of many nonlinear circuits by dividing the region of operation into subregions, within each of which the circuit may be considered linear to a good approximation. The method is applied to a high-speed transistor switching circuit as an illustrative example.

I. Introduction

TRANSISTORS have shown considerable promise as elements of high-speed switching circuits in computer and control systems requiring rapid pulse handling.

Their operation in such circuits is of a type which has received relatively little attention in the literature as yet, and certain complexities arise which are comparatively new. The purpose of this paper is to present an approximate method for analyzing nonlinear circuits and to show how the method may be applied in accordance with the best present experience to point-contact transistor circuits. These methods will serve to elucidate the three main aspects of electronic switching action which it is desirable to be able to calculate, namely the action upon triggering, the output voltage, and the rise and fall times of the output wave.

The method for treating the nonlinear circuit will be simply to approximate it with two or more linear circuits. To accomplish this, a set of independent variables

of the circuit is picked in terms of which the solution is desired. ("Independent variable" here means that such variables can be chosen independently as initial conditions. The number of independent variables is the number of "degrees of freedom" of the circuit. For the transistor, the emitter input current and voltage will be used as independent variables.) Nonlinearity usually manifests itself as circuit parameters which depend on current or voltage. An approximation is then made by dividing the plane or space of the independent variables into regions within each of which it is considered that the circuit parameters are sufficiently constant to give a satisfactory result if exact constancy is assumed. Thus, the nonlinear problem is broken up into several ordinary linear problems, and the solutions are patched together at boundaries of linear regions.

It is to be noted that the method outlined above is of general applicability for circuits having voltage and current-dependent elements, and of course, the approximation can be made as close as desired, at the cost of increasing labor, by subdividing the operating region into smaller and smaller subregions. The effects in the transistor due to minority-carrier storage may not come within the voltage or current dependent class, since the effect may depend on previous history. In any case, calculation of the storage effects is less satisfactory than the others, at present. This will be discussed in Section VIII.

A remark concerning negative resistance is pertinent here. Terms like "negative-resistance oscillations" as

applied to circuits with a negative slope in the *v-i* characteristic at dc are common, but it should be noted that such terms are misleading in that undue emphasis is placed on the dc input characteristic as a method of classification. Any two-terminal lumped-constant impedance, even though gain elements are present, such as a feedback amplifier, can be represented as a network of resistances, capacitances, and inductances if negative resistances are allowed.[1] Therefore, all impedances can be treated by the same general method, even though they may not show a negative resistance at dc. The method presented in this paper can be applied to any circuit in any region where it can be approximated with a linear network.

The plan of the paper is to develop the theory in Sections II–VI, and apply it to illustrative transistor circuits in Sections VII and VIII. Readers interested chiefly in results may read these last two first.

II. Transient Behavior Within Linear Regions

After the total region of nonlinear operation is divided up into linear regions, the behavior within any one is, of course, simply that of a linear, lumped-constant network. Methods for obtaining time-dependent solutions of such a network are well-known. The natures of these solutions when viewed in the independent variable co-ordinate plane (or space, in the multidimensional case) are not so well-known, however, although they have been studied in connection with the nonlinear theory of oscillation in the two-dimensional case.[2]

For our purposes, the types of "trajectories" or paths in a two-reactive-element circuit will be briefly illustrated in both time-dependent and voltage-current plane representations because our application will make use of only two variables. Any two independent parameters which are sufficient to determine the state of the circuit at any instant may be used in this case; but since our transistor application will make use of the emitter voltage-current plane, we will use it for discussion.

In a circuit containing two independent reactive elements, the independent variables all satisfy a second-order differential equation with time. The characteristic equation is therefore of the form[3]

$$s^2 + As + B = 0, \tag{1}$$

where A and B depend only on the circuit constants.

If the roots of this equation are s_1 and s_2, and $s_1 \neq s_2$, then the time-dependent transient solutions of the differential equation have the form

$$
\begin{aligned}
i &= a_{11}e^{s_1 t} + a_{12}e^{s_2 t}, \\
v &= a_{21}e^{s_1 t} + a_{22}e^{s_2 t},
\end{aligned} \tag{2}
$$

when v and i are referred to an origin at the point of equilibrium, that is, the dc portion of the solution is omitted. The a's depend on both the initial conditions and the circuit constants.

The following cases are possible:

Root Conditions	Nature of Solution	Name in Voltage-Current Plane
1. Real		
$s_1, s_2 > 0$	increasing exponential	node, unstable
$s_1 > 0, s_2 < 0$	increasing exponential	saddle point, unstable
$s_1, s_2 < 0$	decreasing exponential	node, stable
2. Complex		
real part > 0	increasing exponential	focus (spiral), stable
real part < 0	decreasing exponential	focus (spiral), stable

That is, the transient consists of oscillations or exponentials, either stable or unstable. These familiar time curves are shown in Fig. 1. Any system admitting an increasing transient is, of course, unstable.
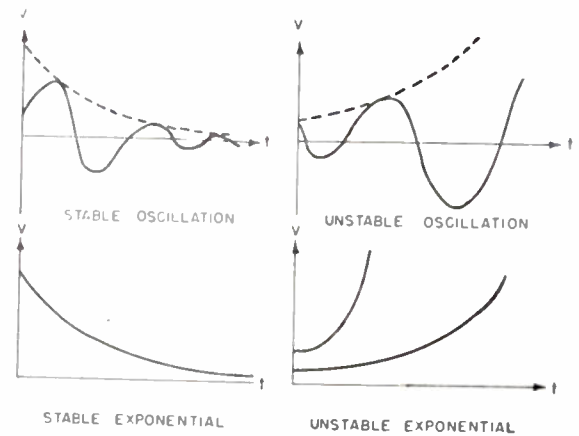
Fig. 1—Two-dimensional linear circuit motions versus time.

A trajectory in the voltage-current plane is a combination of two motions at right angles in a plane. Both motions have the same characteristic roots, but differ in coefficients, which are determined by the initial conditions. Figs. 2 and 3 show representative trajectories (from a family filling the plane) of the types listed above. The curves show only the transient or alternating components of current and voltage in each case. This point is stressed because, in the transistor case, the equilibrium points do not fall at zero dc emitter volt-

[1] H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co., Inc., New York; p. 185; 1947.

[2] See, for example, A. A. Andronow and C. E. Chaikin, "Theory of Oscillations," Princeton University Press, Princeton, N. J.; 1949.

M. Minorsky, "Introduction to Nonlinear Mechanics," Edwards Bros., Ann Arbor, Mich.; 1947.

W. Reichardt, "Uniform relationship between sinusoids, relaxation vibrations, and relaxation discontinuities," and "Negative resistances, their characteristics and effects," *Elek. Nach. Tech.*, vol. 20; pp. 213–225; September, 1943; and vol. 20, pp. 76–87; March, 1943.

T. Von Karman, "The engineer grapples with nonlinear problems," *Bull. Amer. Math. Soc.*, vol. 46, pp. 615–683; August, 1940.

P. S. Hsia, "Graphical analysis for nonlinear systems," *Proc. IEE* (London), vol. 99, pt. 2, pp. 125–134; April, 1952.

A review of work on nonlinear problems and excellent bibliography is given in M. L. Cartwright, "Nonlinear vibrations," *Brit. Ass. Advan. Sci.* (London), vol. 6, pp. 64–74; April, 1949.

[3] We use the notation and terminology of M. F. Gardner and J. L. Barnes, "Transients in Linear Systems," John Wiley & Sons, New York, N. Y., vol. 1; 1942. See, for example, p. 171.

ge and current, but rather are determined by the inter-
:ctions of load lines with the dc characteristics.

. More exact information about the trajectories of the
ransients in linear networks can be obtained as fol-
ows: In the general case of n variables, by suitable
ransformations, the first derivative of each independent
ariable can be obtained as a linear function of all the
idependent variables themselves.[4]
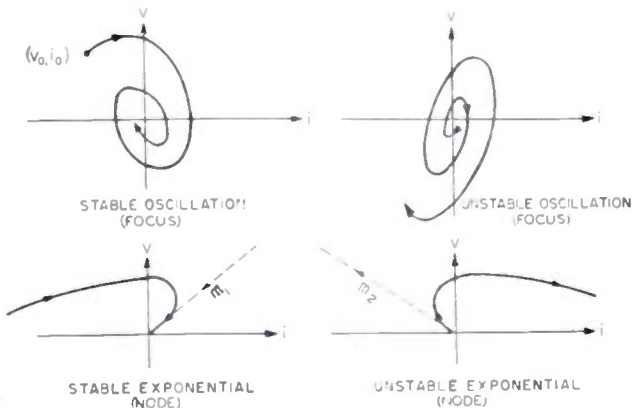
Fig. 2—Two-dimensional linear circuit motions in
voltage-current plane-focus and node.

In two dimentions which we will continue to use for
lustration, these equations are therefore of the form

$$\frac{di}{dt} = k_{11}v + k_{12}i,$$

$$\frac{dv}{di} = \frac{k_{21}v + k_{22}i}{k_{11}v + k_{12}i} \qquad (3)$$

$$\frac{dv}{dt} = k_{21}v + k_{22}i$$

vhen referred to the equilibrium point as origin, and
;ive the direction of the trajectories of the transients
.t any point. The k's depend only on circuit constants.

From these equations it can be seen at once that only
one trajectory can pass through any point, since the
lope dv/dt is determined by the co-ordinates (i, v) alone,
ind that a path passes through every point (except
:he origin, which is a path itself); thus, the entire future
notion is determined by the initial voltage and current.
=urthermore, all trajectories cut a given straight line
:hrough the origin (constant v/i) at the same slope. In
particular, if v is plotted vertically, the paths are al-
ways horizontal along the line $v = -(k_{22}/k_{21})i$, and
always vertical along $v = -(k_{12}/k_{11})i$.

More can be learned about the paths in the case of
exponential curves (real roots) from (2). We define

$$s_1 = 1/2(-A + \sqrt{\Delta})$$
$$s_2 = 1/2(-A - \sqrt{\Delta}), \qquad (4)$$

[4] For an indication of a general procedure see E. A. Guillemin,
"Communication Networks," John Wiley and Sons, New York, vol.
1, p. 223; 1931.

where $\Delta = A^2 - 4B$.

Thus $s_1 > s_2$ when they are real.

Therefore, since

$$\frac{v}{i} = \frac{a_{21} + a_{22}e^{(s_2-s_1)t}}{a_{11} + a_{12}e^{(s_2-s_1)t}} \text{ by (2),} \qquad (5)$$

$$t \rightarrow \infty, \qquad \frac{v}{i} \rightarrow \frac{a_{21}}{a_{11}} \equiv m_1$$

and

$$t \rightarrow -\infty, \qquad \frac{v}{i} \rightarrow \frac{a_{22}}{a_{12}} \equiv m_2. \qquad (6)$$

Thus, if both $s_1, s_2 < 0$ (stable node), the path ends at
the origin ($t = \infty$) with a slope of m_1, and at its other
end ($t = -\infty$) approaches a slope of m_2.

If $s_1, s_2 > 0$ (unstable node), the path starts at the
origin ($t = -\infty$) with a slope of m_2 and at its far end
($t = \infty$) approaches a slope of m_1.

In case $s_2 < 0 < s_1$, both ends of the path flee the origin,
the end corresponding to $t = \infty$ approaches a slope of m_1,
and to $t = -\infty$ approaches a slope of m_2.
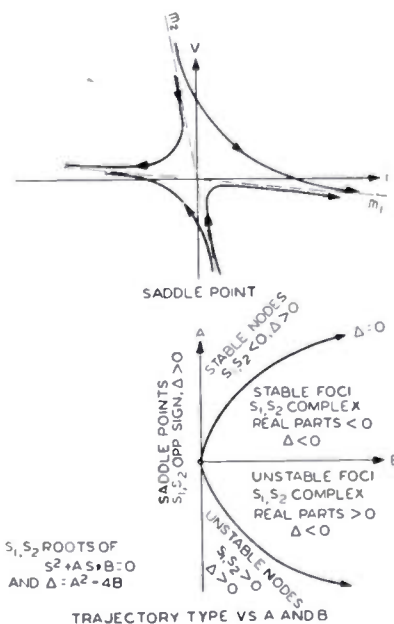
Fig. 3—(top) Saddle-point motion; (bottom) Stability diagram
for two-dimensional circuit.

It can be shown that $v = m_1 i$ and $v = m_2 i$ are actually
paths, the only straight lines which are.

It can also be shown that

$$m_1 = \frac{k_{22}}{s_1 - k_{21}} = \frac{s_1 - k_{12}}{k_{11}}$$

$$m_2 = \frac{k_{22}}{s_2 - k_{21}} = \frac{s_2 - k_{12}}{k_{11}}, \qquad (7)$$

which serve to compute the m's from circuit constants.

Also, it is seen that if $v = m_1 i$ (i.e., traveling on or near the m_1 line), $di/dt = s_1 i$, and $dv/dt = s_1 v$; and if m_2 replaces m_1, s_2 replaces s_1.

From these facts about the trajectories it is relatively simple to sketch them graphically, and thus obtain a clear and simple picture of the motion of the "system point" (i, v) over the plane. Fig. 3 also gives a diagram showing the nature of the paths as determined by A and B, and thus serves as a stability chart for the two-dimensional case.

Although an expression can be derived for the relation between v and i along a trajectory with t eliminated, it is too complicated to be of much use; so that when accurate plotting is desired, it is best to use both time-dependent equations to calculate v and i at the same time. For future reference (2) is given for any two-dimensional case in terms of the initial voltage v_0, current i_0, and the constants previously defined.

For $\Delta > 0$,

$$i(t) = \frac{1}{m_2 - m_1}\left[(m_2 i_0 - v_0)e^{s_1(t-t_0)}\right.$$
$$\left. - (m_1 i_0 - v_0)e^{s_2(t-t_0)}\right]$$

$$v(t) = \frac{1}{m_2 - m_1}\left[m_1(m_2 i_0 - v_0)e^{s_1(t-t_0)}\right.$$
$$\left. - m_2(m_1 i_0 - v_0)e^{s_2(t-t_0)}\right], \quad (8)$$

where

$$i(t_0) = i_0, \qquad v(t_0) = v_0.$$

For $\Delta < 0$,

$$i(t) = M e^{-A/2(t-t_0)}\sin\left[\frac{\sqrt{|\Delta|}}{2}(t-t_0) + \theta\right],$$

$$v(t) = N e^{-A/2(t-t_0)}\sin\left[\frac{\sqrt{|\Delta|}}{2}(t-t_0) + \phi\right], \quad (9)$$

where

$$M = -\frac{\sqrt{\left[i_0\left(\frac{A}{2} + k_{12}\right) + k_{11}v_0\right]^2 + \frac{|\Delta|}{4}i_0^2}}{\frac{\sqrt{|\Delta|}}{2}},$$

$$N = -\frac{\sqrt{\frac{|\Delta|}{4}v_0^2 + \left[v_0\left(\frac{A}{2} + k_{21}\right) + k_{22}i_0\right]^2}}{\frac{\sqrt{|\Delta|}}{2}},$$

$$\cos\theta = -\frac{i_0\left(\frac{A}{2} + k_{12}\right) + k_{11}v_0}{\sqrt{\left[i_0\left(\frac{A}{2} + k_{12}\right) + k_{11}v_0\right]^2 + \frac{|\Delta|}{4}i_0^2}},$$

$$\sin\theta = \frac{i_0}{M}.$$

$$\cos\phi = -\frac{v_0\left(\frac{A}{2} + k_{21}\right) + k_{22}i_0}{\sqrt{\frac{|\Delta|}{4}v_0^2 + \left[v_0\left(\frac{A}{2} + k_{21}\right) + k_{22}i_0\right]^2}},$$

$$\sin\phi = \frac{v_0}{N}.$$

These equations give only the transient portion of the response following a step function which takes the system initially to (i_0, v_0). For steady-state or other forced solutions, the Laplace transform method can be used, as will be discussed in Section IV.

So far, all our trajectories have remained within a single region. However, when a boundary of a region is reached, the current and voltage at that point (referred to the origin of the new region) are used as new initial values and the path plotted on into the new region in the manner just described.

STABLE NODE

INSTABLE FOCUS

COMBINATION LIMITING OSCILLATION
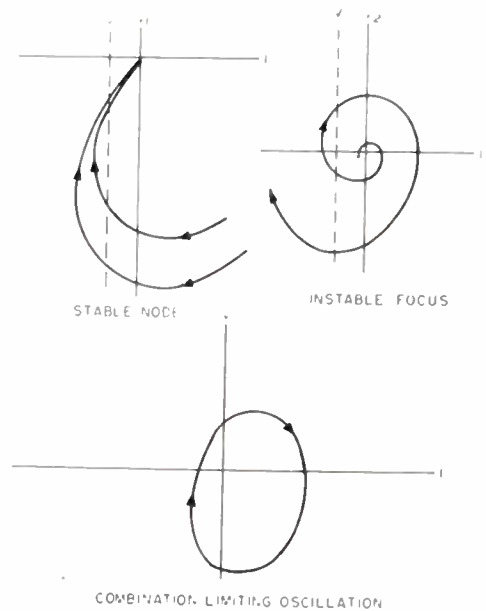
Fig. 4—Combined node and focus resulting in oscillation.

Fig. 4 is intended to show how two regions are combined in a simple case of a type which can arise in transistor circuits. The bottom picture represents the v-i plane of the emitter, with cutoff to the left, and active region to the right. To the right of the origin the system behaves like the upper-right diagram, having unstable spirals with (unstable) equilibrium point as shown. To the left of the origin in the bottom picture the system behavior is like that at the upper left, with stable node paths also having an origin to the right of the dotted line. These two planes are then combined so that the dotted lines coincide. In this case, a path may then be found in the bottom plane so that a portion of a spiral is followed to the right of the origin, and a portion of a node to the left, with the end-points coinciding on the v-axis. Oscillation around the total path will then take place. Such a path is called a "limit cycle."

Conditions under which a limit cycle exists are discussed in the references cited. In this case, physical reasoning makes its existence plausible, since the increasing transient on the right due to gain in this region is balanced by the decreasing one on the left and stable oscillations may be expected.

Fig. 5 gives an illustrative case to show how such an analysis applies to a circuit with two stable points, in the two-dimensional case. This applies in a general way
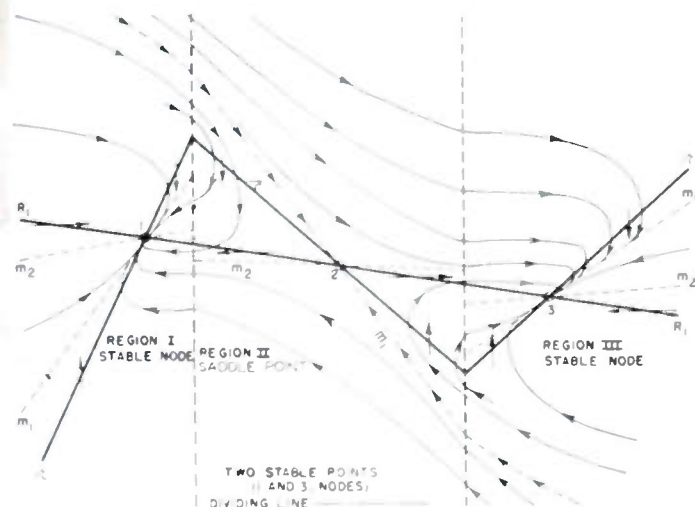
Fig. 5—Essential features of transient behavior of typical nonlinear negative-resistance circuit with two stable points. Assumptions: 1. Two reactive elements, 2. Linear in vertical strips separated by dashed lines, 3. Paths horizontal along R_1-line and vertical along r-line.

to the transistor circuit of Fig. 10. Three linear regions are shown, in each of which the resistance r is different, but constant within any one.

The load line R_1 is shown as constant in all three regions, but it need not be in the general case (for example, it might be diode controlled) and the other parameters of the circuit may also vary from region to region.

The paths in each region have been roughly sketched in with relation to the load and "negative-resistance lines, under the assumptions that the two outside points, 1 and 3, are stable nodes and that the middle one is a saddle point. This is a case frequently met in flip-flop circuits. Note that the plane is divided into two regions, from each of which the system ultimately arrives at one of the stable points. Furthermore, the stable point reached is not necessarily the closest one at the start. In a practical case, when C_1, the external capacitance, is reasonably large, and τ, the internal time constant of the device, is small, the m-lines may nearly coincide with the r and R_1 lines and the paths will be nearly horizontal except near the r-lines. This effect is discussed more fully in Section VII.

It is of interest to discuss the operation when the R_1 line is suddenly lifted over the "peak" on application of a trigger voltage. There remains only one stable point, number 3, and paths can then be drawn by the methods described above which show that the system will pass over the peak and arrive at the stable point.

In Fig. 6, note that if the trigger reverts the system to Fig. 5 before the point has crossed the dividing line, the point will return to 1 instead of going on to 3.

In a manner similar to those briefly discussed, the operation of "negative-resistance" switching circuits can be quickly graphed to get a qualitative picture of the action, and can be calculated accurately if desired.

It is clear that an analysis of the type presented can be applied to systems with more degrees of freedom by using as many independent variables as necessary. The
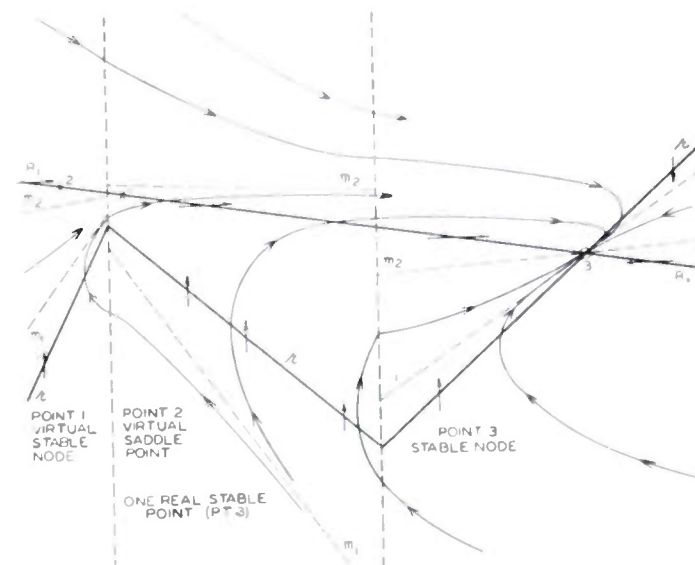
Fig. 6—The circuit of Fig. 5 with R_1-line raised by input trigger, leaving only one stable point.

system can still be considered linear in certain regions, and solutions patched together at the borders. Unfortunately, of course, a multidimensional space instead of a simple plane picture of the operation must then be used. However, a plane picture can be shown with many trajectories passing through every point, each trajectory corresponding to different initial conditions of the remaining nonplanar variables. Several planes may be used also.

III. Approximation to Transistor Input Impedance

In order to apply the method of analysis presented in Section II to the transistor, it is necessary to decide what parameters to use, what regions can be considered linear to good approximation, and what network behavior and constants to assign in each region. The details of what follows will be restricted to point-contact units, although junction units may be treated similarly.

Experimentally, clues to the linear regions are found in the behavior of the static characteristics of typical transistors. It is well known that when the emitter of an n-type transistor is made negative it has little or no effect on the collector current, which is then a minimum in magnitude. In this state the unit is said to be in collector current cutoff.[5] Likewise, when the emitter cur-

---

[5] It has been pointed out that from a duality standpoint there is a certain consistency in speaking of this state as collector voltage saturation. Since we will be concerned primarily with current as a variable, however, the above terminology will be retained.

rent is greater than a certain amount, depending on collector load, the unit is in a current-saturation, or overload, condition. In between, the unit is in its active state. These facts give rise to a dc emitter input characteristic of the circuit in Fig. 7, roughly as shown in Fig. 8(a), where an external base resistance R_b is included. The current scale to the left of the peak is greatly exaggerated for clarity. This figure suggests the linearized characteristic shown in Fig. 8(b). That is, as a first approximation, the three linear regions will be

TRANSISTOR WITH EXTERNAL BASE RESISTANCE

i_1

$V = R_{11}i_1 + R_{12}i_2$

$V_2 = R_{21}i_1 + R_{22}i_2$

"BLACK BOX" REPRESENTATION

$R_{ij} = r_{ij} + R_b$
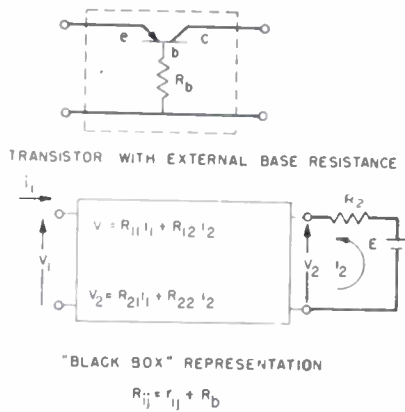
Fig. 7—Four-pole box containing transistor and external base resistance.
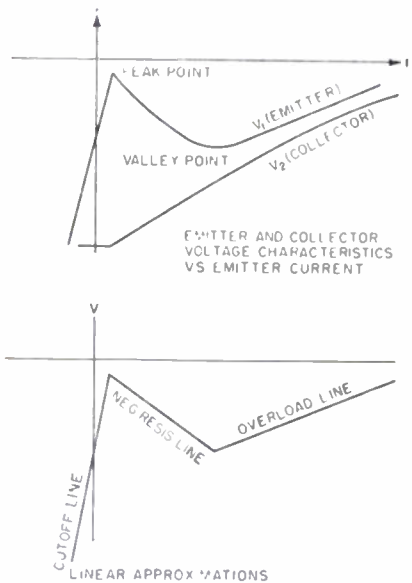
Fig. 8—DC characteristic of circuit of Fig. 7 and its linear approximation.

separated by vertical lines through the "peak" and "valley" points, and the emitter voltage and current will be employed as parameters. This approximation is tentative, at present, subject to additional experimental evidence. More discussion of this point is given in Section VIII. Motions will, therefore, appear as traces in the emitter current-voltage plane.

Now assumptions must be made regarding the nature of the constants in the 4-terminal box in each region. It would suffice to know the 4-terminal impedances Z_{11}, Z_{12}, Z_{21}, and Z_{22}. If the transistor is regarded as a 4-terminal network, the four low-frequency, small-signal parameters of the transistors, r_{11}, r_{12}, r_{21}, and r_{22},[6] can be measured over each region and a value of each r averaged over each region can be used. Knowledge of the variation of the Z's with frequency is all that is left. A first approximation which has been found to give reasonable answers results from consideration of the variation of current gain with frequency. Many transistors show a 6 db per octave asymptotic fall-off of current gain with frequency. This suggests that r_m/r_c[6] acts like a resistor and capacitor in parallel. Our procedure consists of assuming that all the variation with frequency resides in r_m alone, specifically that

$$z_m = \frac{r_m}{1 + i\omega\tau}, \qquad (10)$$

where $\tau = 1/2\pi f_c$, f_c being 3-db point of the current gain versus frequency.

This assumption is equivalent to assuming r_m is shunted by a condenser C of such magnitude that the time constant $r_m C = \tau$.

Since $Z_{21} = z_m + R_{12}$,

$$Z_{21} = \frac{r_m}{1 + j\omega\tau} + R_{12} = \frac{R_{12}j\omega\tau + R_{21}}{1 + j\omega\tau}, \qquad (11)$$

where R_{21} is the dc value of Z_{21}.

It is of interest to compute the input impedance looking into the emitter on this basis. It turns out to be

$$Z_{in} = \frac{r + \omega^2\tau^2 r_1 - j\omega\tau r_2}{1 + \omega^2\tau^2}, \qquad (12)$$

the definitions r, r_1, and r_2 given later in (17).

This input impedance is exactly given also by either of the two equivalent circuits shown in Fig. 9. Note that from the definitions r_2 is always negative. This makes
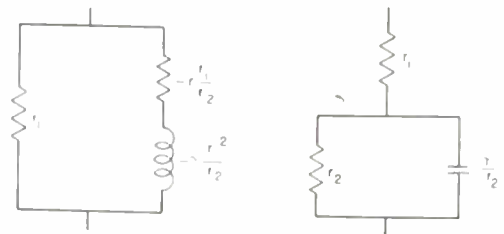
Fig. 9—Equivalent circuits of transistor at emitter terminals, with collector load and external base resistance.

the capacitance negative and the inductance positive. Also, r is the dc input resistance of the circuit, and, therefore, the slope of the static curves in Fig. 8.[7]

[6] For a definition of the r's used here, see W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., New York, N. Y., p. 40; 1950.

[7] Note that negative resistance defined here is "pure" in that no change with frequency is allowed. Frequency dependence is taken care of by additional elements like inductance or capacitance in Fig. 9.

IV. A Basic Transistor Circuit

We are now in a position to apply our analysis to a particular circuit of considerable interest in transistor applications. This is the circuit shown in Fig. 10. It is the basic type used in a number of high-speed switching applications.[8] We will see in Section V that it is also the practical equivalent of the basic circuit used for measuring current gain, in which case R_2 is small.
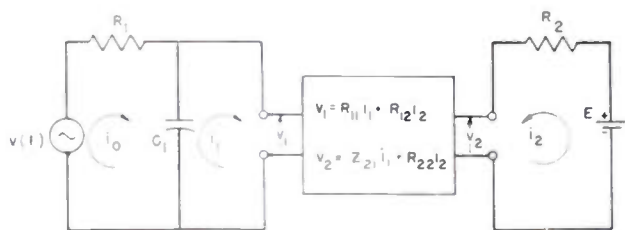
Fig. 10—Basic circuit. Transistor including external base resistor.

We will take R_2 to be small enough that any output capacitance can be neglected. R_1, however, is often quite large, and C_1 must be included. It adds the second reactive element to that within the transistor whose approximate nature has just been discussed in Section III.

We may be interested in the behavior of this circuit when $v(t)$ is impressed.

There are several equivalent approaches. We have already given in (8) and (9) a general solution for the unforced transients in the two-dimensional case, which corresponds to $v(t)$ constant. However, A, B, and the k's must be computed for our particular case. The k's will be computed in Section VII as this is a convenient method of finding the transients in terms of i_0 and v_0 when all applied voltages are constant.

For the general case, we will write the circuit equations and their Laplace transforms to indicate how solutions may be found for any form of applied trigger. If the trigger is applied in a different way from $v(t)$, the equation will be different, of course, but the method is the same. In our case, the equations will be solved only for the steady-state response to an applied sine wave $v(t)$ to show the variation of current gain with frequency and to compute the constants A and B.

If it is desired to work chiefly with the emitter voltage and current, which is often more convenient, either equivalent circuit of the emitter discussed will suffice.

The mesh equations for the circuit of Fig. 10 in schematic layout are given below:

$$
\begin{array}{ccc}
i_0 & i_1 & i_2 \\
R_1 + \dfrac{1}{j\omega C_1} & -\dfrac{1}{j\omega C_1} & 0 & = v(t), \\[2ex]
-\dfrac{1}{j\omega C_1} & R_{11} + \dfrac{1}{j\omega C_1} & R_{12} & = 0, \\[2ex]
0 & \dfrac{R_{12}j\omega\tau + R_{21}}{1 + j\omega\tau} & R_2 + R_{22} & = E.
\end{array}
\tag{13}
$$

[8] See for example A. E. Anderson, "Transistors in switching circuits," Proc. I.R.E., vol. 40, pp. 1541–1558; this issue.
J. H. Felker, "Regenerative amplifier for digital computer applications," Proc. I.R.E., vol. 40, pp. 1584–1597; this issue.

The Laplace transforms are

$$
\begin{array}{ccc}
I_0 & I_1 & I_2 \\
R_1 + \dfrac{1}{C_1 s} & -\dfrac{1}{C_1 s} & 0 & = V(s) - \dfrac{v_1(0)}{s}, \\[2ex]
-\dfrac{1}{C_1 s} & R_{11} + \dfrac{1}{C_1 s} & R_{12} & = \dfrac{v_1(0)}{s}, \\[2ex]
0 & \dfrac{R_{12}\tau s + R_{21}}{1 + \tau s} & R_2 + R_{22} & = \dfrac{\tau v_{a0}}{1 + \tau s} + \dfrac{E}{s},
\end{array}
\tag{14}
$$

where $L[i(t)] \equiv I(s)$, and so on, and v_{a0} is the initial voltage on the capacitor across r_m. It can be expressed in terms of $v_1(0)$ and $i_1(0)$.

$$
v_{a0} = E - (v_{10} - R_{11}i_{10})\frac{R_2 + R_{22}}{R_{12}} - i_{10}R_{12}. \tag{15}
$$

From these equations the response to any $v(t)$ can be calculated. The characteristic equation, $s^2 + As + B = 0$, is found by equating the determinant of the system to zero. It is found that

$$
A = \frac{\tau_1 r + \tau(r_1 + R_1)}{\tau\tau_1 r_1},
$$

$$
B = \frac{R_1 + r}{\tau\tau_1 r_1},
\tag{16}
$$

where

$$
r = \frac{\begin{vmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{vmatrix} + R_2 R_{11}}{R_2 + R_{22}},
$$

$$
r_1 = \frac{\begin{vmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{vmatrix} + R_2 R_{11}}{R_2 + R_{22}},
\tag{17}
$$

$$
r_2 = -\frac{r_m R_{12}}{R_2 + R_{22}},
$$

$$
\tau_1 = R_1 C_1,
$$

$$
R_{11} = r_{11} + R_b, \text{ and so on.}
$$

Note that $r = r_1 + r_2$ and that r_2 is always negative. For computation, the following equivalent forms are often more convenient:

$$
r = \frac{r_{11} - \alpha r_{12} + R_b\left(1 - \alpha + \dfrac{r_{11} - r_{12}}{r_{22}}\right) + \dfrac{R_2(r_{11} + R_b)}{r_{22}}}{1 + \dfrac{R_b}{r_{22}} + \dfrac{R_2}{r_{22}}},
$$

$$
r_1 = \frac{r_{11} - \dfrac{r_{12}^2}{r_{22}} + R_b\left(1 + \dfrac{r_{11} - 2r_{12}}{r_{22}}\right) + \dfrac{R_2(r_{11} + R_b)}{r_{22}}}{1 + \dfrac{R_b}{r_{22}} + \dfrac{R_2}{r_{22}}}.
\tag{18}
$$

It is often the case that some or all of the terms with r_{22} in the denominator can be neglected. We have already mentioned that r is the emitter input resistance.

V. Current Gain versus Frequency

In order to show that our approximation to the frequency response gives an experimentally reasonable result, we will compute the current gain versus frequency as measured in a practically realizable circuit arrangement.

If the ac part of the response $i_2(t)$ to $v(t)$ is calculated from (14) in the usual manner, we find

$$I_2(s) = -\frac{V(s)(R_{12}\tau s + R_{21})}{(R_2 + R_{22})\tau\tau_1 r_1(s^2 + As + B)}; \qquad (19)$$

if $v(t) = V_1 e^{j\omega t}$, this gives

$$\frac{|I_2|^2}{|V_1|^2} = \frac{R_{21}^2 + \omega^2\tau^2 R_{12}^2}{[(R_2 + R_{22})\tau\tau_1 r_1]^2[(B - \omega^2)^2 + \omega^2 A^2]}. \qquad (20)$$

When measuring the current gain, it is desirable that R_1 be made as large as possible to give constant input current, R_2 be small in comparison with r_{22}, $R_b = 0$, and $C_1 = 0$. Under these conditions $I_1 = V_1/R_1$, and the above yields

$$\frac{|I_2|}{|I_1|} = \frac{r_{21}}{r_{22}(1 + \omega^2\tau^2)^{1/2}} = \frac{\alpha}{(1 + \omega^2\tau^2)^{1/2}}. \qquad (21)$$

This shows that our assumption gives the required variation of apparent α, current gain, with frequency. However, because of unavoidable stray C_1 and the like, it is often very difficult to meet the required conditions given above experimentally, especially when units with high-frequency response or base resistance are under test. In such a case, (20) shows that a complicated curve can arise, having a peak apparent α near $\omega = \sqrt{B}$. Such a peak is actually frequently observed with high-frequency units, especially if biased near oscillation. We have seen that $A = 0$ is the borderline condition for oscillation, when $B > 0$, as it is here (Fig. 3). Equation (20) shows that the peak climbs as A approaches zero, and this behavior is also observed.

We have also seen in (9) that the circular frequency of self-oscillations is given by $\sqrt{|\Delta|}/2$. When A is small, this frequency is simply $\sqrt{B}$, so that the transient oscillates at the frequency of the peak. This action is observed experimentally also. All this is to be expected, and from the point of view of the emitter equivalent circuit, merely amounts to saying that a resonance occurs between the external capacitance, C_1, and the internal equivalent inductance at a certain frequency, and that the circuit rings at this frequency.

VI. Circuit Stability

A few remarks about stability are pertinent at this point. Fig. 3 shows that if $B < 0$, a "saddle-point" type of path family always results. In our case this condition is equivalent to $R_1 + r < 0$, or $r < -R_1$. A picture similar to Fig. 3 results. This condition is what is usually called "dc unstable." If, on the other hand, $R_1 + r > 0$, then the type of path family depends only on A. The borderline stability case, $A = 0$, gives us

$$\tau_1 r + \tau(r_1 + R_1) = 0 \qquad (22)$$

by (16). If, furthermore, $R_1 \gg r_1$, which is often the case, (22) reduces to simply

$$\tau = -rC_1. \qquad (23)$$

This gives the relation between internal and external parameters necessary for the borderline stability. Since $\tau > 0$, oscillation cannot take place unless $r < 0$, as is to be expected.

Equation (23) has been tested experimentally in conjunction with J. R. Harris. As an example, unit EK362 was biased to borderline stability in two cases with different C_1; f_c, r, and C_1 were measured, and C_1 calculated from (23). The results were as follows:

I_e ma	V_c v	f_c mc	r Ω	C_1 (meas) μμf	C_1 (calc) μμf
2.3	5.8	50.0	−169	21	19
1.4	4.0	32.4	−138	41	36
			diff	20	17

In view of the experimental difficulties, these results are considered a reasonably satisfactory check on the assumptions of Section III for this unit.

VII. "Negative-Resistance" Relaxation Oscillations

In this section the operation of the circuit of Fig. 10 as a simple oscillator of the type known as a "negative-resistance" relaxation oscillator will be discussed. It will be seen how the familiar simplified explanation of such an oscillator when C_1 is large is contained in the general solution as a special case.

As has been mentioned, the circuit of Fig. 10 has a dc input characteristic as shown in Fig. 8, which is usually called an "N-type" negative resistance. We will discuss this circuit in the case when $|R_1| > |r|$ and the load line of R_1 crosses the dc characteristic in the negative-resistance part, like line A in Fig. 13, for example.

If the external capacitance C_1 is large,[9] such a circuit will usually exhibit oscillations of the "relaxation" type. The usual qualitative explanation of the oscillations states that the current-voltage point of the system moves up the left leg of the "N" until it reaches the peak, then snaps quickly over to the right leg, moves down that side to the valley, snaps to the left side again, and so on. The speed of motion on the legs is determined by the time constant of the external capacitance in parallel with both load and input resistance.

This description is adequate for the case of large external capacitance, and gives good results when the "snap" time is small compared with that on the legs. However, an explanation for such behavior and information as to speed and path of the snap transition are needed. This is because the "snap" time becomes comparable with the rest of the cycle as the external capacitance is lowered in high-speed switching circuits. Furthermore, it is possible to have a condition in which no spontaneous oscillations take place at all, because the intersection of the load line of R_1 and the negative-resistance characteristic determine a stable point.

---

[9] This is the case treated by G. E. McDuffie, "Pulse duration and repetition rate of a transistor multivibrator," Proc. I.R.E., vol. 40, pp. 1487–1490; this issue.

Our analysis answers all these questions, under the assumptions made in Section III.

The question of stability has already been answered in Section VI by (22) and (23). If $r < 0$ but τ is large enough (or τ_1 small enough) to make the left side of (22) (which is the numerator of A) positive, no spontaneous oscillations can take place.

In order to examine the trajectories, (3) is needed for our circuit.

To calculate (3) for our case, the mesh equations (13) of the circuit and then the nodal equations can be differentiated and the result solved for di_1/dt and dv_1/dt, or one of the equivalent circuits can be used somewhat more conveniently. In either case we find

$$\frac{di_1}{dt} = \frac{\tau_1 - \tau}{\tau\tau_1 r_1} v_1 - \frac{\tau R_1 + \tau_1 r}{\tau\tau_1 r_1} i_1,$$

$$\frac{dv_1}{dt} = -\frac{1}{\tau_1}(v_1 + R_1 i_1), \tag{24}$$

so that

$$k_{11} = \frac{\tau_1 - \tau}{\tau\tau_1 r_1}, \qquad k_{12} = -\frac{\tau R_1 + \tau_1 r}{\tau\tau_1 r_1},$$

$$k_{21} = -\frac{1}{\tau_1}, \qquad k_{22} = -\frac{R_1}{\tau_1} = -\frac{1}{C_1}. \tag{25}$$

In (24) the constant term has been eliminated by transforming the origin of (i, v) co-ordinates to the intersection of the load line of R_1 and the r-line, that is, to the dc point of equilibrium. This amounts, of course, to considering the ac part of the motion, which takes place around the dc equilibrium point. We have done this throughout.

It can be seen from (24) that in the circuit under discussion, trajectories are always horizontal ($dv_1/di_1 = 0$) when crossing the load line $v_1 = -R_1 i_1$, and that they are vertical $\left(\dfrac{di_1}{dv_1} = 0\right)$ when crossing

$$v_1 = \frac{\tau R_1 + \tau_1 r}{\tau_1 - \tau} i_1.$$

If τ_1 is large compared with τ, because of large C_1, the last equation reads approximately $v_1 = r i_1$. Also, as $C_1 \to \infty$, the field of paths tends to become horizontal everywhere, except near $v_1 = r i_1$, since by (24) $dv_1/di_1 \to 0$ unless $v_1 = r i_1$.

If this is investigated a little further, we can see how our general solution degenerates when $C_1 \to \infty$, to the intuitive picture of the current "snap" over to the r-line. From (4) it can easily be shown that

$$s_1 \to -\frac{B}{A}$$

$$s_2 \to -A + \frac{B}{A} \tag{26}$$

as $(4B/A^2) \to 0$. (If $A < 0$, interchange 1 and 2.)

As $C_1 \to \infty$, $B \to 0$, and $A \to (r/\tau r_1) < 0$ if $r < 0$ by (16). Therefore $s_2 \to 0$ and $s_1 \to -(r/\tau r_1)$. Thus by Fig. 3 the

trajectories become unstable nodes with nearly horizontal parts except near the r-line. If $r > 0$, the same holds except that we have stable node paths with $s_1 \to 0$. We have seen in Section II that stable node paths approach the origin with slope m_1. Since $s_1 \to 0$ in the stable case (when $r > 0$), (7) shows that $m_1 \to r$. Thus each path moves horizontally until it approaches the r-line, turns rapidly, and approaches the origin with slope r.

In either case, the dominant time constant is $-\tau r_1/r$ which is about τ if r/r_1 is around unity (a typical case, although it can be considerably less).

This can be seen in a more illuminating way. When $C_1 \to \infty$, the paths are nearly horizontal, v_1 is nearly constant, and (24) gives

$$\frac{di_1}{dt} + \frac{r}{\tau r_1} i_1 = \text{const.},$$

that is,

$$i_1 = i_{10} e^{-(r/\tau r_1)t} + \text{const.} \tag{27}$$

Thus during a *current* change, the time constant is dominated by the transistor. While, however, the system moves along the r-line, $v_1 = r i_1$, and (24) gives

$$\frac{dv_1}{dt} = -\frac{v_1}{R_1 C_1}\left(1 + \frac{R_1}{r}\right), \quad \text{or} \quad v_1 = v_{10} e^{-t/RC_1}, \tag{28}$$

where R is the parallel combination of R_1 and r. Thus, along the r-line, the time constant is dominated by the external capacitance, that is, the circuit degenerates to a simple resistance-capacitance network.

We have seen in the foregoing how our general equations agree with the familiar qualitative description in the case of large C_1, and give the time constants during snap transitions. To find the exact times, $i(t)$ and $v(t)$ can be calculated by means of (8) or (9). This will be illustrated in the next section for a high-speed case in which the approximation $C_1 \to \infty$ cannot be made.

While the emitter voltage and current are following a given trajectory, we may wish to compute the output at the collector. Fig. 10 shows that

$$i_2 = \frac{v_1 - R_{11} i_1}{R_{12}},$$

$$v_2 = E - i_2 R_2 = E - \frac{R_2}{R_{12}}(v_1 - R_{11} i_1), \tag{29}$$

which gives v_2, i_2 in terms of v_1 and i_1 at any instant. Under our assumptions, R_{11} and R_{12} are independent of frequency, so that these equations always hold. However, it must be remembered that R_{11} and R_{12} must be average values over the region of interest in the large signal case, just as in the case of all our parameters. In particular, for a transistor, r_{11} and r_{12} change so rapidly in the neighborhood of $i_1 = 0$ that care must be exercised in using the equations there.

VIII. APPLICATION TO A RISE-TIME AND WAVEFORM CALCULATION

As an illustration of the application of the methods discussed, they will be applied to a calculation of the

waveforms and transition times of a regenerative amplifier to which Fig. 10 approximates. This amplifier has been used by Felker in high-speed computing applications, and has been described by him.[10]

It is believed that, although the regions to which a given parameter belongs are not yet known in every case, particularly during the return time when minority carrier storage is effective, it will be seen that the method developed affords a useful tool for understanding and designing circuits essentially dependent on nonlinear device properties.

Since the details of operation of the regenerative amplifier in its system application have been discussed by Felker, we will confine ourselves here to briefly giving only the facts pertinent to our calculation.

For our purposes, the circuit operation can be understood by reference to Figs. 11, 12, and 13. Fig. 11 shows the relative timing of the input pulse applied to the emitter and the reset clock pulse applied to the base of the transistor. Both pulses are here assumed to be square waves. It will be shown neglecting transition times of signals leads to a very good approximation.
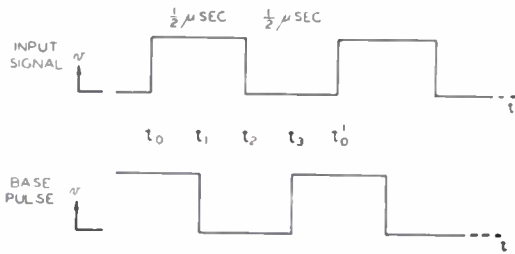
Fig. 11—Relative timing of input and base waves.

The base pulse is applied through a diode, and its chief effect from our point of view is to raise and lower the entire dc emitter characteristic of the transistor as
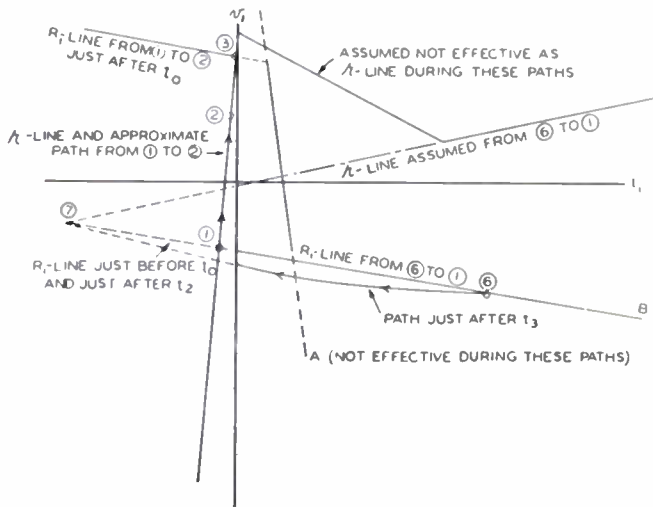
Fig. 12—Illustrates r and R_1-line configurations and paths between t_0 and t_1, and just after t_3. Assumptions: (1) dc emitter input characteristic (r-lines) is controlled by base wave. (2) R_1-line configuration is controlled by the emitter input wave, but the R_1-segment in use at any instant is controlled by the emitter voltage at that instant.

[10] J. H. Felker, "Regenerative amplifier for digital computer applications," PROC. I.R.E., vol. 40, pp. 1584–1597; this issue.

the pulse rises and falls. However, no external base resistance is effective during the base pulse and the effective supply voltage is raised in magnitude so that the parameters of the unit will be somewhat different during this time. The fall of the emitter characteristic is shown in going from Fig. 12 to Fig. 13, and the rise in going from Fig. 13 to Fig. 12.
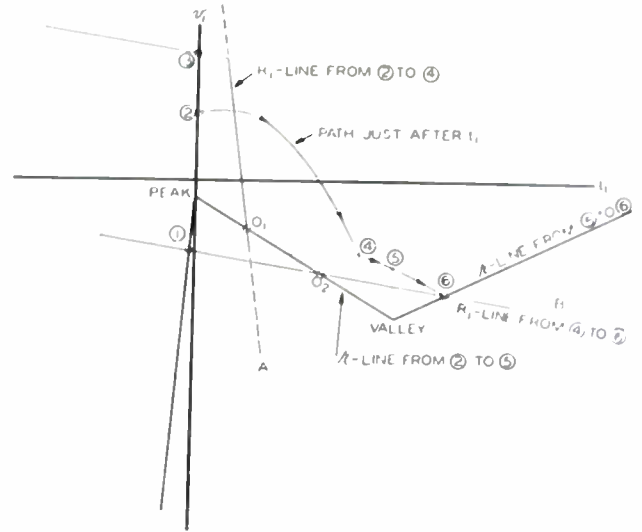
Fig. 13—Illustrates r and R_1-line configurations and paths just after t_1.

The input pulse is applied through diodes in such a way as to shift the load line (R_1 in Fig. 10) from a low value, (B), to a high one, (A), and back again.

In these terms, the action is as follows with some simplifying assumptions. Initially, just before t_0 (Fig. 11), the emitter voltage and current are at the stable point (1), Fig. 12. At t_0, the input pulse suddenly establishes a new load line passing through point (3). The emitter point starts upward towards the new stable point (3). During this motion there is no effect on the output, since the transistor is in cutoff; this means the rise of the input pulse has little effect on the action. At t_1, the base pulse suddenly shifts the dc characteristic to its position as shown in Fig. 13. The emitter point may not have nearly reached (3) at this time, depending on the timing and circuit constants. The picture is drawn as if it had traveled only part way, namely, to (2). Motion now takes place which ends at the stable point (6). At t_3, the base pulse suddenly shifts the dc characteristic upward again as in Fig. 12. The emitter point, initially at (6), then follows a path toward the stable point (7), stopping when cutoff is reached, and ending at (1) to begin again.

The motion is therefore divided into three parts, each of which must be calculated, although during the first there is no output. These three parts are between points (1) and (2), (2) and (6), and (6) and (1).

Between (1) and (2) the transistor is in its current-cutoff condition, and the r effective here is the back resistance of the emitter. Thus, a sufficient approximation is to consider the circuit as a simple RC network where

R is the parallel combination of R_1 and r, as in Section VII. Hence

$$v_1 = v_{10}e^{-t/RC_1},$$

where the origin is at point (3) and r_{10} is the voltage of (1) referred to this. In our numerical case, point (2) was chosen at emitter current and voltage zero, which is a good enough approximation for our purpose. This becomes the new v_{10} to be used at point (2) for entry into the next phase of the motion. Here the emitter current has remained essentially zero throughout. As mentioned above, no output has appeared as yet.

We are now ready to carry the motion into the active region of the transistor and discuss the path between (2) and (6) in detail. Note that in this circuit, the input pulse merely sets the stage and the circuit is then allowed to carry on by itself. Therefore, (8) and (9) may be used to compute the motion, rather than (14), since no $v(t)$ is applied during this motion.

Because of the external diode action, line A will be effective as the load line R_1 as long as the emitter voltage does not rise above the voltage of point (3), or below the voltage of point (1). When the voltage of point (1) is reached, line B becomes effective as R_1 in our equations. We then have a new path based on the active region intersection of B and the dc characteristic, 0_2, as origin. This path continues to point (5), vertically above the valley, at which place we assume the r and the τ of the transistor change.

The circuit under discussion was developed for use with the M-1734 high-speed switching transistor. At present, its parameters in the active region are as follows:

	Approx. Mean	Approx. Range
r_{11}	700 Ω	600–800Ω
r_{12}	500	400–600
r_{21}	44 K	35 K–55 K
r_{22}	20 K	15 K–25 K
f_c	20 mc/sec	10–50 mc/sec

This give an average $\alpha = 2.2$.

Using these mean values in (18), we find $r = -1000\Omega$, $r_1 = 1200\Omega$, and $\tau = 8 \times 10^{-9}$ sec. The co-ordinates of the important points are as follows, in terms of the emitter voltage and current:

	Current ma.	Voltage v
0_1	0.34	−0.84
0_2	0.56	−1.1
(6)	6.8	−1.7
(7)	−5.3	−0.47
(1)	0.0	−0.3
Peak	0.0	−0.5
Valley	2.8	−3.3

The external parameters for the motion between (2) and (4) are $R_1 = 20K$, $R_i = 500\Omega$, $R_2 = 500\Omega$, and $C_1 = 2 \times 10^{-11}$ farad, giving $\tau_1 = 4 \times 10^{-7}$ sec. Here $v_{10} = 0.84$ volt, and $i_{10} = -0.34$ ma, the voltage and current of point (2) referred to 0_1. With these data, we find $A = -6.3 \times 10^7$ an $B = 5.2 \times 10^{15}$ from (16), $\Delta = -1.7 \times 10^{16}$ by (4). Using (9), we calculate the path, and find point (4) is reached in 0.022 μsec. These and subsequent calculations quoted were carried out by R. D. Padgitt.

For the motion between (4) and (5), v_{10} and i_{10} were found to be 0.5 volt and 1.9 ma, and $R_1 = 100$. The time between (4) and (5) was found to be about 1.7×10^{-3} μsec.

We are now ready to continue to point (6). Point (6) is a stable point in this region, since $r > 0$ during this motion, and therefore both $A > 0$, and $B > 0$. In our illustrative calculation, τ was raised to 3.2×10^{-8} sec. although the best value is not yet known. Appropriate values of r and r_1 are $r = 410\Omega$, $r_1 = 410\Omega$, giving $A = 6.5 \times 10^8$, $B = 1.9 \times 10^{16}$, $\Delta = 3.5 \times 10^{17}$, and (8) shows v_1 and i_1 approach (6) exponentially. The point arrives within 0.1 ma of (6) from (5) in 0.1 μsec. Ninety per cent of the total current change of 6.8 ma from (2) to (6) takes place in about 0.065 μsec.

Our emitter now remains at (6) until t_3, when the base pulse produces the picture and subsequent motion shown in Fig. 12. In this case, of course, (7) is used as an origin with initial co-ordinates at (6). Parameters just the same as those used between (5) and (6), except $r = r_1 = 250\Omega$ to account for zero R_b, give a time of 0.04 μsec to arrive at the cutoff region (emitter current zero). The voltage is then so close to (1) that this is considered the end of the motion calculated to our degree of approximation.

The entire emitter cycle has now been completed. If it is desired to compute the output at the collector, (29) can be used. The R_{11} and R_{12} to be used are, of course, the average values within any region just as our other parameters. There is also a rise in collector voltage at t_3, resulting from the base pulse. (No such change occurs at t_1 since the unit is in cutoff.) This change can be calculated approximately by well-known dc methods. Since nothing new is involved, these cal-
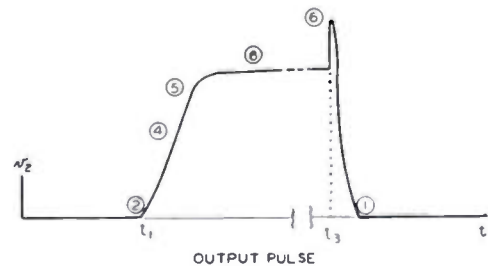
Fig. 14—Qualitative diagram of collector output pulse showing corresponding emitter points.

culations have not been made numerically. However, qualitatively, the output voltage wave according to our approximations looks like Fig. 14, which shows the points of Figs. 12 and 13 at the corresponding times. The rise time for 90 per cent of full change will be approximately the same as that of the emitter current, i.e., about 0.065 μsec.

Comparing this picture with experience, the calculated rise time is somewhat longer than usual, mostly due to the time between (5) and (6), which suggests that τ may remain low between those points, instead of getting larger, as we assumed.

The tail of the pulse, however, differs from the most often observed case in that v_2 as observed usually remains relatively constant after t_3 for a period up to 0.1 μsec before it decreases. It is during this time, with the collector in saturation, that the phenomenon of minority carrier storage is most apparent since the voltage across the collector barrier is small, and this voltage cannot change until the stored carriers are removed from the body of the semiconductor. This presumably causes the period of slow collector voltage change referred to. The calculated fall time of 0.04 μsec is also too short. However, we have seen how a nonlinear circuit can be approximated by changing linear parameters and circuits at suitable currents and voltages, and such changes can also be made during certain periods of time. It is not yet known whether the storage effect can be accounted for by a suitable choice of parameters in our circuit, or if a circuit with different frequency dependence should be chosen during the storage time.

Our conclusion is that the calculation of the output wave results in a rise which is reasonably correct quantitatively, but that accurate calculation of the fall requires further study. It is believed that inaccuracies in both cases are due to insufficient knowledge of values of transistor parameters and the regions or times during which given average values are effective. Further experimental work along these lines should make it possible to use the methods described to predict the complete performance of nonlinear transistor or other circuits to satisfactory engineering requirements.

Control Applications of the Transistor*

E. F. W. ALEXANDERSON†, MEMBER, IRE

Summary—This paper explores the possibilities of using the transistor for control functions which are, in present-day practice, performed by magnetic amplifiers, amplidynes, and thyratrons. It is found that a main transistor controlled by auxiliary transistors can be made to perform in a way resembling the phase-controlled rectifier. It can be energized from an ac power supply and it can be operated with an efficiency which approaches the rectifier. In several ways it is superior to the magnetic amplifier: It can carry current in both directions and it can serve as a two-way rectifier or a frequency changer. It is hoped that these possibilities will lead to development of transistors in larger units.

INTRODUCTION

THE PUBLICATIONS of the Bell Telephone Laboratories are the source of information for this paper, particularly the writings of Shockley, Sparks, and Teal on the *p-n* junction transistor. Deductions from these discoveries and theories can be made which lead to the design of transistor circuits based on analogy with vacuum-tube circuits. Examples of such deductions are contained in papers on transistor circuits by Wallace, Raisbeck, and Pietenpol. The model in this paper is the phase-controlled rectifier.

ANALOGY WITH THE CONTROLLED RECTIFIER

A question of more than theoretical interest is, can the transistor take the place of the controlled rectifier? In this connection it should be observed that a class A amplifier of either the electronic or the transistor type has an efficiency of power conversion which approaches 50 per cent as a theoretical maximum, whereas the rectifier has an efficiency which approaches 100 per cent and is often well over 90 per cent. The reason for the high efficiency of the rectifier is that it presents either an open circuit with voltage and no current or a short circuit with current and almost no voltage.

The efficiency of the transistor when used as a control device with ac power supply may be estimated by extrapolating data which has been published on the transistor. In the volume of reference material published by the Bell Telephone Laboratories in 1951 is an article by Wallace and Pietenpol which gives the characteristic curves of a transistor. Fig. 4 in that article shows the volt-ampere characteristics of the collector circuit at various values of emitter current. A load line indicates a typical variation between maximum resistance and minimum resistance. The extreme points on this load line may be assumed to be the operating points of a controlled transistor with ac power supply.

One of the extreme points shows 40 volts at approximately 0.1 ma. This corresponds to 400,000 ohms. The other extreme point on the load line shows 1 volt at 4 ma which represents 250 ohms. Thus there is a control range of 1,600 to 1.

If we now assume that the load current is 4 ma and the drop in the transistor 1 volt, as indicated by these characteristics, we may impress 40 volts and utilize 39 volts in the load circuit. Thus we have a load loss in the transistor of 2½ per cent. We may furthermore assume that this takes place only on the positive half-wave and that the negative half-wave is suppressed by giving the transistor its maximum resistance. We find then that there is an additional idling loss of 0.1 ma at 40 volts

which is also $2\frac{1}{2}$ per cent. Thus we may estimate an over-all efficiency of 95 per cent with 5-per cent loss in the transistor. This is analogous to a motor with $2\frac{1}{2}$-per cent load loss and $2\frac{1}{2}$-per cent core loss. While this is an extrapolation which may not be valid for transistors designed for higher power, the indications are highly promising.

TRANSISTOR WITH AC POWER SUPPLY

Fig. 1 shows the basic circuit for a transistor with ac power supply. The load is indicated as a resistance. The
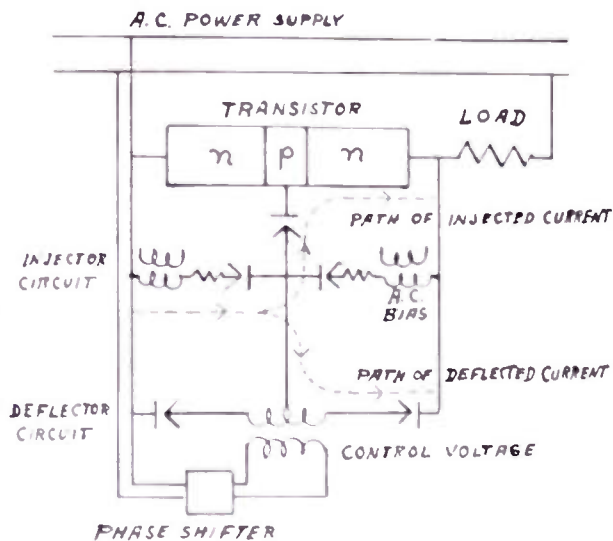
Fig. 1—Basic control circuit for transistor with ac power supply.

transistor which is in series with the load controls the current flow on both the positive and negative half-wave. The terminals of the *n-p-n* transistor therefore change places as collector and emitter on successive half-waves. Full current flow is delivered if the transistor is given minimum resistance from the beginning of each half cycle. On the other hand, the control may be so adjusted that the transistor exhibits maximum resistance during the first portion of each half cycle and changes to minimum resistance during the latter portion of the half cycle. The starting of the current flow is then retarded and the effective current flow is reduced. This is phase control in the accepted sense. But it is not controlled rectification. It is phase control in an ac circuit.

The control circuit which produces this effect is shown in Fig. 1. It has two distinct parts which will be referred to as injector circuit and deflector circuit. The desired effect is produced by the combined function of these two circuits.

The injector circuit is a rectifier fed from the terminals of the transistor. In addition to the transistor voltage an ac bias may be introduced by transformation as shown. The ac bias has the object of providing a power supply for the rectifier circuit even if the control of the transistor is so effective that the voltage drop between its terminals approaches zero. The ac bias may not be needed, and it has been omitted in Figs. 2 and 3.

The output of the injector circuit is a positive direct current which may be injected into the base terminal of the transistor in order to give it maximum conductivity. The transistor is then in condition to carry full current in both directions with minimum resistance loss.

The deflector circuit has the object of diverting the injector current so that it does not reach the base terminal of the transistor. The deflector circuit may be active or inactive, depending upon an ac control voltage introduced by transformation. When the deflector circuit is active, no current reaches the base terminal and the transistor exhibits maximum resistance.

The control voltage of the deflector circuit is the means for applying phase control. When the control voltage is in phase with the power voltage, the deflector circuit becomes inactive and the injector current is admitted to the base terminal of the transistor. Thus, the transistor is in condition to pass full current. If, on the other hand, the phase of the control voltage is retarded, the result is that the injector current is deflected and the transistor exhibits maximum resistance until a later moment when the control voltage passes through zero.

The functions of the control may be analyzed as shown in Tables I and II.

TABLE I

	Phase control advanced	
	Positive halfwave	Negative halfwave
Deflector circuit	Inactive	Inactive
Transitor resistance	Low	Low
Load current	Full	Full
Power flow	AC full load	

TABLE II

	Phase-control retarded			
	Positive halfwave		Negative halfwave	
	1st quarter	2nd quarter	3rd quarter	4th quarter
Deflector circuit	Active	Inactive	Active	Inactive
Transitor resistance	High	Low	High	Low
Load current	Low	High	Low	High
Power flow	AC reduced			

TWO SOURCES OF CONTROL

In the simple ac phase-controlled circuit described above the injector circuit is always active, but the current which it delivers may be injected into the base terminal or it may be deflected by the ac control voltage.

The usefulness of a transistor with an ac power supply can be greatly increased by providing a second source of control. The injector circuit lends itself well to

this. It is a rectifier which can readily be developed as a controlled rectifier. A pair of triodes may be used as controllable rectifier elements. The diagrams of Figs. 2 and 3, however, show auxiliary transistors which may be more desirable for two reasons. The transistor is a constant current device at any one emitter excitation. This is clearly shown by the published characteristic curves referred to above. An auxiliary transistor therefore lends itself well to delivering the desired injector current regardless of voltage fluctuations in the power supply to the injector circuit. A second reason for preferring the auxiliary transistor to a triode is its ruggedness and reliability.
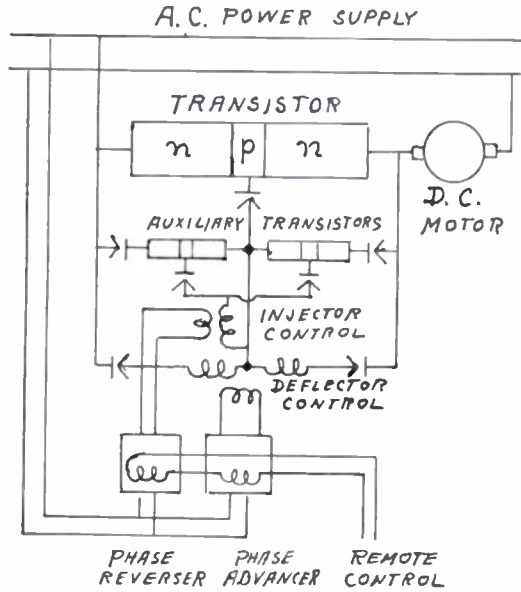
Fig. 2—Remote control of transistor for operating dc motor.
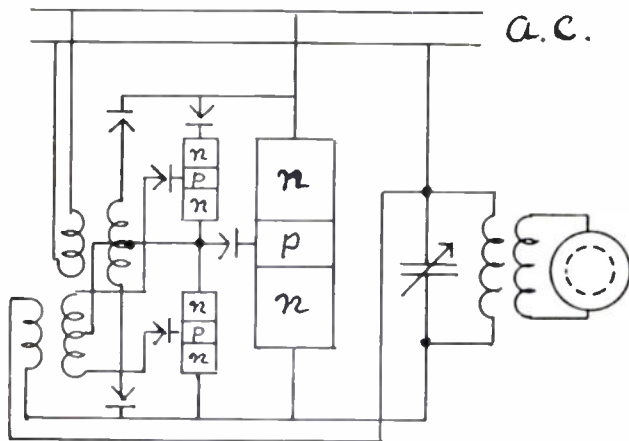
Fig. 3—Transistor as frequency changer operating induction motor.

Tables III, IV and V analyze the functioning of the two sources of control. It is desirable to operate a dc motor with an ac power supply and control the speed from standstill to a maximum in both directions of rotation. The control of the injector circuit changes the transistor into a rectifier and determines whether the output voltage should be positive or negative. This control will be referred to as polarity control. The injector current is always positive, but it becomes a polarity

control because the control delivers the injector current on the positive half-waves and suppresses it on the negative half-waves, or vice versa.

While the polarity control performs this selective function, the phase control determines how high the rectified voltage should be, and thereby determines the speed of the motor.

TABLE III

Motor control full speed forward

	Phase control advanced	
	Polarity control positive	
	Positive halfwave	Negative halfwave
Deflector circuit	Inactive	Inactive
Injector circuit	Active	Inactive
Transistor resistance	Low	High
Load current	Full positive	Low negative
Motor speed	Full forward	

TABLE IV

Motor control full speed reverse

	Phase control advanced	
	Polarity control negative	
	Positive halfwave	Negative halfwave
Deflector circuit	Inactive	Inactive
Injector circuit	Inactive	Active
Transistor resistance	High	Low
Load current	Low	High
Motor speed	Full reverse	

TABLE V

Motor control reduced speed forward

	Phase control retarded			
	Polarity control positive			
	Positive halfwave		Negative halfwave	
	1st quarter	2nd quarter	3d quarter	4th quarter
Deflector circuit	Active	Inactive	Active	Inactive
Injector circuit	Active	Active	Inactive	Inactive
Transistor resistance	High	Low	High	High
Load current	Low	Full	Low	Low
Motor speed	Reduced forward			

REMOTE CONTROL OF DC MOTOR

Fig. 2 illustrates an application of combined polarity and phase control. The problem is to control a dc motor in response to a remote signal in such a way that the motor runs forward at a speed proportional to the intensity of a positive signal and in reverse at a speed

proportional to the intensity of a negative signal. When there is no signal, the motor should be at standstill. Such a signal may be transmitted by radio.

The polarity control is used to determine the direction of rotation. When the signal current is positive, it releases a control of the injector circuit which produces a full injector current during the positive half cycles and blocks the injector current during the negative half cycles. A translating device is therefore needed which delivers an ac voltage in phase with the power voltage when the signal current is positive and a voltage of opposite phase when the signal current is negative. Such a translating device is indicated by a box symbol in Fig. 2. It may consist of a push-pull system of saturable reactors. This ac voltage is impressed upon the control circuit of the rectifier which delivers the injector current. The motor will thus receive power from the positive or negative half-waves of the power supply, depending on the positive or negative polarity of the remote signal.

The second requirement is that the motor speed should be proportioned to the intensity of the signal. This is done by phase control acting in the deflector circuit. The phase advance should be proportional to the intensity of the signal. This is accomplished by a well-known phase-shift device dependent upon magnetic saturation. The signal is used as saturating current, and has the effect of advancing the phase regardless of whether the signal current is positive or negative. In Fig. 2 this translating device is represented by another box symbol.

The two translating devices for the remote signal indicated on Fig. 2 are labelled "phase advancer" and "phase reverser." The first advances the phase in the deflector circuit regardless of the polarity of the signal. The second reverses the ac control voltage of the injector circuit in response to the polarity of the signal.

FEEDBACK CONTROLS

Motor control by transistor is adapted to feedback systems such as positioning controls. A position signal is brought back from the motor-driven apparatus and matched against the master control. An error of position then generates a signal which is fed into the remote-control circuit.

This system is also adapted to positioning control of an oil motor. The dc motor fed by the transistor turns an oil valve for forward or reverse operation of an oil motor. A feedback signal responsive to the position of the oil valve is matched against the remote-control signal. Then there is another feedback responsive to the position of the equipment driven by the oil motor. This feedback is matched against the master control.

An error of position generates a remote-control signal. This signal activates the transistor and starts the dc motor. When the valve has turned to a position corresponding to the strength of the signal, the negative feedback from the oil valve neutralizes the signal and the dc motor stops, leaving the oil valve in an open position.

The oil motor therefore continues to run until the error has been corrected. The motion of the oil motor diminishes the error, but before the error has been reduced to zero, another chain of events takes place. The negative feedback from the oil valve overpowers the error signal and the transistor is activated in negative direction, thereby reversing the dc motor which in its turn closes the oil valve. The equipment driven by the oil motor will, therefore, approach the correct position gradually and not overrun.

This system is also adapted to control by radio. The radio signal may contain supermodulated frequencies, which give positive or negative indications which are rectified and fed into the remote-control circuit which activates the transistor. When positioning control by radio is wanted, this can be accomplished by matching the radio-controlled signal against a feedback signal from the motor-driven equipment.

FREQUENCY CHANGER

Fig. 3 shows an application of the transistor as a frequency changer. The load circuit is an inductance capacity circuit tuned to the desired frequency and feeding an induction motor which may be operated at variable speed, depending upon the adjustment of the variable shunt capacitor.

The phase control is connected to the power system and the polarity control to the output circuit. When the positive voltage of the power circuit coincides with the positive voltage of the load circuit, the transistor becomes conductive in the positive direction and delivers power to the load circuit. When the negative power voltage coincides with the negative output voltage, the transistor also becomes conductive, but then in the negative direction. Thus a negative current flows from a negative source, giving a positive transfer of energy.

CONCLUSIONS

In the two applications here described, a single transistor functions as a reversible rectifier and as a frequency changer. It thus exhibits properties not possessed by any device known so far. The possibilities of polyphase and composite transistor circuits are too numerous for exploration at this time. This promises to become a new branch of industry and research, if transistors can be developed for reasonably high power. In pointing out these possibilities, the author hopes to contribute an incentive for such a development.

BIBLIOGRAPHY

1. W. Shockley, "Holes and electrons," *Phys. Today*, vol. 3, pp. 16–24; October, 1950.
2. G. L. Pearson, "The physics of electronic semiconductors," *Trans. AIEE*, vol. 66, pp. 209–214; 1947.
3. W. Shockley, "The theory of *p-n* junctions in semi-conductors and *p-n* junction transistors," *Bell Sys. Tech. Jour.*, vol. 38, pp. 435–489; July, 1948.
4. W. Shockley, M. Sparks, and G. K. Teal, "The *p-n* junction transistors," *Phys. Rev.*, vol. 83, no. 1, pp. 151–162; July, 1951.
5. R. L. Wallace, Jr. and G. Raisbeck, "Quality as a guide in transistor circuit design," *Bell Sys. Tech. Jour.*, vol. 30, pp. 381–418; April, 1951.

Power Rectifiers and Transistors*

R. N. HALL†

Summary—The behavior of donor, acceptor, and ohmic contacts prepared by fusing impurity metals to germanium is discussed. Power rectifiers having rectification ratios as high as 10^7 can be made by fusing donor and acceptor contacts to opposite surfaces of a germanium wafer. An analysis of their electrical characteristics is presented which agrees well with the measured performance. The properties of transistors prepared in a similar manner and which are capable of 100 watts of output power are described.

I. INTRODUCTION

TWO DISTINCT CLASSES of germanium *p-n* junction devices are now being manufactured. Although both conform very accurately to theoretical expectations, their electrical characteristics are significantly different and their fields of application differ accordingly.

One type of unit[1,2] makes use of a small bar of germanium whose impurity content changes abruptly at one or more places along its length as shown in Fig. 1.
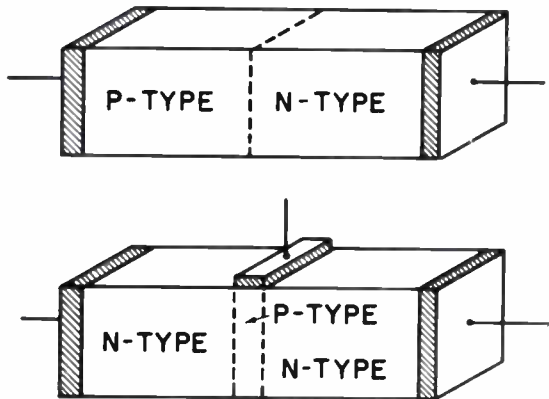
Fig. 1—*P-n* junction rectifiers and transistors.

Although rectifiers and transistors constructed in this manner exhibit remarkably good low-level electrical characteristics, they are not well suited to high-power applications.

A good example of the second method of making *p-n* junction devices is the General Electric G-10 rectifier, shown schematically in cross section by Fig. 2(a). This germanium power rectifier is prepared by fusing donor and acceptor impurity elements to opposite sides of a small wafer of germanium.[3] The excellent electrical and thermal contact with both sides of the wafer, which is possible with this type of construction, permits efficient

operation at very high current densities. Several kilowatts may be reliably handled by a rectifier made from a piece of germanium the size of a dime.
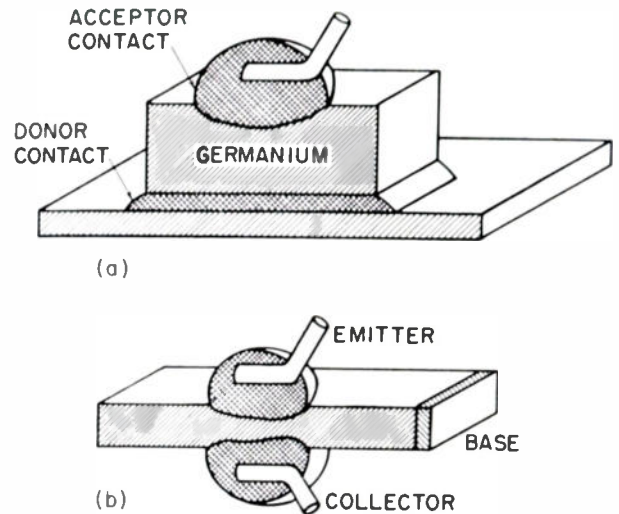
Fig. 2—Fused impurity contact rectifiers and transistors.

Fused impurity contact transistors, capable of delivering a few watts of output power, have been made by fusing a small acceptor impurity contact to each side of an *n*-type germanium wafer,[4] as shown in Fig. 2(b). An investigation of the limitations of this type of transistor shows that the attainment of a substantial increase in power output cannot be achieved simply by increasing the areas of the emitter and collector contacts. For satisfactory high-power operation, all three electrodes must be placed close to each other, and each must have a large area of contact to the germanium. An electrode arrangement which meets these requirements is illustrated in Fig. 3. Emitter and col-
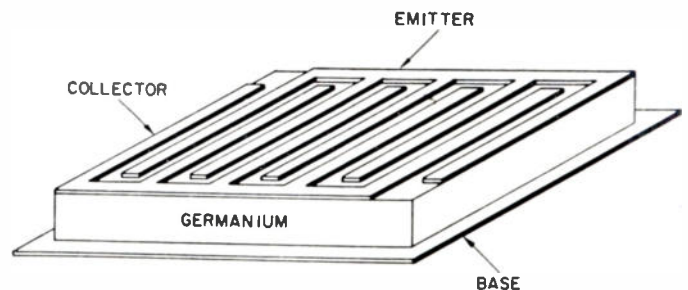
Fig. 3—High-power transistor construction.

lector electrodes are intermeshing grid structures fused to one side of a germanium wafer, while the base connection covers the other side. Somewhat better electrical performance may, perhaps, be obtained from

[1] F. S. Goucher, *et al.*, "Theory and experiment for a germanium *p-n* junction," *Phys. Rev.*, vol. 81, p. 637; 1951.
[2] W. Shockley, M. Sparks, and G. K. Teal, "*P-n* junction transistors," *Phys. Rev.*, vol. 83, p. 151; 1951.
[3] R. N. Hall and W. C. Dunlap, Jr., "*P-n* junctions prepared by impurity diffusion," *Phys. Rev.*, vol. 80, p. 467; 1950.
[4] J. S. Saby, "Fused impurity *P-N-P* transistors," PROC. I.R.E., vol. 40, pp. 1358–1361; this issue.

some of the other electrode arrangements which are possible. This structure works reasonably well, however, and was chosen because of its simplicity and because it illustrates clearly the principles of power transistor operation. Transistors capable of 100-watt output power have been constructed in this manner. Their electrical characteristics and principles of operation will be described below.

II. Construction of Fused Impurity Contact Junctions

When an impurity metal is fused to a piece of germanium, the impurity penetrates by diffusion a small distance into the unmelted germanium. Between this diffusion layer and the bulk of the impurity metal is usually found an oriented deposit of germanium saturated with the impurity metal, whose crystal structure is a continuation of that of the undisturbed germanium. Although the measured diffusion constants for n- and p-type impurities in germanium are very small, it appears that with the heating cycles commonly employed, the diffusion layer contains a sufficient number of impurity atoms to define the boundary conditions at the junction.[5] In any event, since it is followed by a much thicker layer of saturated germanium, the junction may be regarded as an abrupt transition from undisturbed material to germanium which is very heavily doped.

Indium is preferred as an acceptor contact material because its softness eliminates problems due to differential expansion. When it is bonded to high-purity n-type germanium, the p-n junction which is formed is capable of blocking inverse voltages as high as 1,000 volts. When operated in the forward direction, this type of junction functions as a highly efficient hole emitter because of the very large impurity ratio on the two sides of the junction.

Indium fused to p-type germanium forms a nonrectifying contact which is ideal in the sense that it is a source of holes only. In other words, the interface between the indium and the undisturbed germanium has a negligible surface recombination velocity so that electrons are not generated there, nor do they recombine preferentially there when injected in the neighborhood of the indium contact.

Similar statements apply to donor contacts which may be made by fusing antimony to n- or p-type germanium. In this case, however, the recombination velocity at the contact, although small enough to be ignored in the present treatment, is not always negligible. P-n junctions formed between p-type germanium and antimony have inferior reverse characteristics, presumably because of damage resulting from differential expansion between the germanium and the brittle antimony-germanium eutectic.

Ohmic contacts result when the recombination velocity at the surface between the electrode and the germanium is very high. Most metals which are in neither the third nor fifth columns of the periodic table, such as tin or lead, give rise to ohmic contacts when fused to germanium. An electroplated connection to a sandblasted germanium surface provides a similar high recombination-rate contact.

Donor, acceptor, and ohmic contacts are terms which apply to three different and clearly defined types of electrical connection which may be made to germanium. Although contacts of each type may be closely approximated in practice, it will be appreciated that electrodes of intermediate character can be obtained. For example, a continuous range of tin-indium alloys may be used to give a variety of contacts ranging from ohmic to purely acceptor in character.

III. High-Level Injection in Germanium

Fused-contact rectifiers and transistors are generally constructed from nearly intrinsic n-type germanium. The flow of appreciable current densities requires the injection of both holes and electrons into the germanium in concentrations much greater than are found under equilibrium conditions. Units of the type shown in Fig. 1, on the other hand, are generally operated under conditions of low-level injection where the minority-carrier concentration is always small compared with the impurity content. The treatment which is valid for the analysis of p-n junction bars must be modified somewhat to take into account some of the large signal effects which are important in fused contact devices.

The three equations of Shockley[6] which govern the flow of carriers within the volume of a semiconductor are written below in a form which is more convenient for our purpose.

$$J = e\mu(p + bn)E - eD\nabla(p - bn), \quad (1)$$
$$\text{(drift)} \qquad \text{(diffusion)}$$

$$D\nabla^2(p + n) - \mu\nabla \cdot [(p - n)E] = \left[1 + \frac{1}{b}\right]R, \quad (2)$$

$$\nabla \cdot E = \frac{4\pi e}{K}(p - n + N). \quad (3)$$

Equation (1) states that a current flows because holes and electrons drift under the influence of an electric field, E, and because they diffuse in the direction of decreasing concentration gradient; p and n are the hole and electron concentrations, their mobilities being given by μ and $b\mu$, respectively. The diffusion constant, D, is related to the mobility, $D = kT\mu/e$.

The steady-state distribution of electrons and holes is affected by the rate at which they disappear by recombination as described by (2). The recombination func-

---

[5] W. C. Dunlap, J. S. Saby, and A. C. English discussed the structure of fused impurity contact junctions at the IRE and AIEE Conference on Semiconductor Device Research; June, 1952.

[6] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. Jour.*, p. 451; 1949.

tion, R, is the rate at which volume recombination exceeds thermal generation. It depends upon hole and electron concentration, crystal perfection, and temperature as has been described elsewhere.[7]

In the analysis of the low-level performance of junction devices, it may be assumed that R varies linearly with the departure of the minority-carrier concentration from its equilibrium value. When very large numbers of holes and electrons are injected in equal numbers into germanium, it is again found that the recombination rate is linearly related to the carrier concentration, but with a rate constant corresponding to a much shorter lifetime. This lifetime, which is found to be in the neighborhood of 40 μsec, is the quantity which determines the diffusion distance for electrons and holes under conditions of high-level injection. The variation of lifetime with injection level is indicated in Fig. 4.
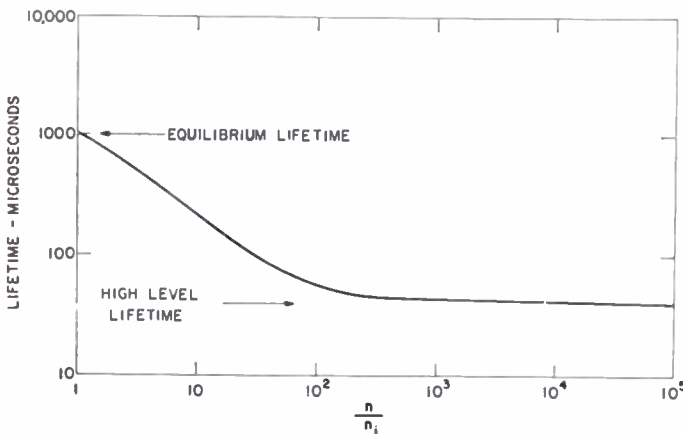
Fig. 4—Change in lifetime with injection level for pure germanium at room temperature.

Poisson's equation is the third relation between the hole and electron concentrations and the electric field. It states that the field strength within the crystal changes rapidly wherever the space charge produced by the excess, N, of donor impurity atoms over acceptor atoms is not balanced by an equal difference between the hole and electron concentrations.

If the impurity content of a germanium sample is not balanced accurately in this way, the resulting space charge is sufficient to generate potentials of hundreds of volts in a small fraction of a millimeter. Thin blocking layers of this kind are created when a p-n junction is biased in the reverse direction, all free charge carriers within the layer being swept out by the high electric field. Elsewhere in the semiconductor, the electric field strength must be small because of the presence of free charge carriers. Poisson's equation requires that the electron and hole concentrations vary together in order to preserve space-charge neutrality. The assumption that the hole and electron densities change in equal

amounts is a very good approximation and greatly simplifies the analysis of junction devices.

IV. Fused Impurity Contact Rectifier Analysis

Most germanium p-n junction devices consist of regions of uniform germanium separated by very thin layers within which the impurity content changes rapidly. It is convenient to treat each uniform region as a boundary value problem in the analysis of the characteristics of the entire device. Electron and hole flow in the uniform regions is governed largely by diffusion and recombination, since the electric field strength is very small. Strong dipole fields exist in the transition layers, on the other hand, due to the rapid change in impurity content, and these dipole layers produce corresponding changes in the hole and electron concentrations on either side.

The change in carrier concentration in going across a transition layer may be obtained by integrating (2), recognizing that recombination is negligible because of the small distance involved. We find the relation

$$n_1/n_2 = p_2/p_1 = \exp(e\Psi/kT), \tag{4}$$

where Ψ is the potential difference between sides 1 and 2 of the dipole layer. This, of course, is the relation which is to be expected for particles having a Boltzmann energy distribution.

Under equilibrium conditions, the concentrations are determined by the impurity content on each side, and the strength of the dipole layer may be evaluated from (4). The application of an external voltage to the crystal alters the potential difference and produces corresponding changes in the carrier concentrations on each side. The boundary conditions of a problem together with (4) are sufficient to determine the hole and electron flow within each of the uniform regions of the device.

The equilibrium distribution of holes and electrons in a fused impurity contact rectifier is indicated in Fig. 5. In the interior of the high-purity wafer of germanium, both types of carriers are present in numbers which are comparable with the intrinsic concentration. When n-type material is used, the transition from n-type to p-type germanium takes place near the acceptor electrode as indicated. The application of a voltage, V, in the reverse direction produces a space-charge blocking layer of thickness $w = (VK/2\pi eN)^{1/2}$ near the acceptor electrode. A small reverse current flows because this layer, which is usually less than a mil in thickness, captures all of the minority carriers which diffuse to it from the remainder of the germanium. For voltages greater than a few tenths of a volt, this component of the reverse current is constant. For thin wafers, it is equal to the product of the volume of the germanium wafer by the rate at which holes are generated in it by thermal agitation. An additional component resulting from surface conduction and crystal imperfections may become significant at high inverse voltages. An upper limit to the inverse voltage that a junction can support

[7] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87 p. 387; 1952.

is set by the appearance of the Zener current[8] at a critical voltage which is inversely proportional to the impurity content of the germanium. Very high purity n-type

IMPURITY DISTRIBUTION

HOLE AND ELECTRON DISTRIBUTIONS

Fig. 5—Carrier distribution in fused impurity contact rectifier.

material is necessary for the construction of high inverse voltage rectifiers.

The application of a forward voltage to the rectifier causes the injection of holes and electrons into the interior of the germanium in equal numbers. Holes injected by the acceptor electrode at $x = -d$ are prevented from entering the donor region at $x = d$ by the potential hill arising from the impurity gradient there, and vice versa. Some flow of injected carriers into the opposite impurity contact is observed at high forward voltages as will be described later.

The carrier concentrations n_d and n_a at the donor and acceptor surfaces of the pure germanium are shown by (4) to depend exponentially upon the height of the potential hill between the pure germanium and the heavily-doped material where the carrier concentration is so large that it is unaffected by current flow. Most of the applied voltage, V, appears as a lowering of the potential hills at the donor and acceptor sides of the rectifier by amounts V_d and V_a. The remainder consists of the internal voltage drop, V_i. When the rectifier thickness is small, this component can be made negligible, as will be shown later.

[8] K. B. McAfee, E. J. Ryder, W. Shockley, and M. Sparks, "Observations of Zener current in germanium *p-n* junctions," *Phys. Rev.*, vol. 83, p. 650; 1951.

Equation (4) relates the change in potential at each electrode to the carrier concentration there:

$$n_a = n_i \exp (eV_a/kT)$$
$$n_d = n_i \exp (eV_d/kT). \tag{5}$$

A second requirement upon the boundary conditions is that the ratio of the concentration gradients at the two sides of the rectifier be equal to the mobility ratio, b.

$$\nabla n \big|_{x=-d} = - b\nabla n \big|_{x=d}. \tag{6}$$

This relation follows because the net current density is everywhere the same and the flow of current across each side of the rectifier is due to carriers of one kind only.

As explained in Section III, the recombination rate when large numbers of carriers are present is proportional to the carrier concentration. Neglecting thermal generation, we may write $R = n/\tau$ where τ is the high-level lifetime. Because n and p are equal, the middle term of (2) may be omitted, and the recombination equation takes the simple form.

$$\nabla^2 n = n/L^2, \tag{7}$$

where $L^2 = 2D\tau/(1 + 1/b)$. L, the high-level diffusion length, is about five times smaller than the equilibrium diffusion length for minority carriers in high-purity germanium. The solution to (7) which satisfies the boundary conditions (5) and (6) is

$$n = n_i \frac{(b + 1) \cosh(d/L) \cosh(x/L) - (b - 1) \sinh(d/L) \sinh(x/L)}{\sqrt{(b + 1)^2 \cosh^4(d/L) - (b - 1)^2 \sinh^4(d/L)}} e^{\frac{e(V_a + V_d)}{2kT}}. \tag{8}$$

The current flow through the rectifier is largely accounted for by the rate at which holes and electrons recombine within the wafer,

$$J = \int_{-d}^{d} eR\,dx,$$
$$= \frac{2en_iL}{\tau} \frac{\tanh(d/L)}{\sqrt{1 - \frac{(b - 1)^2}{(b + 1)^2} \tanh^4(d/L)}} \tag{9}$$
$$\times \exp [e(V_a + V_d)/2kT].$$

When the half-thickness of the wafer is small compared with L, this expression simplifies to

$$J \approx \frac{2edn_i}{\tau} \exp (eV/2kT) \tag{10}$$

since

$$V_a + V_d \approx V.$$

The factor of 1/2 which appears in the exponential of these equations distinguishes the behavior of fused impurity contact rectifiers from that of simple *p-n* junction rectifiers[1] operated under conditions of low-level injection.
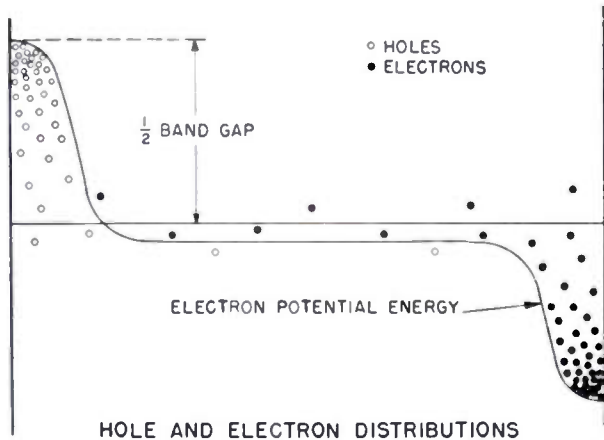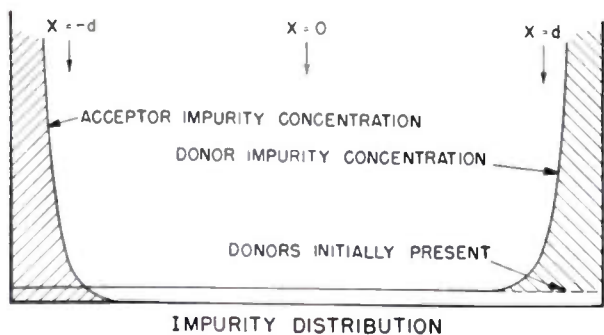
Forward-current data from a typical fused contact rectifier are shown in Fig. 6. Good agreement with theory is observed up to current densities of several hundred amperes per square centimeter. Beyond this point, the current increases less rapidly with voltage than the above theory would predict. It is believed that this deviation is the result of a decrease in mobility due to hole-electron scattering which should become important at injection levels greater than 10^{17} cm^{-3} as is the case with impurity scattering.
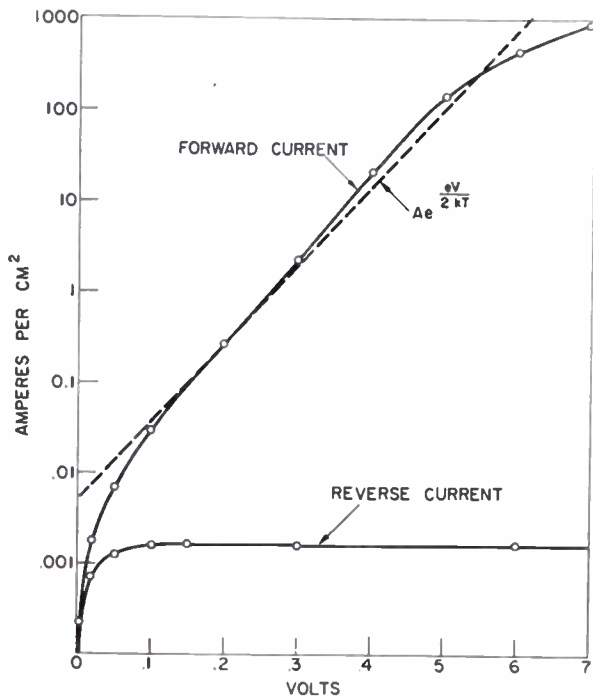
Fig. 6—Current-voltage characteristics of fused impurity contact rectifiers at 25°C. Equation (10) fitted to the dashed line gives $\tau = 38$ μs for $n_i = 2 \times 10^{13}$ cm^{-3} and $d = 0.03$ cm.

A current component due to injected carriers flowing over both potential hills is to be expected when the applied voltage is sufficiently great. This component should vary as $\exp(eV/kT)$ and an estimate of its magnitude shows that it could very well account for the slight rise above the theoretical curve which can be noticed in Fig. 6.

A more critical test of the theory is obtained by measuring the rectifier characteristics over a wide range of temperatures. Fig. 7 shows a plot of log (J) against $1/T$ for a series of forward voltages. The data are compared with the corresponding curves calculated from (10), modified slightly to indicate the effect of thermal generation which is significant at low voltage and high temperature. Except for a leakage current at very low temperatures and the high-current deviations described above, it can be seen that the agreement is good.

The change in carrier lifetime with injection level shown in Fig. 4 must be taken into account for forward voltages below 150 mv. Low-level analysis may be used for the calculation of the reverse characteristic and

the current flow for small forward voltages and yields the following expression for a rectifier made from an n-type wafer of thickness $2d$:

$$J = (eDn_i^2/NL_p)\tanh(2d/L_p)[\exp(eV/kT) - 1]. \quad (11)$$

L_p is the equilibrium diffusion length for holes in germanium having an excess donor impurity content N.

The voltage drop within the germanium, V_i, may be calculated by solving (1) for the electric-field strength and integrating. The resulting expression is cumbersome and since an estimate of this voltage drop is all that is generally required, it is sufficient to consider the special case $b = 1$.

$$V_i = 4(D/\mu)\sinh(d/L)[\tan^{-1}\exp(d/L) - \pi/4]. \quad (12)$$

This expression reduces to the following simple forms for two limiting cases:

$$V_i \approx 2(D/\mu)(d/L)^2 \quad \text{for} \quad d/L \ll 1, \quad (13)$$

$$V_i \approx (\pi D/2\mu)\exp(d/L) \quad \text{for} \quad d/L \gg 1. \quad (14)$$

Since the ratio D/μ is only 26 mv, V_i may be neglected except when $d/L > 1$. It is interesting to observe that V_i does not depend upon current flow except at extremely high current densities where electron-hole scattering causes L itself to decrease. This is because the conductivity of the germanium is increased by carrier injection as rapidly as the current flow increases.

The fused-contact rectifier construction discussed above has yielded units having rectification ratios as high as ten million. Nearly as good results may be obtained when impurities which give a high recombination-rate contact, such as tin, are substituted for antimony. In this case, most of the recombination of holes with electrons takes place at the surface of this electrode giving rise to a somewhat greater voltage drop at very high current densities and an increase in the reverse saturation current.[9]

V. POWER TRANSISTORS

Fig. 3 illustrates a p-n-p transistor made by fusing two sets of acceptor electrodes to an n-type wafer of germanium to create the emitter and collector junctions. The base electrode on the opposite side serves as a source of electrons and as a means of removing heat from the unit. In operation, holes and electrons are injected by the emitter and base electrodes as in the power rectifier described above. The collector, which is biased in the reverse direction, functions as a perfect sink for holes, and competes with the recombination processes for the holes injected by the emitter. For efficient transistor performance, it is important that a substantial fraction of the injected holes be picked up at the collector. In addition to having the electrodes closely

[9] A discussion of the reverse saturation current for rectifiers constructed in this manner was presented by J. A. Hornbeck and P. W. Foy at the IRE and AIEE Conference on Semiconductor Device Research; June, 1952.

FORWARD CHARACTERISTIC OF GERMANIUM RECTIFIER
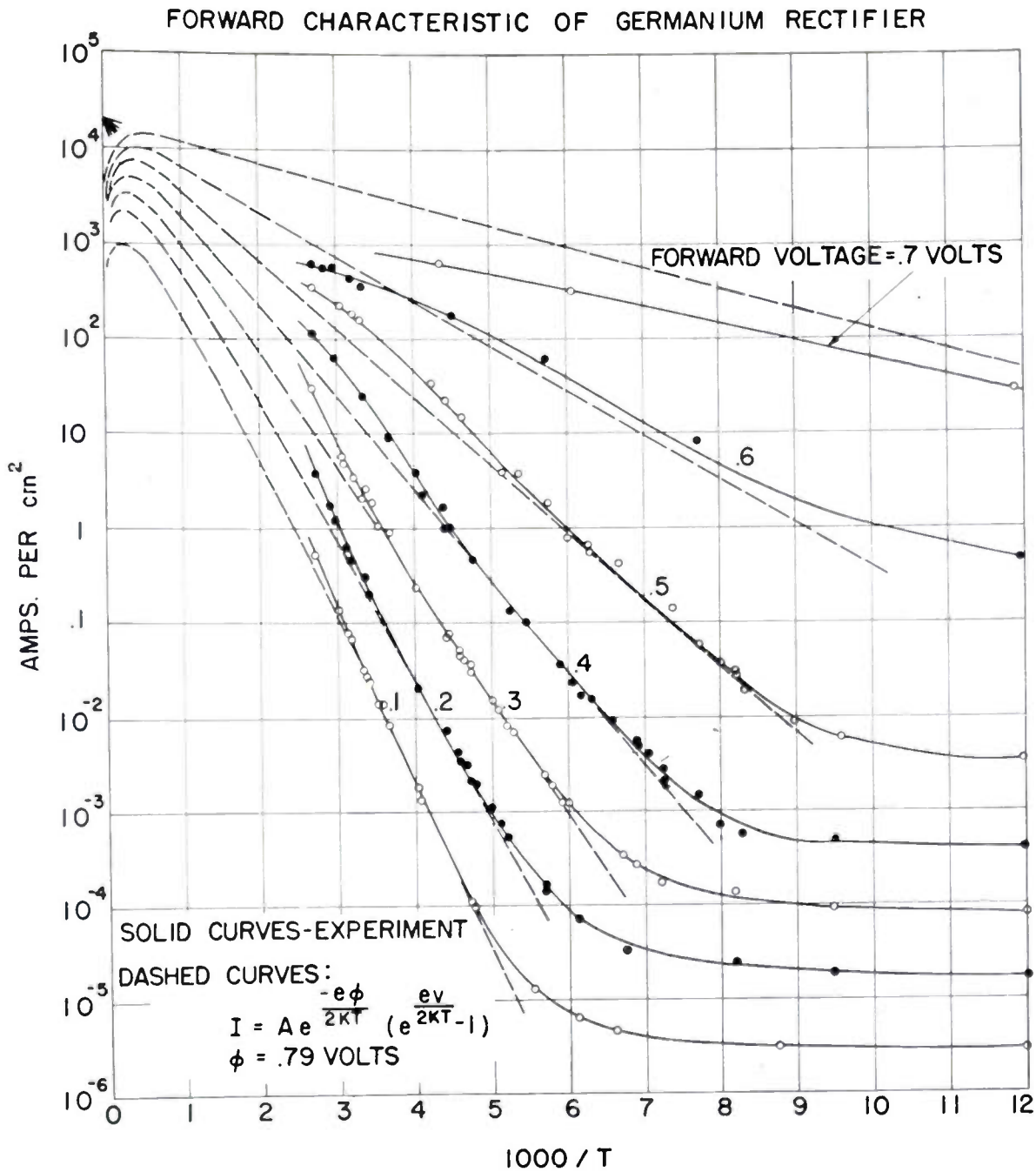
Fig. 7—Forward current versus absolute temperature between 100°C. and −190°C.

spaced, it is essential that none of them be high recombination-rate contacts in order that surface and volume recombination be kept at a minimum.

In analyzing the performance of this type of transistor, several simplifying assumptions will be made in order to bring out the essential design considerations with a minimum of effort. Preliminary considerations show that best performance is to be expected when the wafer thickness is slightly less than the emitter-collector spacing, d. These proportions provide for adequate flow of holes from emitter to collector without the excessive volume recombination which would result from the use of a thick wafer of germanium. It will be

assumed that the distribution of injected carriers in the germanium between the emitter and base contacts is reasonably uniform and that the concentration there is given by $n_d = n_i \exp(eV/2kT)$ as in (5). Since the carrier density must vanish at the collector electrode, we shall use the following expression to represent the distribution of holes and electrons between the collector and the emitter-base region a distance d away:

$$n = n_d \frac{\sinh x/L}{\sinh d/L}. \tag{15}$$

In calculating the flow of holes into the collector contact, it is necessary to take into account the reflection

of electrons from this surface. Since the net flow of electrons must vanish there, the electron diffusion current must be balanced by a drift current arising from the accumulation of a slight excess of electrons near the collector electrode. Because holes are oppositely charged, this electric field exactly doubles the hole flow which would result from diffusion alone. The collector current density is therefore

$$J_c = 2qD\nabla n\big|_{x=0}$$
$$= 2qDn_d/[L\sinh(d/L)]. \tag{16}$$

Since electrons can disappear only by volume recombination, the base current density is given by

$$J_b = \int_0^{d} eR\,dx$$
$$= \frac{eDn_d}{L\sinh(d/L)}\left[\cosh(d/L) - 1\right]. \tag{17}$$

The current gain of the transistor, defined as the ratio of collector current to emitter current, takes the following form,

$$\alpha = 2/(1 + \cosh d/L). \tag{18}$$
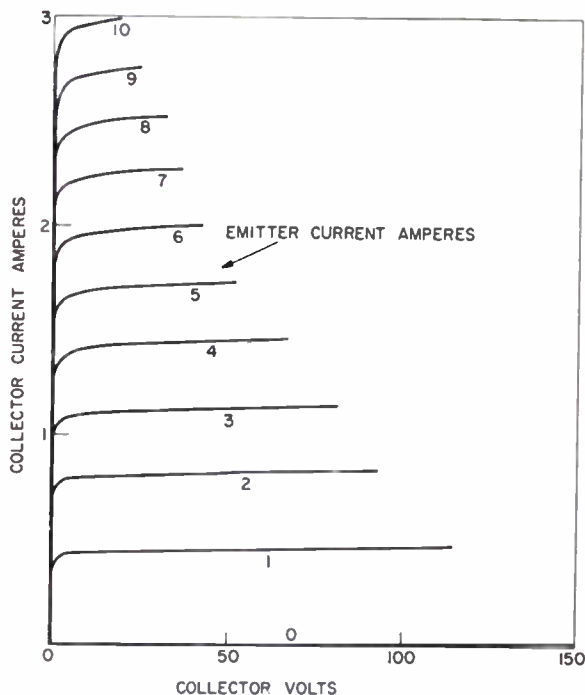
Fig. 8—Power-transistor characteristics.

Since the high-level diffusion length, L, is observed to be roughly half a millimeter, the electrode spacings should be no greater than this if a reasonable current gain is to be attained under conditions of high-power

operation. In actual practice, the gain will be somewhat less than the above formula would indicate because of spreading effects which were not taken into account in the derivation.

The performance of a power transistor having electrode spacings of 1/2 mm and a total area of one square centimeter is shown in Fig. 8. The collector current is observed to be nearly independent of collector voltage, as is typical for junction transistors. The current gain drops from 0.6 at low level to about 0.3 in the region of high-current operation due to the decrease in lifetime with injection level. The power output of this type of transistor is seen to be in the neighborhood of 100 w for class C operation and somewhat less under class A conditions because of the increased collector dissipation.

The current-voltage characteristics of the emitter are very similar to those of a fused-contact rectifier and are almost unaffected by changes in the collector circuit. Since the emitter impedance is low and strongly dependent upon emitter current, it is seldom possible to realize all of the available power gain which is usually between 10 and 20 db under full-power operating conditions.

The power transistor described above is in a very early stage of development. Considerable improvement in current gain and frequency response is to be expected when techniques are developed for applying closely-spaced electrode contacts to the germanium. Present units, with 1/2-mm electrode spacings, are limited to audio-frequency applications by the time required for injected carriers to diffuse this distance through the germanium.

VI. Conclusions

The fused impurity contact technique has made possible rectifier and transistor structures such as those discussed in this paper. Although the electrical contacts which are produced in this manner are well suited to high-power applications, the same type of construction can be used effectively in the preparation of devices intended for small signal operation.[4]

The frequency response of the power units described is limited to about 20 kc by transit-time effects and by the out-of-phase current which arises because carriers must be injected into the germanium and removed from it during each cycle.

A further limitation which these units have in common with other germanium *p-n* junction devices comes from the increase in reverse saturation current with temperature. The performance of power rectifiers is particularly affected by this component of the reverse leakage which increases about 8 per cent per deg C.

A High-Voltage, Medium-Power Rectifier*

C. L. ROUAULT† AND G. N. HALL†

Summary—A high-voltage germanium *p-n* junction rectifier resulting from improved materials and techniques is described. Its characteristics are briefly delineated. A method of rating based upon distribution of internal losses emphasizes versatility of this rectifier.

DIRECT CURRENT is a necessity for a large number of industrial and electronic processes, but is not as readily obtained as might be desired. To the designer of a dc power system, a number of methods are available: motor generator sets, dry plate rectifiers, mercury arc rectifiers, high-vacuum rectifiers, vibrators, and the like, each of which possesses certain advantages. If he were given a choice, unrestricted by economics, it is quite likely that he would choose characteristics for an efficient ac-dc converter as follows: (1) the low conducting IR drop of the dc machine, (2) the peak inverse voltage ratings of the high-vacuum rectifier, (3) the low reverse current of the high vacuum rectifier, (4) the mechanical simplicity of the selenium stack, (5) low first cost.
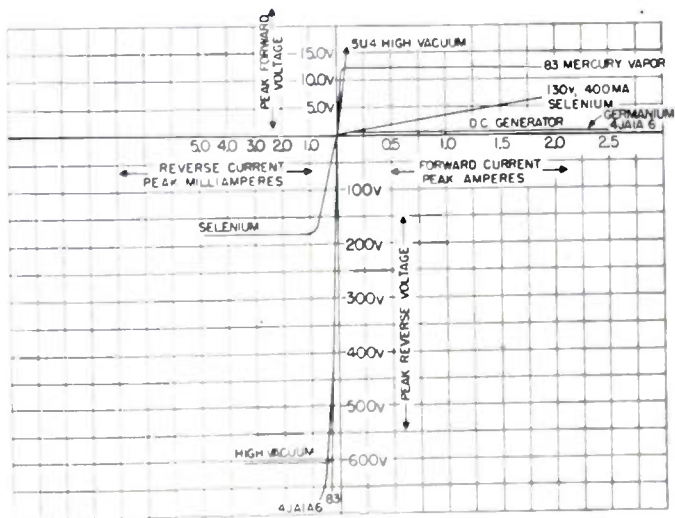
Fig. 1—Comparative characteristics, rectifiers of similar forward current ratings.

Until the advent of the germanium *p-n* junction rectifier recently made commercially avilable, the five points on the "wish list" were not available in one ideal package. The germanium junction rectifier is not ideal, but its characteristics are sufficiently better in order of magnitude to make it appear nearly ideal by comparison with devices whose characteristics are shown on Fig. 1.

Fig. 2—Dynamic voltage and current characteristics, germanium junction diode Type 4JA1A–4JA2A.

On Fig. 2 are shown the characteristics of the junction rectifier alone. Fig. 3 is a picture of this rectifier.

The mechanism of the *p-n* junction rectifier has been described completely elsewhere in the literature,[1] so no further description will be undertaken.

Fig. 3—Germanium power rectifie. Type JA-2A.

This rectifier exemplifies to a high degree of precision the characteristics obtainable in a controlled *p-n* junction made by the fused impurity contact method.[2] In fact, the correlation between the theoretical performance and that actually obtained over a substantial manufacturing run is so good that units of greater ratings than those described in this paper may be projected with confidence.

[1] W. Shockley, "Electrons and Holes in Semi-Conductors," D. Van Nostrand Co., Inc., New York, N. Y.; 1950.
[2] C. W. Dunlap and R. N. Hall, "PN junctions prepared by impurity diffusion," letter to *Phys. Rev.*, p. 467; 1950.

The characteristics shown in Fig. 2 show a reverse current, I_r which is a significant improvement over previous junction rectifiers made on a small-scale basis. Rectifier cells are still selected on the basis of Zener breakdown voltage in the reverse direction, but now the yield of 600 volts (and over) units is economically significant.
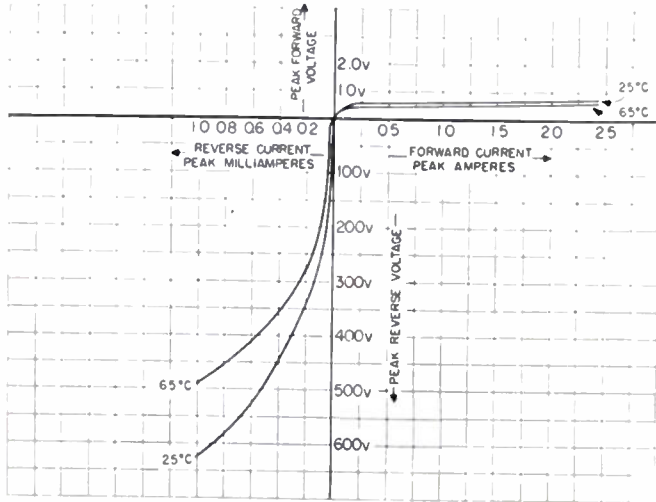
Fig. 4—Temperature characteristics, germanium junction diode Type JA-1A.

The variations in forward characteristic are very small compared to those in the reverse direction. This is one of the scientifically gratifying results of the improvement in manufacturing techniques during the past year, for the theory of the junction rectifier predicts that there should be little relation between these characteristics when units are processed from equivalent materials. The variations noted are primarily the result of process imperfections which are more or less inherent at the present time.

The variations in forward voltage drop is so small that parallel operation is no problem of consequence. The current differences in the back direction for the same grade of rectifier are quite small, so that units may be coupled in series with a modest sacrifice in gross peak inverse voltage rating.

Germanium, like other semiconductor materials, possesses a substantial negative temperature resistance coefficient which imposes a limitation on maximum operating temperate. At a temperature well under 100°C, the semiconductor properties of germanium are degraded, so that for practical purposes a temperature at the junction of 75°C should not be exceeded. In Fig. 4 a typical temperature characteristic of a typical germanium junction rectifier is given.

The power-handling capability is determined by the internal losses which cause self heating above the ambient temperature. This self heating may be expressed as follows:

$$P_i = \frac{2}{T}\int_0^{T/2} v_f i_f dt + \frac{2}{T}\int_{T/2}^{T} v_r i_r dt = P_f + P_r,$$

P_f = power dissipated in forward direction,
P_r = power dissipated in reverse direction,
P_i = power dissipated internally, per cycle,
V_f = forward voltage drop,
I_f = forward current,
V_r = reverse voltage drop,
I_r = reverse current.

For a given internal dissipation, then, one may expend the losses P_r and P_f as follows:

1. $P_f \approx P_r$. Medium voltage, medium current application, receiver, and general purpose power supplies, and the like.

2. $P_f \gg P_r$. Low voltage, high current applications; filament supplies, battery chargers, and the like.

3. $P_f \ll P_r$. High voltage, low current, applications; instrument rectifiers, potential devices, and the like.

These three cases illustrate the versatility of this junction rectifier.

If we limit the reverse current to a predetermined safe value, and the temperature within the limits cited, we may define a power-transmission characteristic given in Fig. 5.

The implication of Fig. 5 is that we may tolerate a lower peak inverse voltage rating, increased back current, and still have a very efficient rectifier for low-voltage, high-current requirements, or any intermediate condition of voltages and currents up to Zener breakdown and within the maximum temperature limitation.
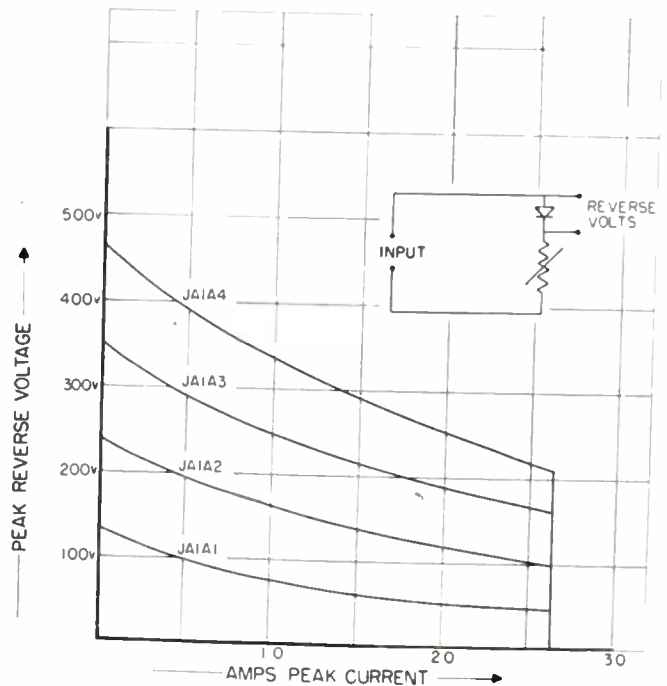
Fig. 5—Power capability characteristics, germanium junction Type 4JA1A unfinned 25° C.

Since the primary limitation on power handling capability is of thermal nature, the effects of overloading are important. Repeated overloads on test units have shown that unless the overload results in fusion of a constituent part that relief of the overload restores the device to normal operation. The self heating cannot al-

ways be depended upon, but is the normal case.

When cooling fins, or other means of cooling, are employed to increase the rate of heat transfer to the surrounding sink, the ratings shown may be markedly increased; but note that a reduction in temperature increases the forward drop, and hence increases the forward losses. For the same reason, the reverse current decreases so that an increase in power transmission capabilities by means of additional cooling may be accompanied by a slight reduction in efficiency, depending upon the distribution of internal losses. In the rectifiers we are discussing, a small portion of the forward drop is due to the ohmic nature of the contacts

rather than the rectifying junction. These general characteristics are shown in Fig. 6.

A rectifier of low forward resistance and high back resistance is of interest not only in the power rectification field, but over the general frequency spectrum as a limiter, clamper, and the like. Certain energy storage effects definitely restrict the usefulness in pulse circuits, but these effects may be reduced somewhat by proper circuitry. The sine-wave response, however, is good up to frequencies on the order of 100 kc, as shown by the rectification efficiency plot given in Fig. 7.

Fig. 6—Characteristics, germanium junction diode Type JA-1A.

Fig. 7—Frequency response, germanium junction diode Type JA-1A.

While this paper described a specific rectifier of modest power ratings, it is planned to cover the power rectifier field with units of increased rating commensurate with the application.

Application of Transistors to High-Voltage Low-Current Supplies*

G. W. BRYAN, JR.†

Summary—A battery-powered, high-voltage, low-current supply is described, which utilizes a transistor as an oscillator. Significant savings in weight and power are realized.

IN THE DESIGN of portable, battery-operated radiac instruments, it is often necessary to use rather high voltages at low currents. Instruments such as Geiger counters, ionization chambers, and scintillation counters require voltages ranging up to 1,000 volts or more. The current drain is usually very small, seldom exceeding 5–10 μa. Voltages and currents of the magnitude mentioned may be readily obtained

from batteries, but their size and weight are excessive. For instance, a 900-volt battery pack of a type commonly used in portable Geiger counter instruments weighs approximately $2\frac{3}{4}$ pounds and occupies a space of roughly 60 cubic inches. Other sources of high voltage at low currents include vibrator, radio-frequency, and blocking oscillator supplies. These types require an input power from 0.1 to 0.3 watt, including filament and plate power. In general, the three last-mentioned types of power sources have low efficiency even when maximum power is taken by the load. However, in the case of ionization chambers and Geiger counters, with the exception of some halogen-filled types, the power required from the high-voltage supply seldom exceeds

200 μw. It is therefore apparent that a great deal of efficiency is lost in the applications mentioned because of the extremely light loading. In the oscillator type of supply, filament power alone may run as high as 0.1 watt. Since battery life and weight are directly related, it is usually necessary to compromise one for the other.

In view of the above, and in an attempt to circumvent some of the undesirable characteristics of the high-voltage sources mentioned, an investigation was made of the possibility of utilizing the transistor as an oscillator in a high-voltage supply. The transistor is an attractive device since it requires no filament power and will operate at very low voltage and input current.

A number of the "point contact" transistors type
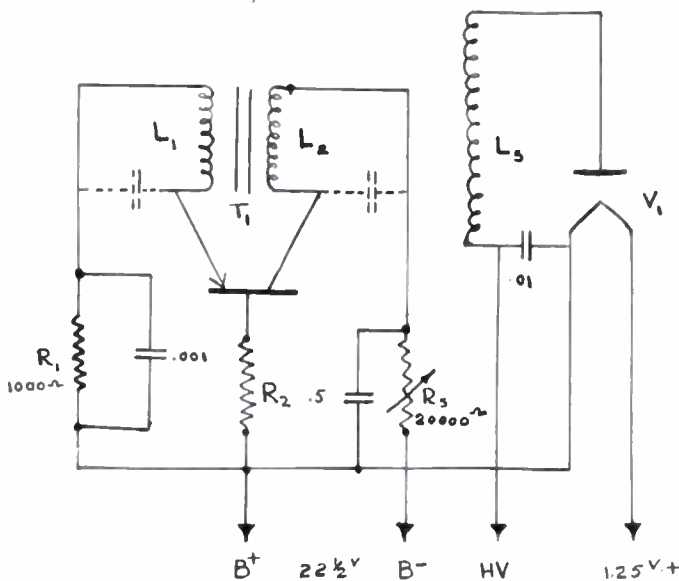
Fig. 1

1768 were arranged first in the familiar Hartley oscillator circuit (see Fig. 1). As may be seen in Fig. 1, the transformer T1 determines the frequency of oscillation, and is arranged to supply the necessary in phase feedback to the emitter. Winding L3 serves as the high-voltage winding. The rectifier V1 is a Victoreen type 5799 high-voltage diode, having a 10-ma filament. Base bias was used in the transistor circuit since this permits operation from a single battery. The circuit was found to oscillate easily at an input of 22.5 volts. However, in order to obtain an output voltage of 500 volts at 1 μa from the rectifier circuit, it was necessary to run the input current up to approximately 9 ma. Thus the power taken from the batteries amounted to 212 mw. This included filament power for the rectifier. Although the input compared favorably with vacuum-tube circuits, this circuit afforded no improvement in efficiency.

Since the aim was to decrease the average input power, it was decided to place the transistor in a relaxation oscillator circuit and utilize the "flyback" principle of operation (see. Fig. 2). The waveform of the oscillator is a sawtooth and the flyback of this wave is used to shock excite the high-voltage transformer T1 primary.

As may be seen in Fig. 2, base bias is used as in the first arrangement. The base-bias resistor R_2 has a negative temperature coefficient. A thermistor or varistor having a resistance of approximately 1,000 ohms at 70° F may be used. The nonlinear resistance allows both a high starting bias which drops rapidly as the operating current flows and easier starting at low battery potentials. However, a fixed resistance may be used if a value is chosen that allows the circuit to start oscillating easily and which maintains as low an average input power as possible. A suggested value would be 600 ohms. In-phase feedback between collector and emitter is obtained through capacitor C1 and the primary of transformer T1. The repetition rate of the sawtooth oscillations is determined primarily by R3, C1, and R1 in parallel with the emitter impedance. R3 is made variable, and therefore controls the high-voltage output. At
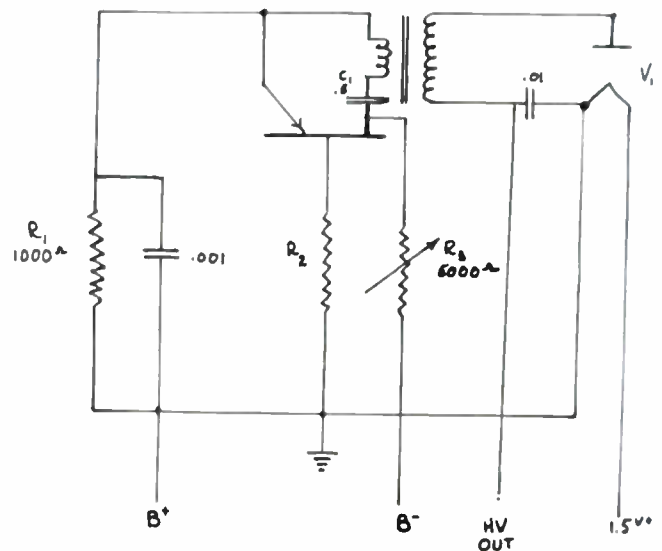
Fig. 2

an output voltage of 500 volts and at a current of 1 μa the input measured 12 volts at 3 ma. Total input power including the rectifier filament amounted to approximately 48 mw. In order to determine the relative efficiency of various transistors under identical circuit conditions, 30 units were checked and the input power measured. The average input was found to be 33 mw. The greatest deviation from the average was ±25 per cent, or approximately 25 mw for the most efficient unit and 41.8 mw for the least efficient unit.

The unit described here was designed to supply a sweep-out potential in a 600-cubic cm ionization chamber capable of measuring rates as high as 500 roentgens per hour. In this application, the high-voltage regulation was adequate. However, the supply described is inherently a high-impedance source of voltage since the apparent impedance is, among other things, a function of the oscillator repetition rate as in other pulse-type high voltage supplies.

If a higher degree of regulation is required in a particular application to compensate for changes in bat-

tery supply voltage or load current, a corona voltage regulator should be placed in shunt with the high-voltage output circuit.

The weight of the transistor supply is about $3\frac{1}{2}$ ounces. The space occupied is 4 cubic inches (see Fig. 3).

Fig. 3—Shows relative size of 500-volt transistor power supply as compared to a 300-volt battery.

Fig. 4

A second transistor supply was designed to supply 920 volts at 5 μa. The circuit is essentially the same, with the exception of the base bias which was found undesirable at higher input voltages (see Figs. 4 and 5).

Both supplies were designed to have a negative output polarity; therefore the rectifier filament is at ground potential and a battery is used to supply filament voltage. However, if a positive high-voltage output is required, a filament winding may be put on the high-voltage transformer in order to obtain filament excitation at a high-voltage above ground.

Fig. 5—Shows size of the 920 volt, regulated supply as compared to a 900-volt battery pack.

The high-voltage transformer $T1$ is equivalent to an audio output transformer having an impedance ratio of 8 ohms to 4,000 ohms. A suitable high-voltage transformer may be constructed according to the following specifications:

1. Core material—Allegheny Ludlum Type A Silicon steel or equivalent.
2. Core dimensions—$3/16 \times 3/16 \times 2$ inches long. Lamination thickness 0.014 inch.
3. Primary coil—186 turns #25 Formex wound in 4 layers. Winding length 1 inch.
4. Secondary coil—11,000 turns #41 Formex wound in 40 layers. Winding length $1\frac{1}{8}$ inches. Layer insulation 0.0005. Kraft paper.

Conclusions

A circuit was designed which utilizes the point-contact transistor as an oscillator in a high-voltage, low-current power supply. Such a supply offers the advantages of small size, light weight, and decreased battery drain, which is of great importance in portable radiac instrument design, as well as other applications where high voltages at low currents are required.

Printed Circuitry for Transistors*

S. F. DANKO†, MEMBER, IRE AND R. A. GERHOLD†, MEMBER, IRE

Summary—The current transistor development activity is being complemented by development of new families of subminiature components. Assembly of such minute parts by conventional hand-wiring techniques poses serious physical and economic production problems. The auto-sembly printed circuit system is presented as a simple, effective, and compatible approach to the fabrication of compact transistor circuits. Auto-sembly offers error-free, low-cost, high-production assembly lines with low demands on tooling and background "know-how."

I. Introduction

THE continually expanding applications of electronics in military and industrial operations has placed great emphasis on the miniaturization of equipment to reduce bulk and weight. The simplest and most obvious approach to such miniaturization involves the elimination of waste space and encroachment on the space usually allowed for the construction and subsequent maintenance of the equipment. Though effective in many instances, this miniaturization approach introduces new problems of high temperature and costly construction. Such problems are further aggravated when additional size reductions are sought using smaller high-temperature quality components. Very substantial relief on the equipment temperature-rise problem is now promised by the transistor which, because of its low-heat generation and low-energy levels of operation of the associated components, reduces by several decades the amount of heat that must be dissipated.

Such miniaturization using transistors and other minute components introduces a critical fabrication problem involving physical manipulation, wiring, and solder assembly of subminiature components. Conventional hand-wiring approach to assembly of such small transistor circuits is tedious, slow, and costly, and is naturally conducive to high rejection in production due to wiring errors, accidental short-circuits, and other production pitfalls characteristic of such finely detailed work.

As a first step to facilitate production and subsequent maintenance, the principle of unitization is employed whereby a complex equipment is divided into units of small size and lesser complexity. However, subminiaturization of even these smaller units is difficult on a production basis using conventional hand wiring and soldering of individual components. Printed circuitry (or more properly "prefabricated wiring") offers a simple yet effective solution to these difficulties. The treatment of printed circuits devised by the Signal Corps Engineering Laboratories and termed "auto-sembly" (see Fig. 1) is considered to offer particular advantages in transistor circuit applications since it is a production technique realistically suited to assembly of small complex circuits using miniature components, yet demanding no manipulative skill or training of assembly personnel.

II. Auto-Sembly

As illustrated in Fig. 1, auto-sembly involves the use of a prefabricated (printed circuit) pattern prepared on an insulating base, which is perforated at various points where component leads are to be joined to the pattern. Conventional JAN or special subminiature components, including the transistor, are inserted into these perforations from the blank side of the chassis, and the pattern side is then momentarily touched to a molten solder bath to effect a "one-shot" mass soldering operation joining each component lead to the pattern.

The details of these operations are as follows:

(1) Conductive Pattern: In the laboratory, the simplest technique of forming a good conductive pattern is through the etching of commercially available copper-clad plastic laminates. In this technique, the pattern desired is screened on to the copper surface using etch-resistant ink (resist); in the subsequent etching operation in 50-per cent solution of ferric chloride, the copper not protected by the ink is etched out leaving the remaining protected copper delineated in the pattern desired. Such patterns can be prepared on various grades of phenolic, melamine, teflon, silicone, polystyrene, polyesters, and epoxy to suit particular physical and electrical circuit requirements. In mass production a similar technique using offset presses for the printing of the etch resistant ink is employed to yield patterns in mass quantities (Fig. 2). Photographic methods of preparing the etching resist are also very commonly used. Stamping and pressed-powder deposition techniques also yield highly conductive and solderable patterns.

(2) Components: For Service applications, no compromise in reliability can be tolerated in any miniaturization approach. The use of standard JAN components or other separately fabricated components of known and proven quality is consistent with this reliability objective. Where commercially available packaged combinations of "printed" resistors and capacitors meet performance and reliability requirements of the equipment, such combinations may be advantageously used with the conventional components to effect additional size and weight reduction.

(3) Assembly: Solder dipping of patterns has been established in production as a practical and effective means of mass assembly of auto-sembly circuitry. Although variations in fluxes and solder compositions are employed in different locations, in general the flux consists of an activated rosin-alcohol solution combined
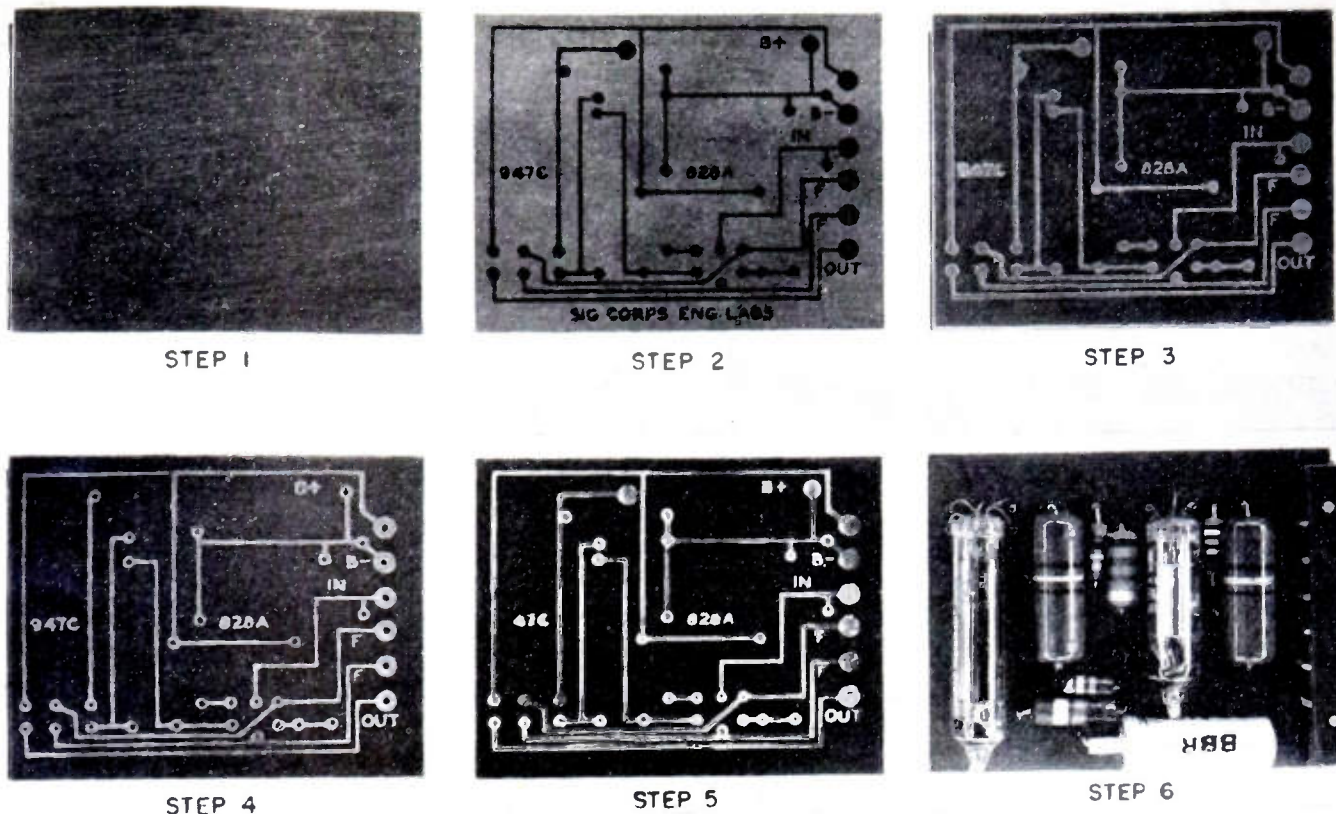
Fig. 1—Process steps in auto-sembly fabrication using etched circuitry. Step 1—Copper surface (foil laminated to plastic chassis). Step 2—Conductor pattern outlined with acid-resistant ink. Step 3—Etched metal-plastic laminate with resist removed. Step 4—Plastic chassis after perforation. Step 5—Underside of chassis after component mounting and solder-dip assembly. Step 6—Top side of assembly showing component layout.

with an oil or wax dissolved in an appropriate solvent; 63–37-per cent (eutectic) tin-lead solder maintained at about 450° F is used as the dipping bath.
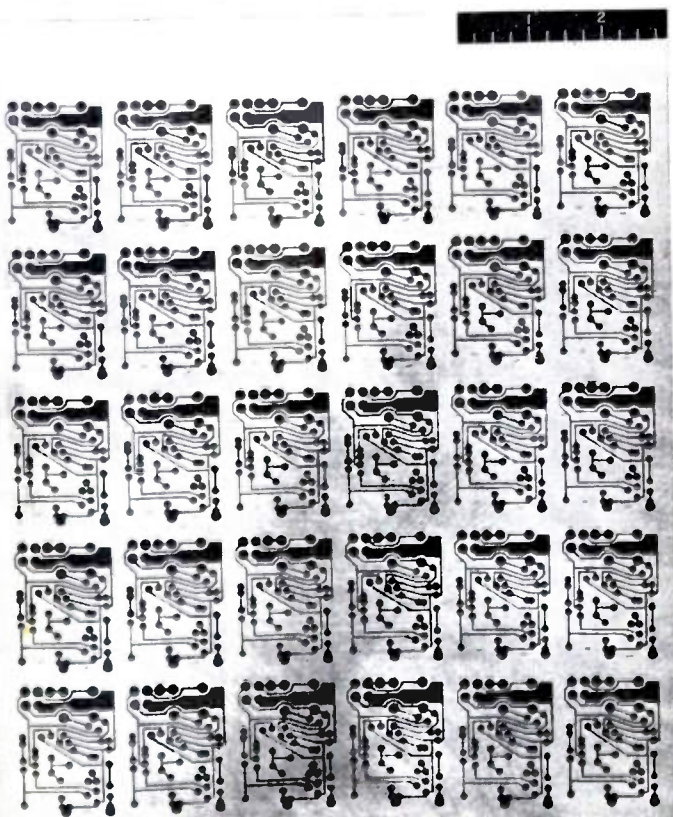
Fig. 2—Printed multipattern sheet prior to etching and division into individual cards.

III. HEAT DISSIPATION

As smaller and smaller components are used in equipments and the equipments themselves are made more compact, a limit to the degree of miniaturization tends to be set by the cooling problem; that is, the problem of transferring the heat generated by the circuit elements to some outside heat sink begins to assume major proportions. For a given circuit, power dissipation in watts per square inch at the surface increases as size is reduced. Above approximately 1 watt per square inch, simple radiation and free convection cooling are no longer adequate. This condition is further aggravated by having other units of the same equipment mounted in close proximity to each other so that each unit package is effectively placed in an ambient temperature that is considerably higher than the external ambient for the completed equipment. This situation necessitates auxiliary means of cooling, such as forced air, liquid cooling with a heat exchanger, evaporative cooling, and so on. It is the additional size, power requirements, and complexity introduced by such auxiliary cooling means which serve to limit the degree and effectiveness of a particular miniaturization effort.

The prime variable to which the equipment engineer must direct his attention is the total power dissipation of the circuit. It is here that the transistor can make its greatest contribution. With a total power dissipation several orders of magnitude lower than that of the electron tube, the transistor permits a degree of miniaturization which was hitherto unobtainable where heat-dis-
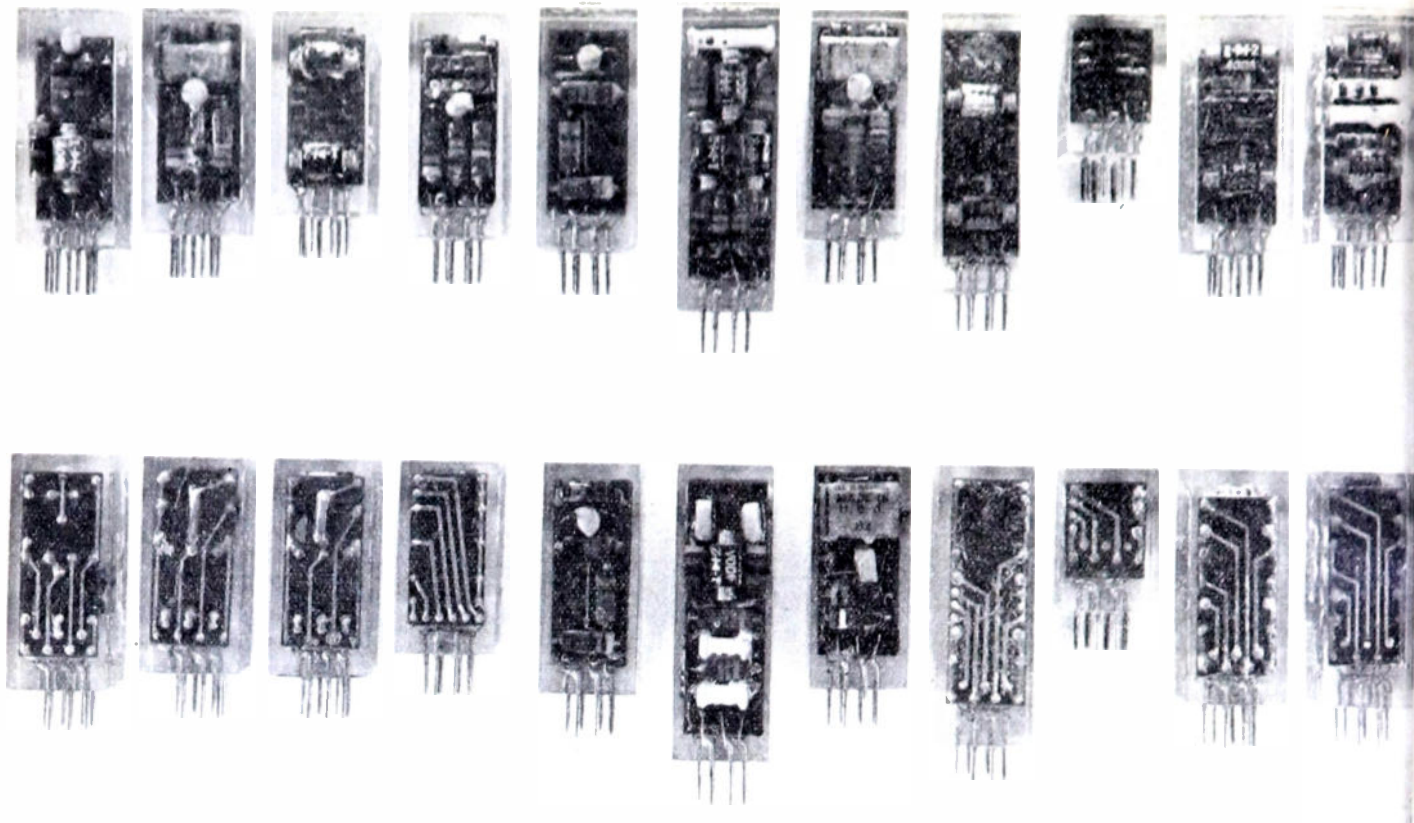
Fig. 3—Bell Telephone Laboratories auto-sembled transistor packages.

sipating elements were involved. From a thermal-design standpoint, the transistor is also an active element. The small size of the transistor results in a rate of heat dissipation at its surface (in terms of watts per square inch) which still equals or exceeds that of many subminiature tubes. Proper consideration must therefore be given the heat dissipation from the transistor in any particular circuit subassembly. In essence, substitution of transistor circuitry for subminiature tubes (assuming equivalent electrical performance characteristics or compensating advantages) will obviate the cooling problem in a given subassembly. Conversely, advantage may be taken of this fact to permit redesign for greater miniaturization as long as the heat problem does not again assume a controlling position.

IV. Auto-Sembled Transistor Applications

Fig. 3 shows a series of computer packages designed and assembled by Bell Telephone Laboratories under a Signal Corps sponsored research and development. Auto-sembly was selected for these packages as being particularly applicable to fabrication of the compact circuitry desired while permitting the use of subminiature components of conventional type and of proven performance and reliability. Also the ease with which transistors, diodes, and other special components could be integrated into the packages by auto-sembly was of paramount importance. The etched-foil prefabricated conductor pattern is shown in many of the views of the lower row which are the backs of the corresponding units above. Enlarged view of one of these circuits (a transistor binary counter) is shown in Fig. 4; two

transistors and a central diode are located at the end opposite the plug.

It may be noted that the Bell Laboratories' packages were cast in resin for purposes of protection and ease of visual examination and display. One disadvantage of
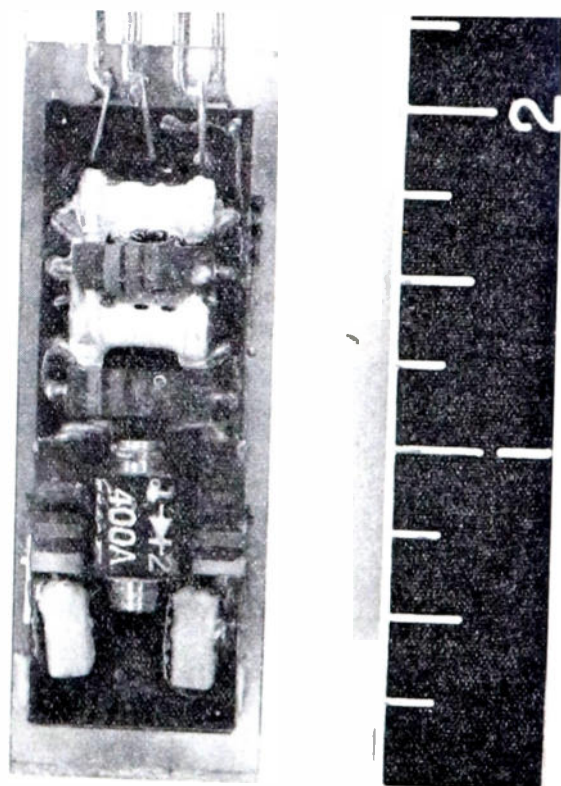
Fig. 4—Enlarged view of a Bell Telephone Laboratories binary counter.

asting resin, particularly for portable equipments, is that it adds considerable weight. The approach presently being investigated at the Signal Corps Engineering Laboratories involves dip coating with a suitable thermo-setting resin which will provide the required
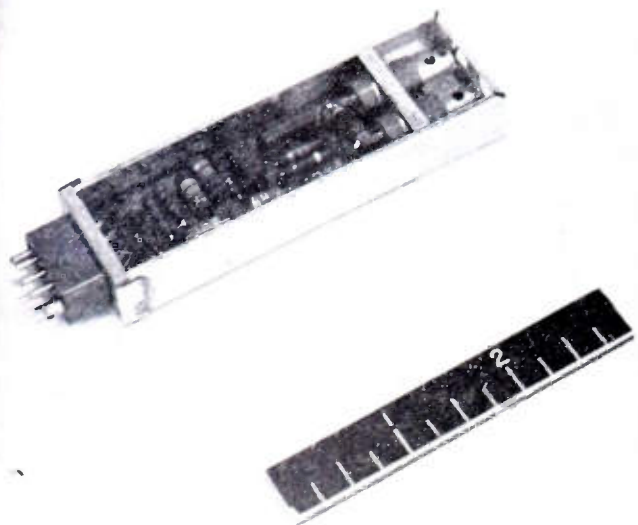
Fig. 5—Two-transistor stabilized binary counter, auto-sembled at the Signal Corps Engineering Laboratories.

noisture protection. The assembly of subminiature components on the prefabricated circuit base has been found to be sufficiently rugged with regard to shock and vibration. The addition of resin coating will provide increased rigidity for these components.
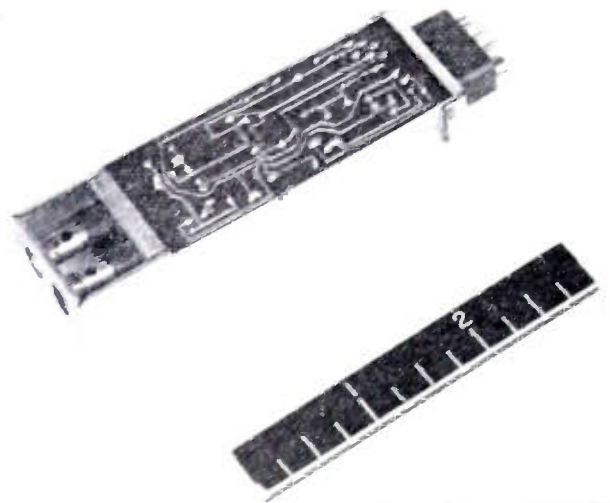
Fig. 6—Back view of binary counter of Fig. 5.

Figs. 5 and 6 show a similar type package, an experimental stabilized general-purpose, two-transistor binary counter, auto-sembled at the Signal Corps Engineering Laboratories using two plug-in point-contact Western Electric transistors. The circuit is described[1] by Trent

[1] R. L. Trent, "The Transistor," Bell Telephone Laboratories, Murray Hill, N. J. Selected reference material on characteristics and applications, Signal Corps Contract DA 36-039 SC-5589 (Task 3); November 15, 1951.

of Bell Telephone Laboratories. The simplicity and compactness of the assembly are readily apparent. It may also be noted in this instance that the ultimate in miniaturization was not effected. Obvious opportunity exists to reduce the size still further by utilizing two shorter cards of prefabricated circuitry with the resistors and diodes on each card staggered opposite each other in the upper portion of the assembly. Thus the subassembly might readily be foreshortened by some 25 per cent over the model shown. The frame of this package facilitates insertion and withdrawal from the equipment. Protection against dirt and moisture is provided by a thin coating of a thermosetting resin.

For assemblies of greater complexity, the auto-sembly technique of successive "decking" permits the fabrication of an extremely compact unit without the myriad difficulties that beset conventional hand-wired production of such items. Fig. 7, a subminiature video
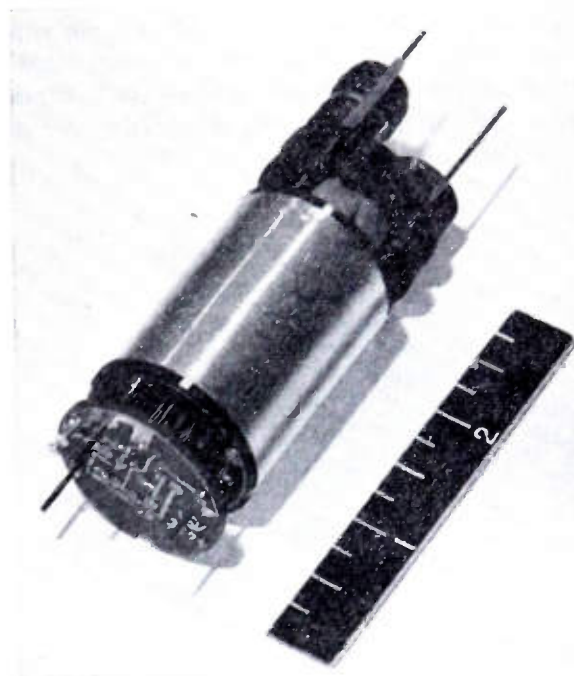
Fig. 7—Crystal video receiver using successive decking technique.

receiver using tubes, is an illustration of this decking technique. The central machined cylinder serves as tube shield and thermal conductor. An appreciation of the deck technique itself may be had by considering the exploded view (Fig. 8). The tubes and associated components are first assembled and dip soldered to their respective decks of circuitry. These assembled decks are then inserted into the succeeding decks and soldered with the individual components that mount on those decks. This particular assembly employs four dual triode subminiature tubes and 31 components including a delay line in a package 1½ inches in diameter and 3½ inches long. The further size advantage of at least a third which could be gained by the substitution of transistors for the tubes in such a circuit is obvious.

A decade counter consisting of 63 components auto-
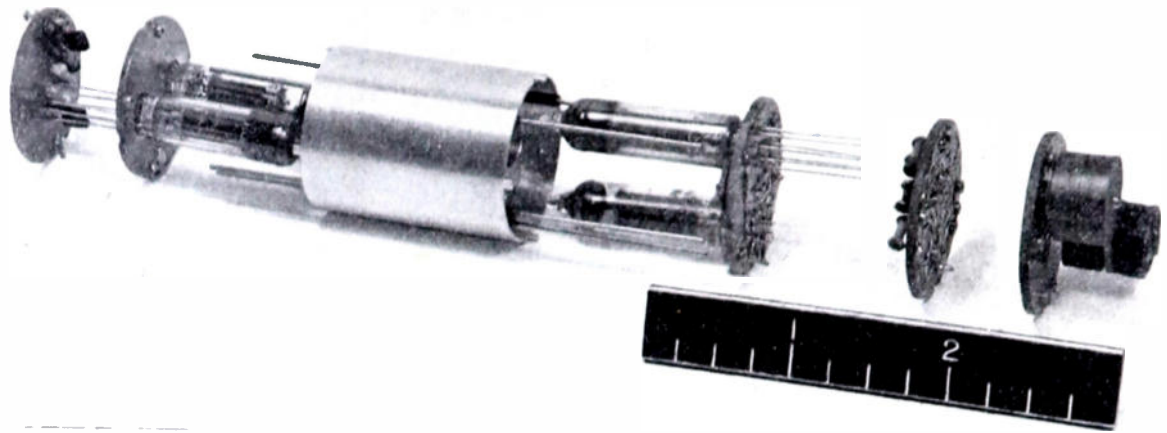
Fig. 8—Exploded view of video receiver of Fig. 7.

sembled on one deck is illustrated in Fig. 9. This counter uses four subminiature twin triodes and a heat collector plate (not shown) which are mounted on a second deck above the component card, roughly doubling the volume of the counter. Here again almost 40-per cent saving in the over-all volume could be effected by using transistors mounted on the lower deck with the other components, thus eliminating the need for a tube deck entirely.
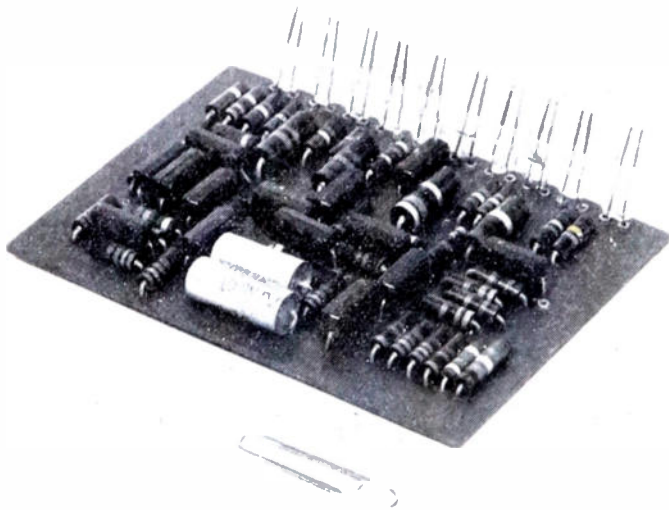
Fig. 9—Auto-sembled decade counter less tube deck.

V. ASSEMBLY ADVANTAGES OF THE TRANSISTOR AND ASSOCIATED COMPONENTS

Transistors are extremely rugged as regards shock, vibration, and microphonics. In addition, while no positive statements are being made as to the ultimate life expectancy of the transistor, extrapolation of aging curves are reported to indicate a life of the order of 75,000 to 100,000 hours. With promise of such performance, the need for sockets to facilitate transistor replacement may well be eliminated and transistors may be auto-sembled directly to the conductor pattern in the same manner as the other accepted long-life components.

The distinctive circuit requirements of the transistor are in turn leading to the development of a specialized family of components. The physical sizes of such components can only be discussed in terms of desired objectives at this time, but very substantial size reductions are expected in view of the low voltage (50 volts or less) and low current (a few milliamperes) circuit requirements. The problems of manipulation, mounting, and assembly of such ultra-small components into circuits will be fully considered in establishing the actual physical design of the component and its terminations.

Transistor circuitry appears to pose no particular problem in laying out the conductive pattern since currents and voltages are low and frequencies of operation are limited to about 20-mc maximum. Auto-sembled circuits operating at voltages up to 300 volts, at frequencies of several hundred megacycles, and at currents up to 5 amperes (with reasonable conductor dimensions) have been successfully fabricated; accordingly, no peculiar problems are expected in the transistor applications of auto-sembly.

VI. CONCLUSIONS

The transistor and the anticipated transistor family of components presage new concepts and new techniques in production fabrication of circuitry. The compatible auto-sembly approach, with its extreme flexibility in application and simplicity in operation, offers a new production philosophy which will lead to lower cost, lower maintenance skills, and higher production benefits both for the military and the industrial electronic industry.

BIBLIOGRAPHY

1. S. F. Danko, "New Developments in the Auto-Sembly Technique of Circuit Fabrication," Signal Corps Engineering Laboratories, Fort Monmouth, N. J. Paper presented at National Electronics Conference, Chicago, Ill.; October 24, 1951. (See Conference Proceedings.)
2. R. L. Swigget, "Printed circuits on foil-clad plastics," *Mod. Plas.*, vol. XXVIII, pp. 99–100; August, 1951.
3. S. F. Danko and S. J. Lanzalotti, "Auto-sembly of miniature military equipment," *Electronics*, vol. 24, pp. 94–98; July, 1951.
4. "A Study of Automatic Unitized Printed Circuit Techniques," Herlec Corp., Signal Corps Contract W36-039 SC-38153; Final Engineering Report; February 15, 1951.

Transistors in Airborne Equipment[*]

O. M. STUETZER†, ASSOCIATE, IRE

Summary—Discussed in this paper are methods and results of a semitheoretical survey made to estimate what would be gained at the present state of the art by using transistors instead of tubes in standard airborne equipment.

THE PROBLEM

SOME MONTHS AGO transistor types became available with performance tolerances not inferior to unselected vacuum tubes. From an exciting toy the transistor had developed into a technical device.

Several properties make this device especially attractive as an active element in airborne equipment. Its size (0.001 to 0.02 cubic inches) is one to two orders of magnitude below that of the presently used vacuum tubes, and so is its weight. Even the point-contact types stand shocks of several thousand g, and vibrations of 100 g do not affect them. The average life is 70,000 hours, and for junction types this figure will surely be exceeded. Considerable savings in power requirements cut down the dreaded heat dissipation in equipment.

In this situation a survey seems to be in order of what advantages and savings we could expect at the present time by using transistors instead of tubes in all possible places in airborne equipment. ("Transistorisation.")

This paper intends to report methods and results of such an analysis without going into details, which are admitted to be controversial in some cases. The results, covering only the present situation, are necessarily of limited importance.

LIMITATIONS OF APPLICATION

The discussion is based on the Bell Telephone Laboratories' transistor types listed in Table I. The values mentioned are averages, and are applicable only for one certain method of operation. Impedance and cutoff frequency tolerances are 30 per cent, those of the current gain 20 per cent, and noise figures stay within 3 db.

Aside from the limitations in frequency response and noise behavior, it has to be remembered that above 80°C the germanium devices cease to function properly. The power-handling capacity of the transistors in Table I is about 100 mw in class A operation, and about 400 mw in class B; 3 watts peak are delivered by the switching units.

For transistorization of equipment it must be kept in mind that transistors cannot simply replace tubes in a given system; the system using tubes must be replaced by a system using transistors. There is at present no sufficient number of personnel available possessing the experience and skill to attempt this task for the whole airborne equipment complex. For the present estimate we have to use the present vacuum-tube stages as system blocks and analyze how many of these can be replaced by transistor blocks performing the same function. This procedure will not give the optimum possible.

ANALYSIS OF AIRBORNE EQUIPMENT STAGES

Counting stages with dual-purpose tubes twice, it was possible to obtain the following figures of each category, as shown in Table II.

TABLE II
PERCENTAGE OF STAGES SUITABLE FOR TRANSISTORIZATION

Communication and navigation equipment	52%
Fire-control equipment	72%
Radar equipment	40%
Special equipment	—
Weighted average	54%

This means only that the above percentage of tubes have functions for the performance of which the above-mentioned transistor types would be sufficient.

It does not imply that in all these cases each tube stage could be replaced with just one transistor stage. In some cases the transistor circuitry involved would be unreasonably complicated, in other cases the operation of the transistor would be very critical. Examples are stages in close proximity to high-dissipation tubes, stages working close to the frequency cutoff of present transistors, or medium-frequency preamplifier stages.

A closer study of one representative example out of each equipment category shows that, to take account of these difficulties, the above figure of 54 per cent should be reduced to 40 per cent of the stages. For these, transistorization would be technically sound at present.

TABLE I
PRESENTLY AVAILABLE OR ANNOUNCED TECHNICAL TRANSISTORS

Type	Special application	R_{in} R_{out} (ohms)		α	Gain (db)	f_{cutoff} (mcps)	Noise F (db)	
1689 1698	Switching	800	10,000	1.5–2			48	
	Low voltage operation	200	8,000	2.2	18	3	54	
1725 1729	Gen purpose.	200	15,000	2.5	20	5		
1734	Video High-speed switching. IF				20 45	20 1	15	NPN-junction. polarities reversed
1752	Audio	200	100	1				
1768	Low drain	150	7,000	2.4	21	1	43	

Savings Through Transistorization

With 40 per cent of the equipment stages transistorized, the limit for savings possible at present would be 40 per cent in size and weight, to be reached only if transistor stages are negligibly small and lightweight compared with the present stages. This, although sometimes assumed, is not the case.

Table III gives some figures of representative characteristic values for "average" stages, obtained by averaging a number of samples selected at random. The most suitable components obtainable at present were used for transistor stages.

TABLE III
Average Data for Tube and Transistor Stages

Stage containing	Weight	Size	Power required	Average life at room temp in light	Average life at room temp on ground
Standard tube	120 g	200 ccm	3 watts	100 hours	5,000 hours
Miniature tube	55 g	60 ccm	2.5 watts	100 hours	5,000 hours
Subminiature tube	35 g	40 ccm	2 watts	500 hours	5,000 hours
Transistor	15 g	10 ccm	.2 watts	75,000 hours	75,000 hours

The weight and size figures are for components only; racks, cables, blowers, and the like are not included. The in-flight life data[1] for vacuum tubes, depressing as they may look, are rather optimistic.

Weight, size, and power figures for transistor stages will be recognized to be much more unfavorable than basically required. They reflect the fact that in actual transistor circuits the theoretical optimum is very rarely reached. As an example, power and components have to be given away to approximate the constant current bias source required for the emitter in amplification circuits. Or, the basic oscillator circuit uses only four components. For an audio oscillator with low power drain and prescribed generator impedance, 8 components are needed.

It is not surprising that the figures, being averages, are found to be half an order of magnitude wrong in special cases.

The figures can now be used to estimate the savings obtainable in military airplanes with "average" equipment. Table IV gives the results. It shows that the savings in size will be larger than those in weight.

TABLE IV
Average Savings by Transistorization

	Weight	Size
Standard stages	35%	38%
Miniature stages	30%	33%
Subminiature stages	22%	30%
Weighted average	28%	33%

The figures are a little optimistic due to the fact that the few final power stages containing large tubes (e.g., magnetrons) are not included in our average data. An estimate shows, however, that the effect is small, the number of tubes in an average airplane being very large.

[1] Oral communication by L. L. Gibbs.

Of much greater importance seems to be that racks, cables, and wiring do not seem to reduce at the same rate as the components with the presently available materials. Although only very tentative figures could be collected on the size and weight of racks and wiring, it is believed that the results in Table IV are finally reduced to about 20-per cent savings in weight and 25 per cent in size.

The savings in power to be furnished or dissipated are easily estimated from Table III. It is roughly the number of stages transistorized times 2.5 watts. Welcome as this may be, it does not mean too much if we keep in mind that the power stages will be unaffected by transistorization and that inverters work with about 20-per cent efficiency. Thus, the batteries and generators in a plane will not change appreciably. However, in such parts of the equipment, where the transistors are in the majority, the cooling problem will be substantially alleviated.

A substantial and important saving becomes obvious when we consider the lifetime figures: The failures in a given time should be reduced by almost 40 per cent. This higher reliability will cause corresponding savings in maintenance.

Conclusions

A semitheoretical survey shows that with present transistors and components conventional airborne equipment could be reduced about 20 per cent in size, 25 per cent in weight, and would have 40 per cent fewer failures.

These are average values. For certain types of equipment they will be radically different. So they would be much more favorable for data link systems using computer technique. For guided missiles, on the other hand, the necessary refrigeration and the influence of cosmic rays darkens the picture.

The savings obtainable at present are quite substantial; still they are somewhat short of certain expectations arrived at from the performance figures of the transistor alone. This is, of course, due to the dominant influence of the remaining vacuum-tube stages. The more the device will be improved in the future to take over from vacuum tubes, the more rapidly the present picture will change in favor of transistorization. Another important contribution will be made by components to be developed especially for use in transistor circuits.

Transistor Trigger Circuits[*]

A. W. LO†, ASSOCIATE, IRE

Summary—This paper concerns trigger and pulse circuits for transistors having emitter-to-collector current gain (such as point-contact units). The circuits are designed to permit reliable operation with transistors which are not completely uniform; they also allow reasonable variation of circuit parameters or bias voltages. Quantitative analysis is made possible by use of simplified circuit theory which divides the nonlinear characteristic of a transistor and its associated feedback and external resistances into quasi-linear regions. By this analysis, one can predict both the type of operation, i.e., monostable, bistable, and astable (oscillatory), and the amplitude and waveform of the output.

A basic monostable circuit consists of a single transistor with a resistance in the base lead to provide feedback and a capacitance or, preferably, a transmission line as the emitter load. This circuit can be used to regenerate periodic or nonperiodic pulses, thereby providing a standardized output pulse shape, or to generate single pulses when initiated; it is possible to provide very intense short pulses (up to one ampere) even with transistors of low power-handling ability. The output pulses may be arbitrarily delayed with respect to the input, and amplitude discrimination against noise or spurious signals is also possible.

Bistable circuits also use a single transistor with external resistances. It is shown that lack of reliability of previously used bistable single-transistor circuits can be overcome by proper arrangement of circuit parameters and bias supplies and, in some cases, by use of a nonlinear resistance (crystal diode) as the emitter load. Laboratory experience indicates that such single-transistor circuits can be made highly reliable and will allow interchangeability of transistors and reasonable variations in power supplies, and so on. In these respects, they have been made superior even to earlier twin-transistor bistable circuits in which reliability had been emphasized. A small, light-weight decade counter was designed with these single-transistor bistable circuits; it uses less than 2 watts of power in contrast to a well-designed electron-tube counter which needs 60 watts, uses twice as many amplifier elements, and requires considerably more weight and space.

INTRODUCTION

CIRCUITS EMPLOYING a single point-contact transistor (or other transistor with emitter-to-collector current gain greater than unity) together with passive circuit elements may be made to have two distinct stable states separated by an unstable negative-resistance region.[1,3] The circuit can be made to trigger from one stable state to the other in a very short time by the application of an input pulse. The trigger action suggests the use of this type of transistor in circuits where the output is to be controlled by the presence of an input pulse and where the faithful reproduction of the input signal is not required.

Earlier investigations in this field resulted in the development of a number of novel circuits. Laboratory experience, however, revealed that the operation of most transistor circuits was so highly critical that their applications in systems where reliability was of

primary importance is limited. Transistors presently available match much more closely in their characteristics than early units, but are still not completely uniform; temperature changes also affect their characteristics. In this paper, special consideration has been given in these respects to the design of practical and novel transistor circuits so as to achieve reliability which has not been previously attained.[1-8]

Part I presents a simplified analysis of transistor trigger circuits which will assist in the understanding of their operation and which will serve as a guide for the design of practical circuits. Part II includes a number of practical pulse-handling circuits, the satisfactory operation of which is not impaired by reasonable variations of circuit parameters, bias voltages, transistor characteristics, and ambient temperature.

PART I. A SIMPLIFIED ANALYSIS OF TRANSISTOR TRIGGER CIRCUITS

Basic Trigger Circuit

Some simple circuits employing a single transistor may perform trigger action by virtue of the current-amplification within the transistor and the positive feedback afforded by base resistance (both internal and external). A rigorous analysis is difficult because of the high degree of nonlinearity involved. However, certain approximations may be made to simplify the analysis without loss of effectiveness.[6] The following simplified analysis, derived from experimental results, may serve effectively to explain and describe the operation of these circuits and to guide in their design.

A basic trigger circuit is shown in Fig. 1.[1,3] The operation of this circuit may be monostable, bistable, or astable (oscillatory), depending on the nature of the emitter load. For practical purposes the behavior of the circuit is adequately described by the two character-

[1] W. M. Webster, E. Eberhard, and L. E. Barton, "Some novel circuits for the three-terminal semiconductor amplifier," *RCA Rev.* vol. 10, pp. 5-16; March, 1949.
[2] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, pp. 367-400; July, 1949.
[3] H. J. Reich and R. L. Ungvary, "Transistor trigger circuits," *Rev. Sci. Inst.*, vol. 20, pp. 586-588; August, 1949.
[4] E. Eberhard, R. O. Endres, and R. P. Moore, "Counter circuits using transistors," *RCA Rev.*, vol. 10, pp. 459-476; December, 1949.
[5] H. J. Reich and P. M. Schulthersis, "Some transistor trigger circuits," PROC. I.R.E., vol. 39, pp. 627-632; June, 1951.
[6] A. E. Anderson, "Some switching aspects of transistors," The Transistor—Selected Reference Material on Characteristics and Applications, Bell Tel. Labs., Inc.; 1951.
[7] E. A. Oser, R. O. Endres, and R. P. Moore, "Transistor Oscillators," to be published.
[8] "The Transistor—Selected Reference Material on Characteristics and Applications," Bell Tel. Labs., Inc.; 1951.
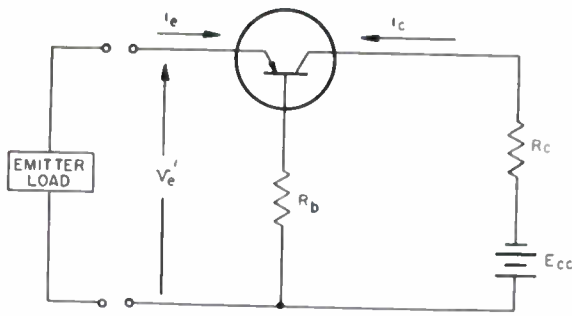
Fig. 1—Basic trigger circuit.

istics[9] shown in Fig. 2: (a) the emitter input characteristic (v_e'–i_e curve), and (b) the current transfer characteristic (i_c–i_e curve). The v_e'–i_e curve, together with the emitter load line, dictates the mode of operation of the circuit, while the i_c–i_e curve determines the
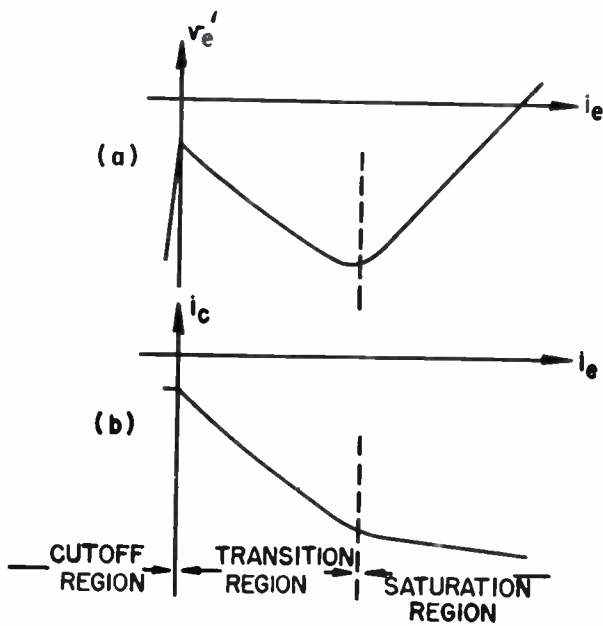
Fig. 2—Trigger circuit characteristics. (a) Emitter input characteristic. (b) Current transfer characteristic.

form of the output signal. Considering i_e as the independent variable, each curve exhibits three distinct regions, namely, a low-conduction "cutoff" region, a negative-resistance "transition" region, and a high-conduction "saturation" region. The curves in each region do not deviate appreciably from straight lines, except near the boundaries. This property indicates that the circuit may be represented by three equivalent circuits, one for each region.

Following conventional notation, the transistor may be represented by its internal resistances r_e, r_b, and r_c, and a current source αi_e. The value of r_b is usually small compared with external base resistance R_b. In the cutoff region, where $i_e < 0$, the transistor is characterized by its

[9] These characteristics are of the whole circuit, including the transistor and the external circuit parameters.

high r_c and zero current-amplification. The equivalent circuit for this region is shown in Fig. 3. The emitter
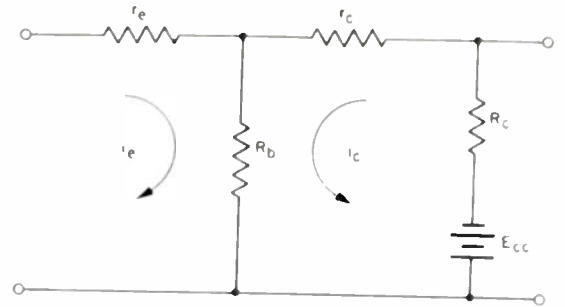
Fig. 3—Equivalent circuit—cutoff region.

input characteristic for the cutoff region is represented by

$$v_e' \cong i_e r_e - \frac{R_b}{R_b + R_c + r_c} E_{cc}, \tag{1}$$

whose slope and intersection are, respectively,

$$\frac{\partial v_e'}{\partial i_e} \cong r_e$$

$$v_e' = -\frac{R_b}{R_b + R_c + r_c} E_{cc}, \quad \text{at} \quad i_e = 0. \tag{2}$$

The current transfer characteristic for the cutoff region is represented by

$$i_c = -\frac{R_b}{R_b + R_c + r_c} i_e - \frac{E_{cc}}{R_b + R_c + r_c}, \tag{3}$$

or

$$\frac{\partial i_c}{\partial i_e} = -\frac{R_b}{R_b + R_c + r_c}$$

$$i_c = -\frac{E_{cc}}{R_b + R_c + r_c}, \quad \text{at} \quad i_e = 0. \tag{4}$$

In the transition region, where $r_e \ll R_b$ and R_c, and $\alpha > 1$, the equivalent circuit is shown in Fig. 4. The emitter input characteristic for the transition region is

$$v_e' = \frac{R_b[R_c + r_c(1 - \alpha)]}{R_b + R_c + r_c} i_e - \frac{R_b}{R_b + R_c + r_c} E_{cc}, \tag{5}$$
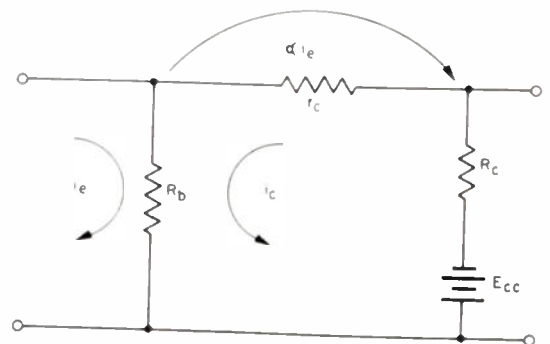
Fig. 4—Equivalent circuit—transition region.

or

$$\frac{\partial v_e'}{\partial i_e} = \frac{R_b[R_c + r_c(1 - \alpha)]}{R_b + R_c + r_c}$$

$$v_e' = -\frac{R_b}{R_b + R_c + r_c} E_{cc}, \quad \text{at} \quad i_e = 0. \tag{6}$$

The current transfer characteristic for the transition region is

$$i_c = -\frac{R_b + \alpha r_c}{R_b + R_c + r_c} i_e - \frac{E_{cc}}{R_b + R_c + r_c}, \tag{7}$$

or

$$\frac{\partial i_c}{\partial i_e} = -\frac{R_b + \alpha r_c}{R_b + R_c + r_c}$$

$$i_c = -\frac{E_{cc}}{R_b + R_c + r_c}, \quad \text{at} \quad i_e = 0. \tag{8}$$

In the saturation region the transistor behaves approximately like a conductor, suggesting the approximate equivalent circuit shown in Fig. 5. The emitter in-
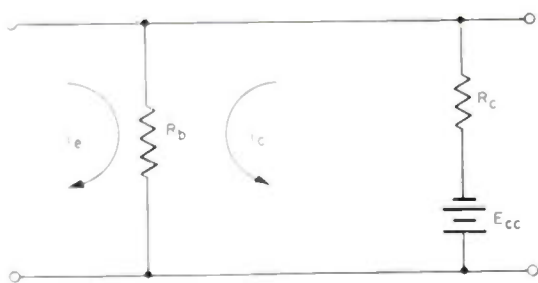
Fig. 5—Equivalent circuit—saturation region.

put characteristic for the saturation region is, approximately,

$$v_e' = \frac{R_b R_c}{R_b + R_c} i_e - \frac{R_b}{R_b + R_c} E_{cc}, \tag{9}$$

or

$$\frac{\partial v_e'}{\partial i_e} = \frac{R_b R_c}{R_b + R_c}$$

$$i_e = \frac{E_{cc}}{R_c}, \quad \text{at} \quad v_e' = 0. \tag{10}$$

The current transfer characteristic for the saturation region is, approximately,

$$i_c = -\frac{R_b}{R_b + R_c} i_e - \frac{E_{cc}}{R_b + R_c}, \tag{11}$$

or

$$\frac{\partial i_c}{\partial i_e} = -\frac{R_b}{R_b + R_c}$$

$$i_c = -\frac{E_{cc}}{R_b + R_c}, \quad \text{at} \quad i_e = 0. \tag{12}$$

The assumption that the transistor acts like a perfect conductor in the saturation region is an approximation suited for a first approach. Actually, the values of r_e, r_b, and r_c, at high conduction are small but not negligible, and they do not remain constant with
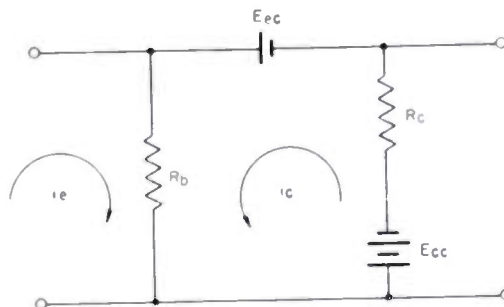
Fig. 6—Modified equivalent circuit—saturation region.

changes of i_e. Laboratory investigation, however, reveals the existence of a small voltage between emitter and collector, which is substantially constant over the entire saturation region. For a better approximation, therefore, the equivalent circuit for the saturation region is modified as shown in Fig. 6. The more exact emitter input characteristic for the saturation region is

$$v_e' = \frac{R_b R_c}{R_b + R_c} i_e - \frac{R_b}{R_b + R_c} (E_{cc} - E_{ec}), \tag{9a}$$

or

$$\frac{\partial v_e'}{\partial i_e} = \frac{R_b R_c}{R_b + R_c}$$

$$i_e = \frac{E_{cc} - E_{ec}}{R_c}, \quad \text{at} \quad v_e' = 0. \tag{10a}$$

The more exact current transfer characteristic for the saturation region is

$$i_c = -\frac{R_b}{R_b + R_c} i_e - \frac{E_{cc} - E_{ec}}{R_b + R_c}, \tag{11a}$$

or

$$\frac{\partial i_c}{\partial i_e} = -\frac{R_b}{R_b + R_c}$$

$$i_c = -\frac{E_{cc} - E_{ec}}{R_b + R_c}, \quad \text{at} \quad i_e = 0. \tag{12a}$$

Based upon this simplified analysis, using (1) to (12(a)), the calculated characteristics and experimental results are plotted for a typical circuit, as shown in Figs. 7(a) and (b). Close agreement between the experimental and analytical curves is observed both in the cut-off region and the saturation region. Agreement
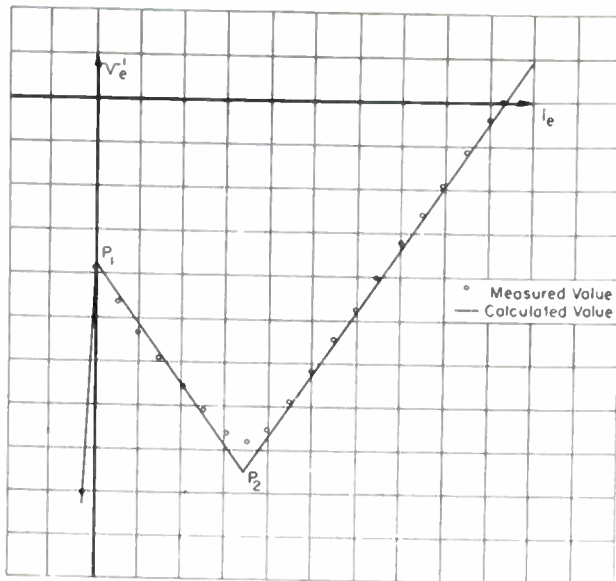
(a) Emitter input characteristic.
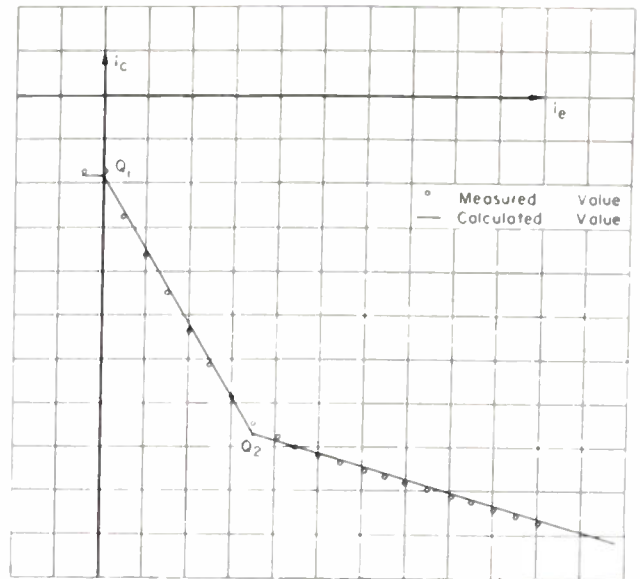
Fig. 7

(b) Current transfer characteristic.

is not expected, of course, where the observed curve departs from linearity. In the transition region the curves depart considerably from a straight line. This is due to the fact that the values α and r_e do not remain constant with change of i_e as assumed in the analysis. However, in most pulse-handling circuits the exact shape of the curve in the transition region is not important as long as the turn-over points P_1, P_2, Q_1, and Q_2 are well defined. From (1) to (12) the calculated co-ordinates of the turn-over points are

$$P_1: i_e = 0, \qquad v_e' = \frac{-R_b}{R_b + R_c + r_s} E_{cc}. \quad (13)$$

$$Q_1: i_e = 0, \qquad i_c = \frac{-E_{cc}}{R_b + R_c + r_c}. \quad (14)$$

$$P_2: i_e = \frac{E_{cc}}{\alpha(R_b + R_c) - R_b}, \quad v_e' = \frac{R_b(1-\alpha)E_{cc}}{\alpha(R_b + R_c) - R_b}. \quad (15)$$

$$Q_2: i_e = \frac{E_{cc}}{\alpha(R_b + R_c) - R_b}, \quad i_c = \frac{-\alpha E_{cc}}{\alpha(R_b + R_c) - R_b}. \quad (16)$$

The points, P_1 and Q_1, where the circuit is triggered from low to high conduction, depend on the value of r_c, which, unfortunately, is affected considerably by temperature and which also varies from unit to unit. The points, P_2 and Q_2, where the circuit is triggered from high to low conducting, depend mainly on α which decreases with increasing i_e. The calculated co-ordinates of P_2 and Q_2, based on the assumption that α remains constant, depart considerably from the measured values in v_e' but check fairly well in i_e and i_c. This is a favorable situation as far as the design of practical circuits is concerned.

The analysis of this type of circuit is not limited to the emitter characteristic alone; rather it is subject to a like development of the base and collector character-

istics. In general, the approach is valid at any branch in the circuit where negative resistance is exhibited. The emitter input approach offers some advantages in simplicity, however.

Monostable Circuits

A basic monostable circuit is shown in Fig. 8, together with its emitter input characteristic and its cur-
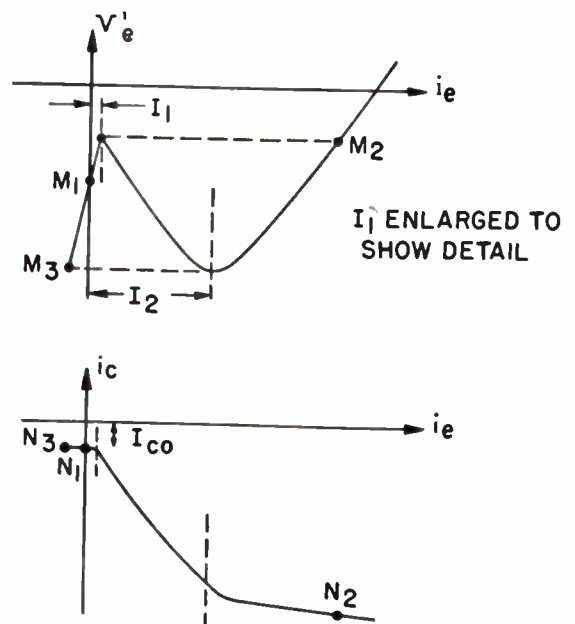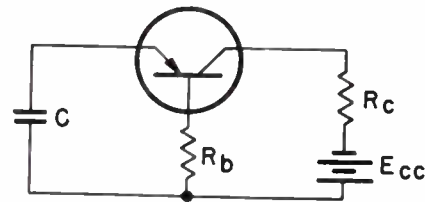
Fig. 8—Basic monostable circuit.

rent transfer characteristic. These two curves may be referred to jointly in the following discussion. The negative-resistance region of the $v_e'–i_e$ curve begins slightly to the right of the v_e'-axis. In the quiescent state the condenser, C, constitutes an open-circuit in the emitter loop. The quiescent operation point is at the point M_1, which is the intersection between the open-circuit load line (the v_e'-axis) and the $v_e'–i_e$ curve. When a trigger source sends into the emitter a positive current greater than I_1, the circuit passes into its unstable negative-resistance region. By virtue of the current-amplification within the transistor and the positive feedback action afforded by the base resistance, the emitter and collector currents increase rapidly, which results in the sudden jump of i_e to the point M_2 in the saturation region. In the process of charging the condenser C in the emitter loop, i_e diminished until its value is less than I_2. At this point the circuit re-enters the unstable region and quickly jumps back to point M_3 in the low-conduction stable state. Eventually, as a consequence of the discharge of capacitor C, the initial quiescent operating point M_1 is attained.

Fig. 9—Equivalent circuit of trigger action.

Referring now to the equivalent circuits of the two states, Figs. 3 and 5, the operation of the circuit is seen to resemble a simple switching problem as shown in Fig. 9. The switches SW_1 and SW_2 are open-circuited in the quiescent state. At $t=0$, corresponding to application of the trigger pulse, both switches are closed. It is readily shown by Laplace-transformation analysis that the emitter and collector currents as functions of time may be expressed as

$$i_e = \frac{E_0}{R_c}\epsilon^{-(1/R_{11}C)t}; \qquad 0 < t < T.$$

$$i_c = -\frac{R_b E_0}{R_c(R_b + R_c)}\epsilon^{-(1/R_{11}C)t} - \frac{E_{cc}}{R_b + R_c}; \qquad (17)$$

$$0 < t < T,$$

where

$$E_0 = \frac{r_c}{R_b + R_c + r_c}E_{cc}. \qquad (18)$$

$$R_{11} = \frac{R_b R_c}{R_b + R_c}.$$

The analysis reveals that i_e jumps from zero to E_0/R_c on application of the trigger pulse, and then decreases exponentially with times at a rate determined by the time constant CR_{11}. When i_e falls below the value I_2, the circuit re-enters the negative-resistance region and triggers back to the low-conduction state. The value of I_2, calculated by equating (5) and (9), is

$$I_2 = \frac{E_{cc}}{\alpha(R_b + R_c) - R_b}. \qquad (19)$$

This second trigger action is equivalent to opening switches SW_1 and SW_2 at $i_e = I_2$. The collector current varies in a like manner (Fig. 10). The amplitude of the
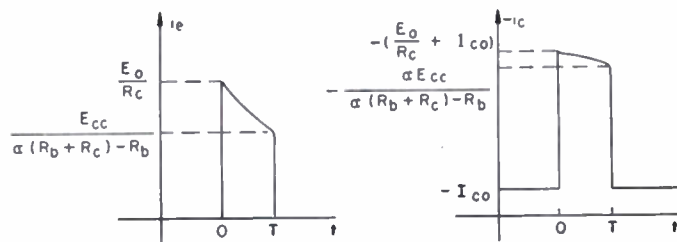
Fig. 10—Emitter and collector current waveforms.

output voltage pulse is approximately E_0 and its width (which is the same as the width of the emitter current pulse) is

$$T = R_{11}C \ln \frac{r_c[\alpha(R_b + R_c) - R_b]}{(R_b + R_c + r_c)R_c}. \qquad (20)$$

Note that the generated pulse width is independent of E_{cc}. Equations (17) and (18) based on the saturation-state equivalent circuit, are valid only in the interval $0 < t < T$. In other words,

$$i_e = \frac{E_0}{R_c}\epsilon^{-(1/R_{11}C)t}, \qquad \text{for } 0 < t < T$$
$$= 0, \qquad \text{for } t < 0, t > T. \qquad (21)$$

$$i_c = \frac{-E_0 R_b}{R_c(R_b + R_c)}\epsilon^{-(1/R_{11}C)t} - \frac{E_{cc}}{R_b + R_c},$$
$$\text{for } 0 < t < T \qquad (22)$$

$$= \frac{-E_{cc}}{R_b + R_c + r_c}, \qquad \text{for } t < 0, t > T.$$

The output pulse departs from the ideal flat-topped pulse according to the exponential term of (18). This departure is more pronounced when R_c is small, which is the case when a high-current pulse is desired. The output pulse width is a function of α and r_c (20) which may vary from unit to unit and which may also be temperature dependent. For precision applications it is desirable to have a circuit whose output waveform is controlled by circuit parameters alone. One such circuit employing a transmission line is shown in Fig. 11.[10]

[10] First investigated by Endres of the RCA Victor Co.

For the following analysis let the emitter load be an open-circuited loss less line with a surge impedance $R_0 = R_b R_c / R_b + R_c$, and of such length that the delay time is $T/2$. Treating the problem in the same manner as was described in connection with its archetype, the emitter and collector currents as functions of time may be expressed as

$$i_e = \frac{E_0}{2R_c}\left[U(t) - (t - T)\right]$$

$$i_c = -\frac{E_0}{2R_c}\left[U(t) - U(t - T)\right] \tag{23}$$

$$-\frac{E_0}{2(R_b + R_c)}\left[U(t) + U(t - T)\right] - I_{c0},$$

where

$$I_{c0} = \frac{E_{cc}}{R_b + R_c + r_c}. \tag{24}$$
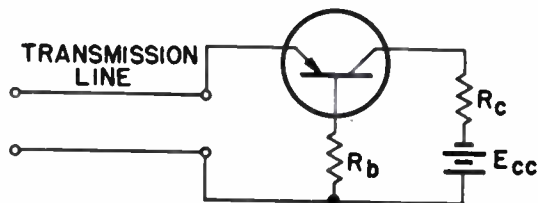
Fig. 11—Transmission-line controlled monostable circuit.

These expressions, based on the equivalent circuit for the saturation region, are valid only when $i_e > I_2$. Since i_e drops from $E_0/2R_c$ to 0 at $t=T$, the expressions are valid only during the time interval $0 < t < T$. The general expressions i_e and i_c are

$$i_e = \frac{E_0}{2R_c}, \qquad \text{for } 0 < t < T$$
$$= 0, \qquad \text{for } t < 0, t > T. \tag{25}$$

$$i_c = -\frac{R_b E_0}{2R_c(R_b + R_c)} - \frac{E_{cc}}{R_b + R_c}, \qquad \text{for } 0 < t < T$$
$$= I_{c0}, \qquad \text{for } t < 0, t > T. \tag{26}$$

Fig. 12—Output waveforms of two monostable circuits. (a) Condenser controlled, transistor no. 1; (b) condenser controlled, transistor no. 2; (c) transmission line controlled, transistor no. 1; (d) transmission line controlled, transistor no. 2.

From (25) and (26) it is evident that the values of i_e and i_c are constant (depending on circuit parameters only) during the time interval $0 < t < T$. The width of the current pulse is controlled solely by the transmission line, which may be precisely designed and built. The advantages of employing a transmission line instead of a condenser as the emitter load of the basic monostable circuit is illustrated in a typical case as shown in Fig. 12.

Bistable Circuit

A basic bistable circuit[4,9] and its operating characteristics are shown in Fig. 13(a) and (b). The emitter load consists of a resistance R_e and a bias voltage E_{ee}.
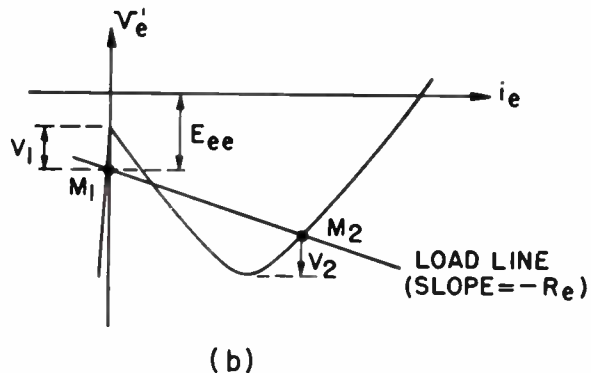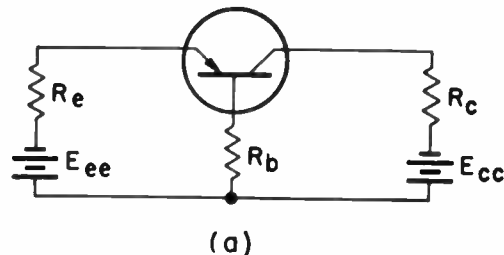
Fig. 13—Basic bistable circuit. (a) Circuit arrangement; (b) Emitter input characteristic

A fundamental requirement for bistable operation is that the emitter load line intersect the v_e'—i_e curve once in each region. The circuit may stay at the low-conduction operation point M_1 (the OFF state), or at the high-conduction operation point M_2 (the ON state). The equivalent circuits of the two states are shown in Fig. 14, and the emitter and collector currents in either
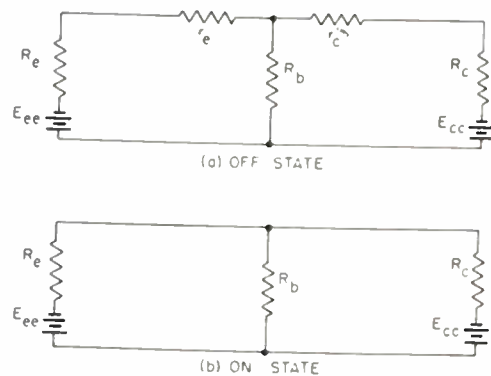
Fig. 14—Equivalent circuits of the bistable circuit.

state may be computed accordingly. Referring to Fig. 13, a positive voltage V_1 may trigger the circuit from

its OFF state to its ON state, and a negative voltage V_2 may perform the reverse operation. The values of V_1 and V_2 may be made small to favor trigger sensitivity, or they may be made fairly large to insure reliable operation in the presence of noise, temperature change, and individual difference between transistor units. In practice, the chosen values of R_e and E_{ee} are a compromise between sensitivity and stability. However, the ultimate limits of the two parameters are

$$R_e < \left| \frac{R_b[R_c + r_c(1 - \alpha)]}{R_b + R_c + r_c} \right|$$

$$\left| \frac{R_b}{R_b + R_c + r_c} E_{cc} \right| < |E_{ee}| < \left| \frac{R_b(1 - \alpha)}{\alpha(R_b + R_c) - R_b} E_{cc} \right|,$$

as is evident from Fig. 13(b), and (5), (13), and (15).

Astable Operation

A simple transistor circuit may perform astable operation if the dc operation point lies in the unstable negative-resistance region and some reactive element is incorporated in the circuit.[1,8,9] In Fig. 15(a), the emitter is current-biased through a high resistance R_e so that the dc operation point is some point P in the negative
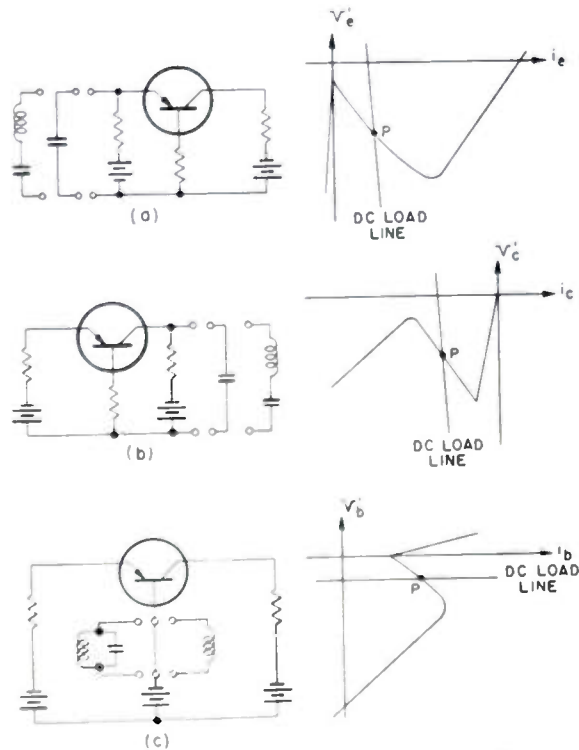
Fig. 15—Basic astable circuits.

resistance region. A condenser C may supply the required reactance for nonsinusoidal RC oscillation, and a series resonant circuit may be employed for sinusoidal wave oscillation. A similar situation exists in the collector circuit (Fig. 15(b)). In the base circuit (Fig. 15(c)), the circuit may be voltage-biased through zero resistance in the base to obtain the proper dc operation point. An inductance or a parallel resonant circuit may be employed to perform the astable operation.

PART II. SOME PRACTICAL TRANSISTOR PULSE-HANDLING CIRCUITS

Monostable Circuits

The basic monostable circuit described in Part I provides a regenerative amplifier with unique application possibilities. A practical pulse standardization circuit
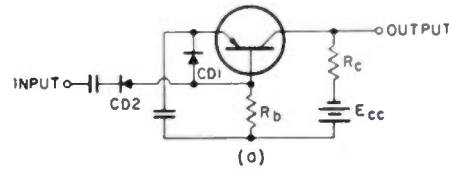
Fig. 16(a)—Pulse standardization circuit. (a) Circuit arrangement.
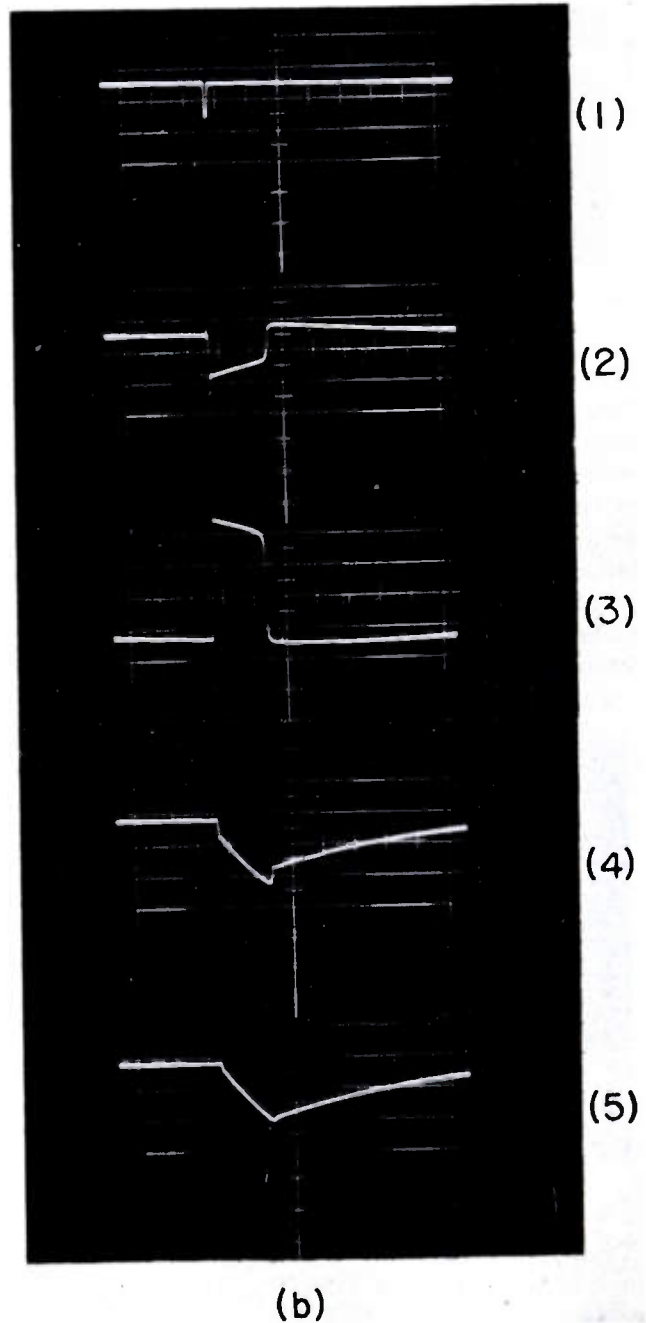
Fig. 16(b)—Pulse standardization circuit. (b) Waveforms. (1) Input pulse. (2) Emitter current. (3) Collector current. (4) Base voltage. (5) Emitter voltage.

is shown in Fig. 16(a). A negative-going trigger pulse, applied at the base of the transistor, initiates the trigger action. Because of the presence of the crystal diode CD_2, the input source is isolated from the transistor circuit except for the short duration when the initiation of trigger action takes place. The output pulse amplitude and width are practically independent of input pulse width and amplitude (Fig. 17).
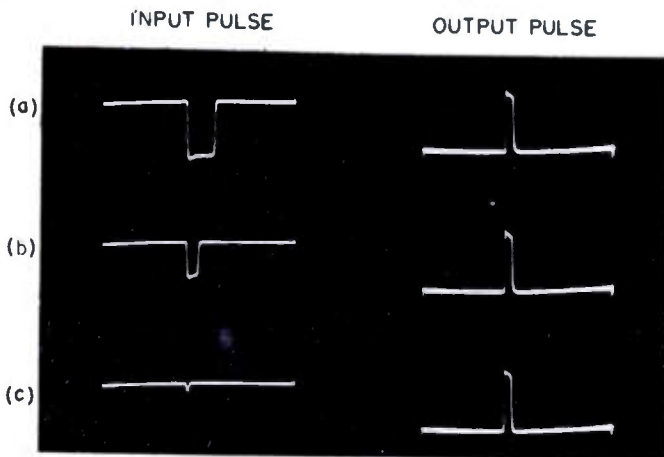
Fig. 17—Input and output waveforms of pulse standardization circuit.

It is observed that after the circuit triggers from its saturation state to its cut-off state the condenser C starts to discharge through the internal emitter resistance r_e, which is high in the quiescent state. This discharge time limits the upper repetition frequency at which the circuit can be operated. To reduce the discharge time, a second diode CD1 (Fig. 16(a)) is placed between emitter and collector to provide a low impedance path for the discharge of the condenser. The effectiveness of this measure is illustrated by the waveform photography of Fig. 18.

Fig. 18—Circuit recovery time, (a) with diode, (b) without diode.

Typical performance of the pulse standardization circuit is summarized in Table I and some output voltage waveforms are shown in Fig. 19.

As the width of the output pulse may be readily prescribed by choosing proper circuit parameters, it is

TABLE I
PERFORMANCE OF PULSE STANDARDIZATION CIRCUIT

	$C=470\ \mu\mu f$	$C=2,200\ \mu\mu f$	$C=0.01\ \mu f$
Output pulse amplitude (volts)	40	40	40
Output pulse width (μs)	0.25	1.1	5
Output pulse rise time (μs)	0.02	0.02	0.02
Output pulse fall time (μs)	0.1	0.1	0.1
Maximum voltage gain	80	80	80
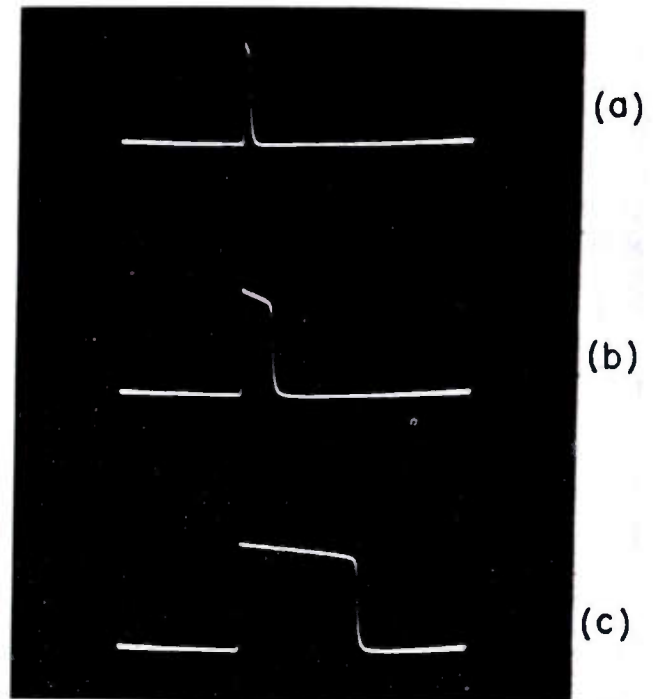
Fig. 19—Output waveforms of pulse standardization circuit.
(a) $C=470$ MMF. (b) $C=2200$ MMF. (c) $C=0.01$ MF.

feasible to utilize the trailing edge of the output pulse to give a delayed pulse. As shown in Fig. 20(a), a peaking coil and a damping diode are arranged in the collector circuit to generate a sharp pulse when the collector current jumps back from its saturation value. Some typical waveforms are shown in Fig. 20(b).

For those applications where a single, manually initiated pulse is desired, the circuit of Fig. 21 will be of particular interest. In switching condenser C into the emitter circuit the charging current of the condenser triggers the circuit, generating a single output pulse.

When the external collector resistance R_c is reduced to a very small value, the monostable circuit is capable of delivering high-current pulses of short duration. This is evident from an inspection of the saturation region equivalent circuit and its analysis. Output pulses of current as high as one ampere (10 volts across a 10-ohm resistance) were observed in such a circuit using ordinary point-contact transistors. The extremely low output impedance of this circuit is of great value in driving low-impedance networks or in operating in a
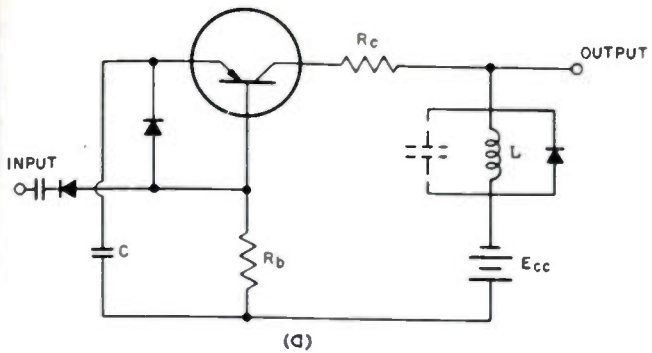
Fig. 20(a)—Pulse-delay circuit. (a) Circuit arrangement.

(b)

(1)

(2)

(3)
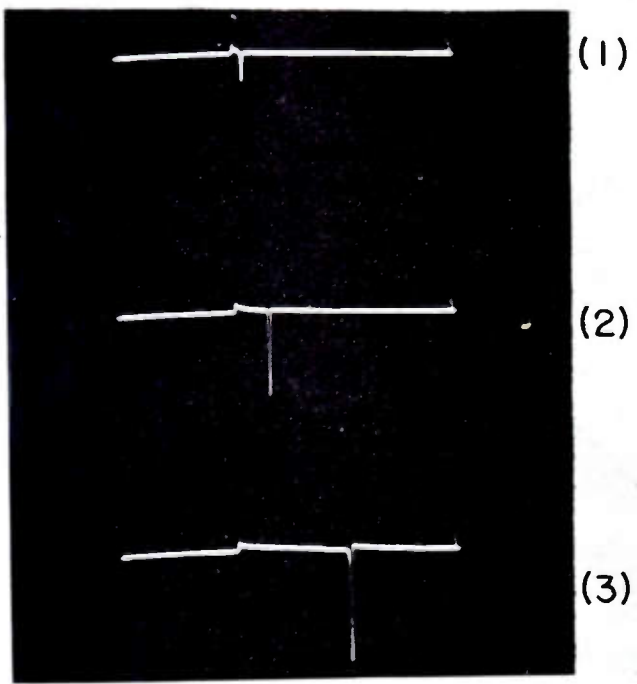
Fig. 20(b)—(b) Output waveforms. (1) $C = 470$ MMF.
(2) $C = 2200$ MMF. (3) $C = 0.01$ MF.

system where distributed capacitance is an important consideration in design.

The power-supply voltage E_{cc} in the basic monstable circuit may be located in the collector and base circuits in order that the dc level of the output circuit might ac-
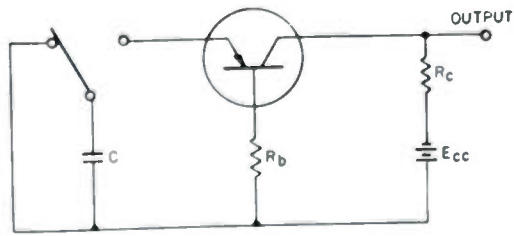
Fig. 21—Single-shot pulse generator.

commodate dc coupling to the other circuits (Fig. 22). This may be found valuable when the monostable circuit is used to drive certain direct coupled circuits.

Frequently, discrimination against spurious signals

is more important than trigger sensitivity. In that event, the emitter of the transistor may be current-biased to insure that the transistor remains in its low-conduction quiescent state until a trigger pulse exceeding a prescribed level is applied. Two current-bias arrangements, and their effects on the emitter input characteristic,
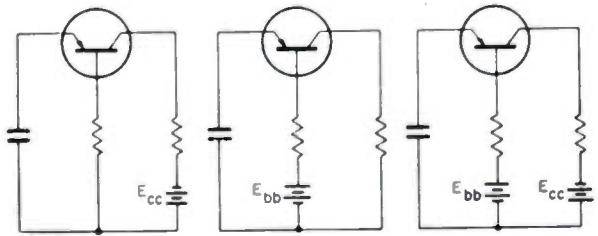
Fig. 22—Location of power supply voltage.

are shown in Fig. 23. The high resistance R, in each case, allows a negative current to flow into the emitter when the transistor is in its quiescent state. The presence of the resistance has little effect on operation of the circuit when the transistor is in its active state.
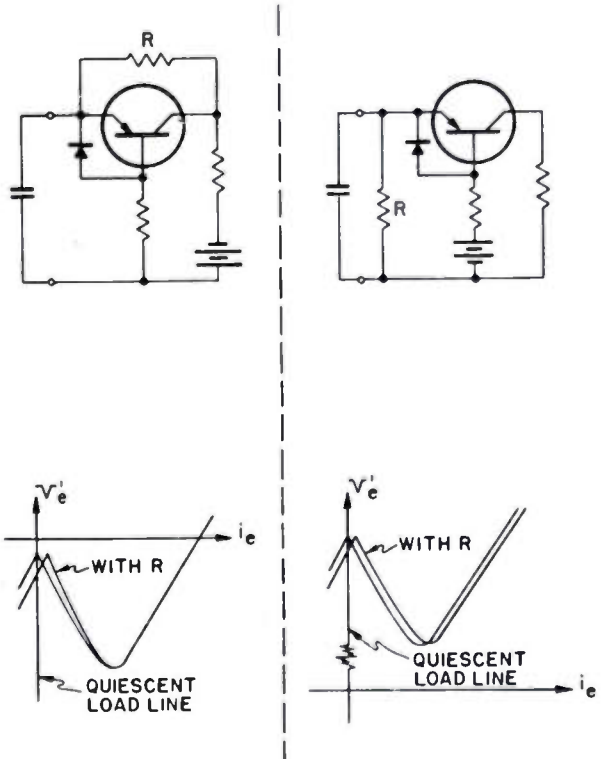
Fig. 23—Effects of emitter current—bias.

As mentioned in Part I, the output waveform of a monostable circuit may be precisely controlled by employing a transmission line, instead of a condenser, as the emitter load. As illustrated in Fig. 12, the transmission-line controlled circuit keeps the width of the output pulse constant, irrespective of individual difference in transistor units. This arrangement is of particular interest in computer systems where pulse width and pulse delay time have to be precisely controlled.

Bistable Circuits

The basic bistable circuit, shown in Fig. 13, requires that the emitter load line intersect the v_e'–i_e curve once in each region.[4],[9] A study of a number of transistors operating in a typical bistable circuit reveals that the v_e'–i_e curve for a given transistor may lie anywhere within the shaded area of Fig. 24(b). In order to insure
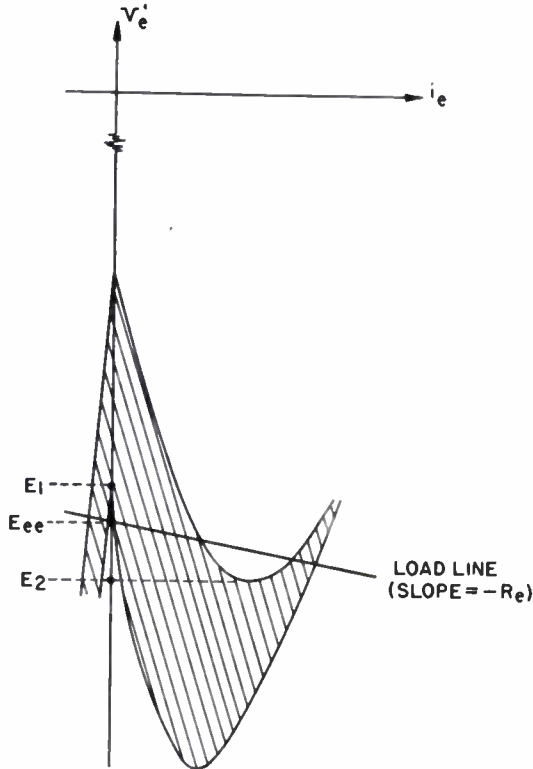
Fig. 24—Emitter input characteristics of basic bistable circuit.

three-point-intersection for all v_e'–i_e curves within the area, R_e must be small and the value of E_{ee} must lie between E_1 and E_2. Since the emitter usually serves as an input terminal for triggering, a low value of R_e implies a low input impedance, while required value of E_{ee} may make circuit unreliable under adverse conditions.

A practical single-transistor bistable circuit using a nonlinear resistance as the emitter load is shown in Fig. 25(a). A base-voltage bias E_{bb} is used to eliminate the use of the critical-emitter bias E_{ee}. The circuit parameters are so chosen that, for the majority of transistors, the center portion of the negative-resistance region of each v_e'–i_e curve lies on the i_e axis. A crystal diode is used as the emitter load, causing the emitter load line to have two straight portions as shown in Fig. 25(b). This nonlinear load line insures the three-point-intersection requirement for each v_e'–i_e curve. Furthermore, at the low-conduction state the direction of emitter current is such that the back-resistance of the crystal diode constitutes a high impedance. Thus the input impedance of the circuit is high in its OFF state; the circuit may be readily triggered into its ON state by the application of a small positive pulse at the emitter.

Any change in circuit parameters and transistor characteristics affects the shape of the v_e'–i_e curve; a large departure from the designed value may render the circuit inoperative. However, the circuit parameters in the practical circuit are much less critical than are the values of E_{ee} and R_e in the basic bistable circuit, especially when E_{ee} and E_{cc} are from the same power supply. Laboratory experience revealed that the practical cir-
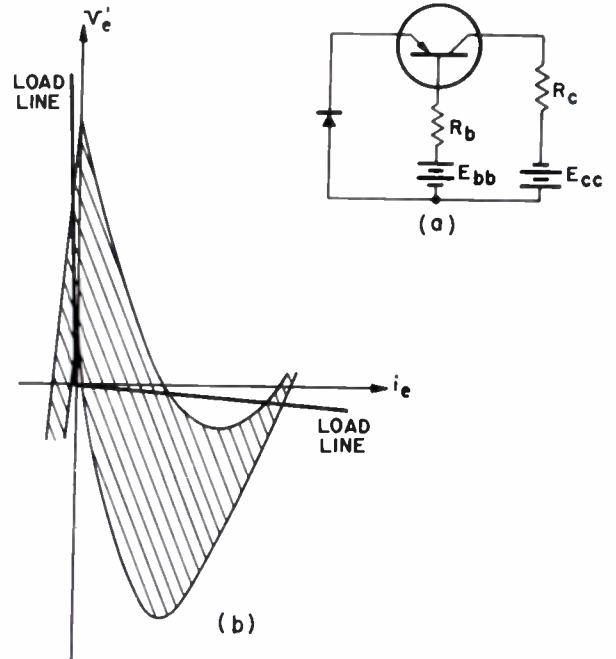
Fig. 25—Practical bistable circuit. (a) Circuit arrangement. (b) Emitter input characteristics.

cuit is highly stable; and reliable operation was observed even when power dissipation within the transistor was high enough to cause heating of the unit. The single-transistor circuit, in fact, was found superior even to earlier twin-transistor bistable circuits in which reliability had been emphasized.

A Decade Counter

To illustrate the practical application of point-contact-type transistors in pulse circuits, an all-transistor decade counter will be described here (Fig. 26). The counter takes the form of a conventional transfer-pulse ring counter. Each decade of the counter consists of
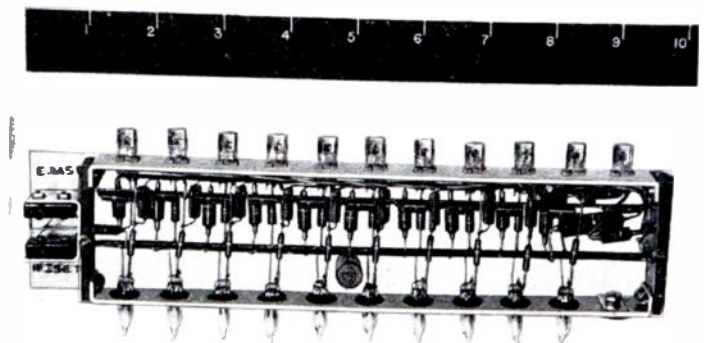
Fig. 26—Transistor decade counter.

ten bistable units interconnected to form a closed ring. One of the ten bistable units is in the ON state while the remaining nine units are in the OFF state. An input counting pulse, fed to all the bistable units in parallel, tends to turn all the units to the OFF state. When the originally ON unit jumps to its normal state, it sends out a transfer pulse which overrides the input pulse at the next succeeding unit and turns the latter to the ON state. Successive input pulses will advance the ON state along the ring in the same manner. The neon lamps shown in the photograph are connected so as to indicate which unit is in the ON state, and thus indicate the count.
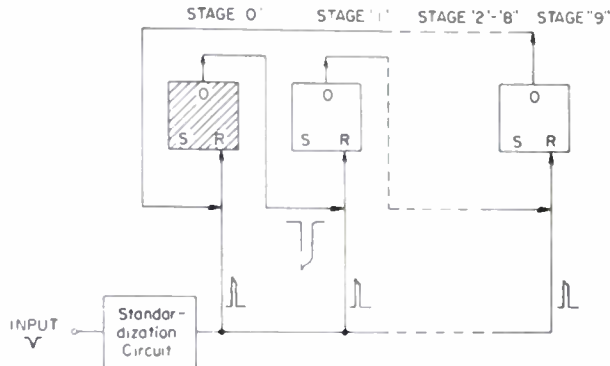
Fig. 27—Block diagram of transistor decade counter.

A block diagram of one decade of the counter is shown in Fig. 27. Before operation all units save the first are reset to their OFF state. In order to take care of counting pulses of varied width and amplitude a pulse-standardization stage is included. The single-transistor bistable circuit described in the last section is used as the bistable unit. However, since the emitter is not used as an input terminal at any time, instead of the employment of a crystal diode, the emitter is directly connected to ground (Fig. 28).
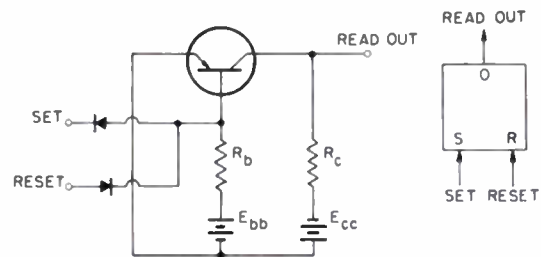
Fig. 28—A bistable unit.

The laboratory model of the counter was operated at rates of higher than 50,000 counts per second. One decade of the counter has a total power consumption of less than 2 watts. Its physical dimensions are $10 \times 3 \times \frac{3}{4}$ inches and weighs 8 ounces. A vacuum tube decade of comparable performance requires 10 twin-triodes, 2 pentodes, and a power consumption of 60 watts.

Acknowledgments

Transistors in Switching Circuits*

A. EUGENE ANDERSON†, MEMBER, IRE

Summary—The general transistor properties of small size and weight, low power and voltage, and potential long life suggest extensive application of transistors to pulse- or switching-type systems of computer or computer-like nature.

It is possible to devise simple regenerative circuits which perform the normally employed functions of waveform generation, level restoration, delay, storage (registry or memory), and counting. The discussion is limited to point-contact type transistors in which the alpha or current gain is in excess of unity and to a particular feedback configuration.

Such circuits, which are of the so-called trigger type, are postulated to involve negative resistance. On this basis an analysis, which approximates the negative-resistance characteristic by three intersecting broken lines, is developed. Conclusions which are useful to circuit and device design are reached. The analysis is deemed sufficiently accurate for first-order equilibrium calculations.

Transistors having properties specifically intended for pulse service in the circuits described have been developed. Their properties, limitations, and parameter characterizations are discussed at some length.

Introduction

IT IS PROPOSED to discuss some of the properties of transistors which are applicable to switching or pulse-type circuits, to develop elementary analysis methods, and to describe a few circuits.

The bounds or limits of the field of switching are difficult to define. The common thread usually involves definite states of being as "open or closed," "off or on," "0 or 1," and so on, rather than a continuum of conditions. Even when consideration is given to more than two states, the thought involves distinct recognition of each state. The field is termed to be nonlinear in distinction to linear manipulation of information. Any number of anomalies in definition may be raised.

Without attempting either to define or to limit the field, some of the functions which are often employed are waveform generation, as rectangular pulses, sawtooth waves, and so on; memory or storage which may be for short, intermediate, or long periods and involves the retention of information for subsequent use; operations involving addition, subtraction, multiplication, and

division; translation of information from one form or code to another; gating, involving the routing of signals according to a predetermined pattern or set of conditions; regeneration of signals in amplitude and waveform; delay, which may be thought of as a form of storage; and timing. Some of these functions are simple; others result from fairly complex structures of simpler functions.

Present trends in electronic switching systems are toward complicated automata as exemplified by digital computers.[1] The reliability, power consumption, and physical size of the electron devices employed largely determines the degree of realizability of such systems. It is believed that the transistor will find a significant application in this field.

The transistor can reduce power consumption by the elimination of heater or filament power. In addition, particularly in broadband applications as in high-speed pulse systems, the "B" power may be reduced by the order of one or two decades, if not more. Transistor circuits with 0.02 μs rise time have been made to operate with an input power of 20 mw, which compares with approximately 2.5 watts (1-watt heater, 1.5-watt plate) for an equivalent tube circuit. Transistors have operated with less than 1-μw input power.[2]

Such power reductions result from the low operating voltages, low internal resistances, and low capacitances of transistors. Low internal impedances greatly reduce the importance of stray wiring capacitances, thereby making mechanical design much simpler and often eliminating the need for isolating or buffer amplifiers.

The transistor can contribute definite reduction in size directly. Fig. 1 is a photograph of a "bead" transistor which has a volume of approximately 1/1000 of a cubic inch and a weight of 5/1000 ounce. Indirectly, the transistor can contribute to size reduction through the use of smaller, lower-voltage, lower-dissipation components. The reduction of power-supply requirements in terms of size, regulation, and capacity is also quite appreciable.

Transistors have been subjected to shocks in excess of 20,000 grams without change in characteristics. Vibration tests have shown no resonances in the transistor shown in Fig. 1 to several thousand cycles. Harmonic accelerations of 100 grams at 1,000 cycles have produced no detectable current modulation.

Life reliability and expectancy are difficult to determine because of the relative infancy of transistors, the definite finiteness of time, the many variables involved, and the rate of development progress. Average life is presently estimated to be in excess of 70,000 hours. Life is a function of the operating conditions, and may be

materially reduced accordingly.

Transistors also have limitations. Noise at present is high for point-contact types as compared to electron tubes; input impedances are low, which may be either advantageous or disadvantageous; power output may be limited; frequency response is relatively low; circuit instability may cause design difficulties; and the devices are sensitive to temperature changes. There is also an absence of a long practical experience with a consequent art background in both devices and circuits.
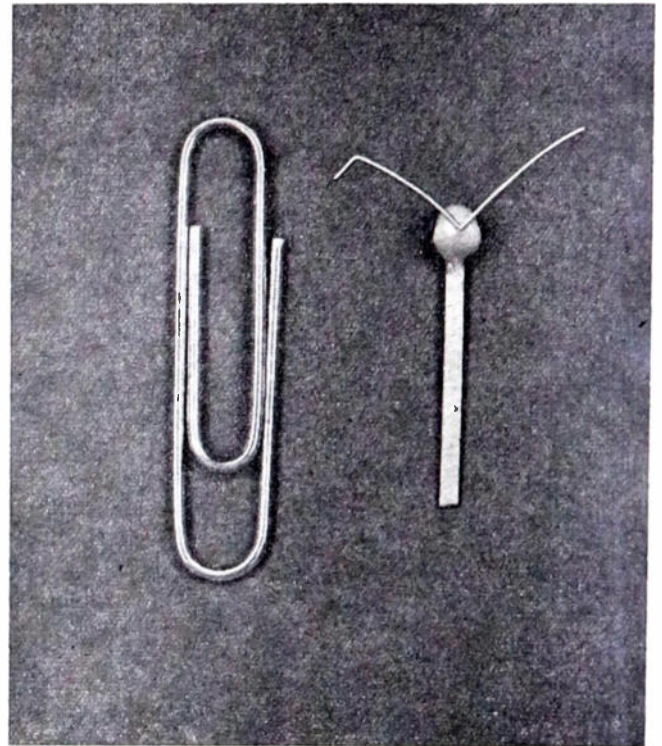
Fig. 1—A miniature switching type transistor (M1689).

A comprehensive review of transistor properties is given in the paper by Morton.[3]

While it is difficult to define the switching field, it is no less difficult to discuss circuit and device properties on a general basis. This is related to the nonlinear nature of the circuits and devices in distinction with the virtually classical linear small-signal field. The lack of a classical method of analysis is a serious handicap in the synthesis of contemporary circuits and devices. When new devices, as the transistor, are to be considered, the problem is multiplied due to the lack of a long background of experience.

It has been assumed that negative resistance is a common thread among "trigger circuits" and oscillators regardless of the device employed electron tube, gas tube, transistor, mechanical structures, and the like. This is not a new or novel idea and there is no intent to present

[1] L. N. Ridenour, "High speed digital computers," *Jour. Appl. Phys.*, vol. 21, pp. 263–270; April, 1950.
[2] R. L. Wallace, Jr. and W. J. Pietenpol, "Some circuit properties and applications of *n-p-n* transistors," *Bell Sys. Tech. Jour.*, vol. XXX, no. 3, pp. 530–563; July, 1951.

[3] J. A. Morton, "Present status of transistor development," *Bell Sys. Tech. Jour.*, vol. XXXI, no. 3, pp. 411–442; May, 1952.

it as such.[4] Rather, it is used as a pattern upon which a certain degree of transistor analysis may be based, leading to simple understanding. The analysis assumes that the negative-resistance characteristic can be broken into three regions; each region is then considered on a linear basis.

Section I will deal with simple circuit properties, Section II, analysis, and Section III, device properties.

I. SIMPLE CIRCUIT PROPERTIES

Common property ascribable to switching functions is that of definiteness of state. The condition of the function is either "off" or "on." Switches are either open or closed; relays are operated or not; tubes are in cutoff or overload; doors are open or closed; and so on. This is common regardless of the phenomena being exploited.

There is an intermediate region between these two conditions, usually characterized by a time which is related to how fast the function may go from one state to another. Functionally, the times of closing and opening are taken to be zero; practically, they are of determining importance. Relays replace hand-operated switches and electronic devices replace relays as speed becomes important. Obviously, no function or system can be faster than its state devices.

All such state devices will have separate attendant properties, such as the degree of reverse coupling between the controlling signal and the controlled signal. Separated into families, however, there are those which are passive and those which are active. The latter are threshold devices in which a small amount of signal or control energy causes the translation of a relatively larger amount of stored energy into dynamic energy which consummates the change in state. As long as the control signal is below the initial threshold, there is no response and any change is directly related to the passive transmission of the control signal alone. When the signal exceeds the threshold, the second state is assumed. Watch escapements, thyratrons, and the whole family of oscillators fall into this category. When the simplest cases of such functions are analyzed, they are found to involve, in one way or another, two stable states separated by a region in which there is positive feedback and gain in excess of unity, with a resultant equivalent negative resistance. The proposition that a negative-resistance characteristic is common to trigger or threshold switching circuits is tacitly assumed. The next step is to examine the transistor for such behavior and to classify the properties.

[4] See, for example, W. Reichardt, "Negative resistances, their characteristics and effects. Sinusoids, relaxation oscillations, and relaxation discontinuities," *Elek. Nach. Tech.*, vol. 20, no. 9, pp. 76–87; March, 1943. Also, "Uniform relationship between sinusoids, relaxation vibrations, and discontinuities," *Elek. Nach. Tech.*, vol. 20, no. 9, pp. 213–225; September, 1943. For transistors, E. Eberhard, R. O. Endres, and R. P. Moore, "Counter circuits using transistors," *RCA Rev.*, pp. 459–476; December, 1949. Also, H. J. Reich and Ungvary, "A transistor trigger circuit," *Rev. Sci. Inst.*, vol. 20, pp. 8 and 586; August, 1949. Also, P. M. Schultheiss and H. F. Reich, "Some transistor trigger circuits," PROC. I.R.E., vol. 39, pp. 627–632; June, 1951.

Negative Resistance in the Transistor

That the transistor[5] can exhibit negative resistance has been demonstrated analytically[6] and experimentally. The resistances seen looking into the emitter and collector of the transistor with grounded base are shown in Fig. 2.

$$R_{IN} = R_{11} - \frac{R_{12}R_{21}}{R_{22}'} = r_\epsilon + r_b - \frac{r_b(r_b + r_m)}{r_c + r_b + R_c}$$

$$R_{IN} = R_{22} - \frac{R_{12}R_{21}}{R_{11}'} = r_c + r_b - \frac{r_b(r_b + r_m)}{r_\epsilon + r_b + R_\epsilon}$$
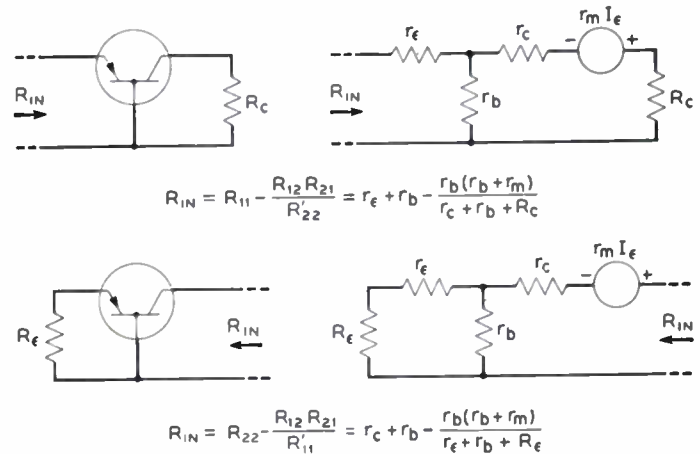
Fig. 2—Emitter and collector driving-point resistances.

In the equations and discussion to follow, the symbol conventions are as follows: External circuit elements are capitalized as R_ϵ, R_b, and R_c. The symbols R_{11}, R_{12}, R_{22}, and R_{21} define the open-circuit transistor resistances; the symbols r_ϵ, r_c, r_m, and r_b define the equivalent circuit element values. Network resistances which contain both device and external elements are primed. For example, $R_{22}' = R_{22} + R_c + R_b$, where $R_{22} = r_c + r_b$. See also footnote references[3,6].

Taking the collector or output driving-point resistance, Fig. 2, for example,

$$R_{in} = (r_c + r_b) - \frac{r_b(r_b + r_m)}{r_\epsilon + r_b + R_\epsilon}. \tag{1}$$

R_{in} can be negative or positive, depending upon the relative magnitudes of the two terms. Actually, of course, r_m has a phase factor and so is frequency dependent. Frequencies wherein r_m is essentially resistive will be assumed. For negative resistance, r_m must be large, R_ϵ small, and r_b not too small or else augmented by external resistance. Negative resistance is thus predicted on a small-signal linear basis. The large-signal behavior may be studied experimentally by adding sufficient resistance as R_c to the first or positive term to insure stability. This is shown in Fig. 3 with the resultant characteristic. External base resistance R_b has been added and R_ϵ is zero.

[5] Discussion is limited primarily to point-contact transistors with α's or current gains greater than unity.
[6] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. XXVIII, no. 3, pp. 367–400; July, 1949.

Fig. 3 illustrates the pattern of a three-valued characteristic. I and III are portions with positive slope, indicating stable operating regions, separated by II, a region of negative slope, indicating the possibility of instability. In this particular case, Region I has high resistance and Region III very low resistance.
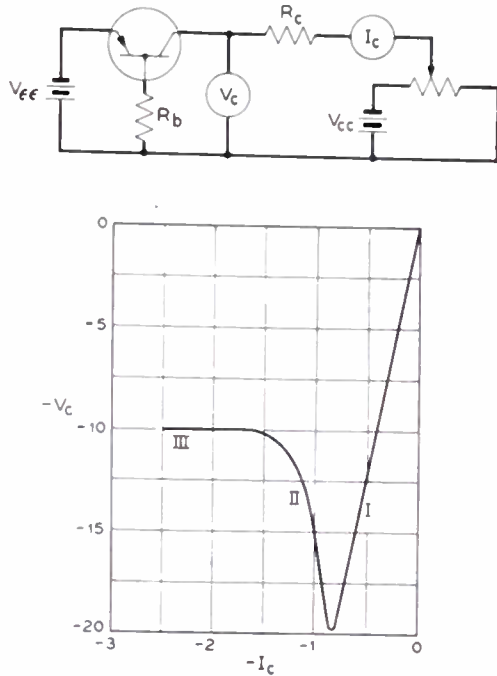
Fig. 3—Collector large-signal negative-resistance characteristic.

An evaluation of the emitter or input characteristic leads to similar results, using the circuit of Fig. 4. R_b has been added here also and R_c taken as zero. The general pattern is again present. Region I has high, positive resistance; Region II, negative resistance; and Region III very low, positive resistance.

Fig. 4—Idealized emitter large-signal negative-resistance characteristic.

Biases and Load Lines—Bistable Operation

The negative resistances of Figs. 3 and 4 are both of the so-called open-circuit stable type. If loads are applied to the circuit terminals of Fig. 2 which are larger in magnitude than the negative resistances, the circuits will be stable, that is, there will be single operating points. This is shown in Fig. 4 by the dashed load lines marked, R_ϵ', R_ϵ'', R_ϵ'''. A load resistance smaller in magnitude than the negative resistance may intersect the characteristic in three positions, as shown by the load line R_ϵ.

The load line R_ϵ can be made to have single or multiple intersections by biasing properly as shown in Fig. 5, where the three possibilities are shown as R_ϵ, R_ϵ', R_ϵ''. Single or multiple intersections result in accordance with the choice of emitter bias, $V_{\epsilon\epsilon}$, as shown. It can be shown that the intersection of load line R_ϵ with the characteristic at b in Fig. 5 is unstable, whereas those at a and c are stable. Experiment in the multiple intersection case shows also that as $V_{\epsilon\epsilon}$ is slowly increased (decreased in absolute magnitude) the load line moves upward and that the assumed operating point, a, moves up along the Region I portion of the characteristic. At the turning point shown on the current axis, the operating point suddenly flips to the high current region, returning along the curve to c as $V_{\epsilon\epsilon}$ is returned to the original value.
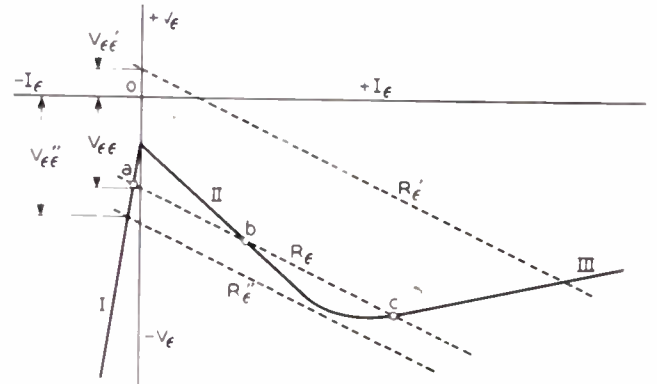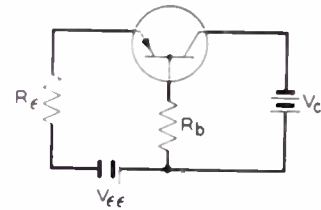
Fig. 5—Emitter negative-resistance characteristic showing possible multiple operating points.

A decrease in $V_{\epsilon\epsilon}$ toward $V_{\epsilon\epsilon}''$ moves the operating point at c downward along the characteristic until it "escapes" past the lower turning point and flips to the Region I portion, returning to a as $V_{\epsilon\epsilon}''$ is returned to the original value. This then is an elementary switching circuit, a bistable trigger circuit, or "flip-flop." A positive emitter pulse will cause the circuit to flip to high current, a negative pulse to low current. The triggers

may be applied to emitter, base or in combination with proper attention to polarity. Trigger sensitivities are shown in Fig. 6. Such a circuit is often used for register or storage purposes. It can store one bit of information for a potentially infinite period, be sampled for presence of such information, and be cleared or restored to the original condition for reuse when the stored information is no longer useful. With the addition of suitable steering circuits it can be made to count by a scale of two.
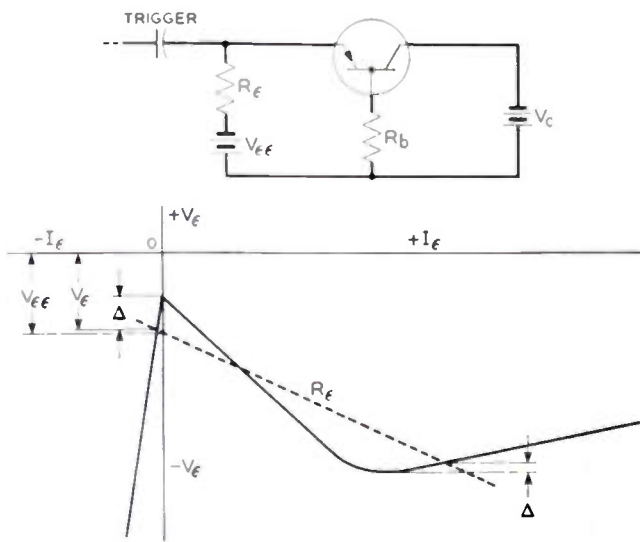
Fig. 6—Bistable circuit showing trigger sensitivity, Δ.

Monostable and Astable Circuits

The addition of a capacitor to the circuit as in Fig. 7(a) leads to either monostable or astable operation. In Fig. 7(b) the normal operating point is stable at a, as discussed previously, by virtue of the bias $V_{\epsilon\epsilon}$. As $V_{\epsilon\epsilon}$ is increased, as by a trigger, the load line is moved up and over the turning point. Without capacitor C in the circuit, the operating point would move to b, with the resultant rapid change in voltage and current. However, a capacitor has in effect voltage inertia; this is equivalent to saying that a capacitor is a short-circuit to a voltage change. Both the capacitance and the rate of change of voltage are assumed high. Thus at the turning point the capacitor effectively short-circuits the emitter and the operating point snaps along dotted line (1) to intersect the characteristics. This point is quasi-stable, and the capacitor is discharged along line (2) to the second turning point where the emitter is again effectively short-circuited and the operating point snaps along (3) to intersect the Region I portion of the characteristic. This point is also quasi-stable and the operating point moves slowly up to the initial or dc stable operating point. A single trigger thus causes a complete cycle of operation. The emitter current shifts rapidly to a high value of current, falls relatively slowly to an intermediate value, then shifts rapidly to a small negative value, and finally returns slowly to the original value. The emitter current and voltage are sketched in Fig. 8. It is a so-called "single-shot" circuit. Alternately, the

rest or dc stable point can be chosen to be in Region III, at high current, by choice of positive instead of negative bias $V_{\epsilon\epsilon}$. Practical considerations as ease in triggering and average power consumption usually indicate a preference for the Region I dc stable point.

When the load line and bias are chosen to result in intersection in the negative-resistance portion, astable operation or continuous oscillation results. This mode is illustrated in Fig. 7(c). Proper bias and $R_{\epsilon} > |-R_{\text{in}}|$, Region II, are required. The operating point formed by the intersection of the load line on the negative-resistance portion of the characteristic would normally be
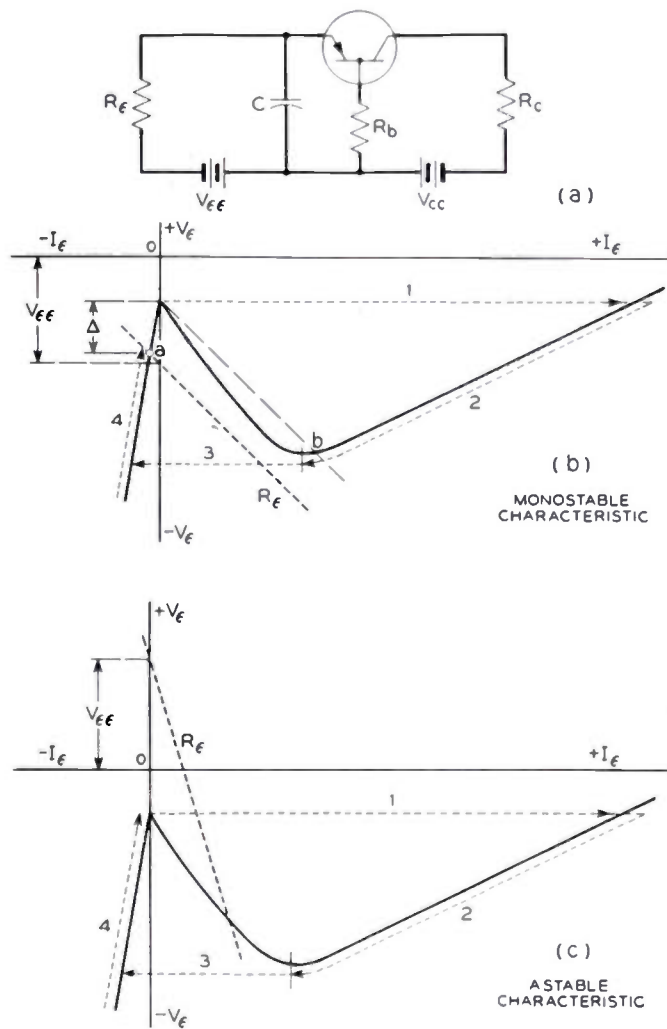
Fig. 7—Monostable and astable characteristics.

stable. However, the capacitor provides an ac short circuit in parallel with R_{ϵ}, causing the path (1), (2), (3), (4) to be followed continuously. Another form of physical explanation of this relaxation oscillation, usually applied to gas tubes, is that the capacitor C is charged slowly through R_{ϵ} to a critical or breakdown value whereupon the tube or device rapidly discharges the capacitor. When the capacitor charge is dissipated, the device discharge can no longer be maintained because of the IR drop in R_{ϵ} and the tube or device open circuits, and the capacitor is recharged.

The above suggests a strong similarity to gas-tube behavior, and this is indeed so. In fact, the modes described above are common to all open-circuit stable negative-resistance devices; only the parameters and device phenomena are different.

The primitive circuits of Fig. 7 have properties basic to several switching functions. These may be deduced from the waveforms of Fig. 8 which are essentially identical to both the monostable and astable cases. The emitter current has a rectangular waveform which suggests the generation of rectangular pulses, and, for the astable case, regenerative amplification for both amplitude and wave shape, pulse rate or frequency division and delay. As shown, the current waveform is not particularly good, having neither a flat top nor a flat base line. Practically, the waveform may be derived from the collector by means of a small load resistor to obtain a flat base line. When the emitter current is negative, there is sensibly no transfer action, hence, the collector current will be constant during the recharge portion of the cycle instead of exponential as shown. The slope of the top is inherent, and may be removed by clipping. Pulse rise time of 0.1 μs, the time required for transition from low current to high current, is quite easily obtainable; 0.02 μs with average input powers of 20 mw have been obtained. Fall time is usually longer than the rise time by factors of 3 or 4. It is to be noted in Fig. 8 that there has been shown a delay between the trigger application and the current transition. Such delay is not peculiar to transistors, but is common to all trigger-type devices and circuits. The delay is shown here exaggerated in order to establish its existence, and is associated with the static charging of the circuit and the dynamic delay of the device concerned. The trigger-transition delay with transistors is usually less than 0.1 μs.
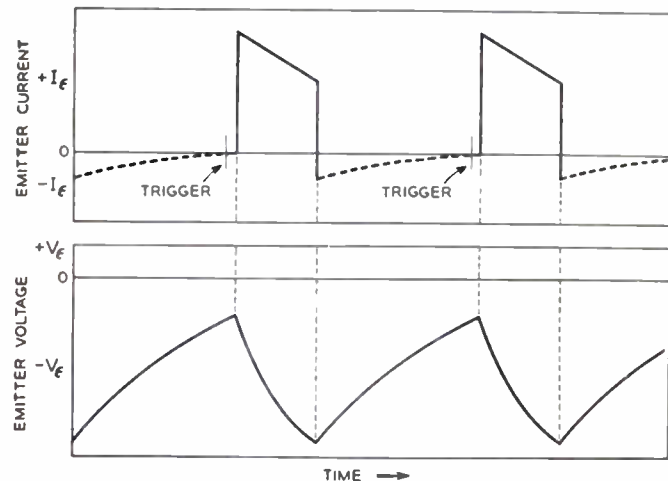
Fig. 8—Idealized monostable relaxation oscillator waveforms.

The voltage waveform of Fig. 8 has a sawtooth form, and may thus be employed to generate linear time bases or sweeps. The normal methods for linearization, such as a high-charging voltage V_{ϵ} and a high-charging resistance R_{ϵ}, or other constant current means, are applicable here as in other device circuitry. Free-running and driven sweeps may be obtained with the astable and monostable circuits, respectively.

Since the collector characteristic shown in Fig. 3 is also open-circuit stable, the same sort of circuits can be constructed using the output characteristic. Bistable, monostable, and astable circuits are shown in Fig. 9.

(a) COLLECTOR BISTABLE CHARACTERISTIC

(b) COLLECTOR MONOSTABLE CHARACTERISTIC

(c) COLLECTOR ASTABLE CHARACTERISTIC
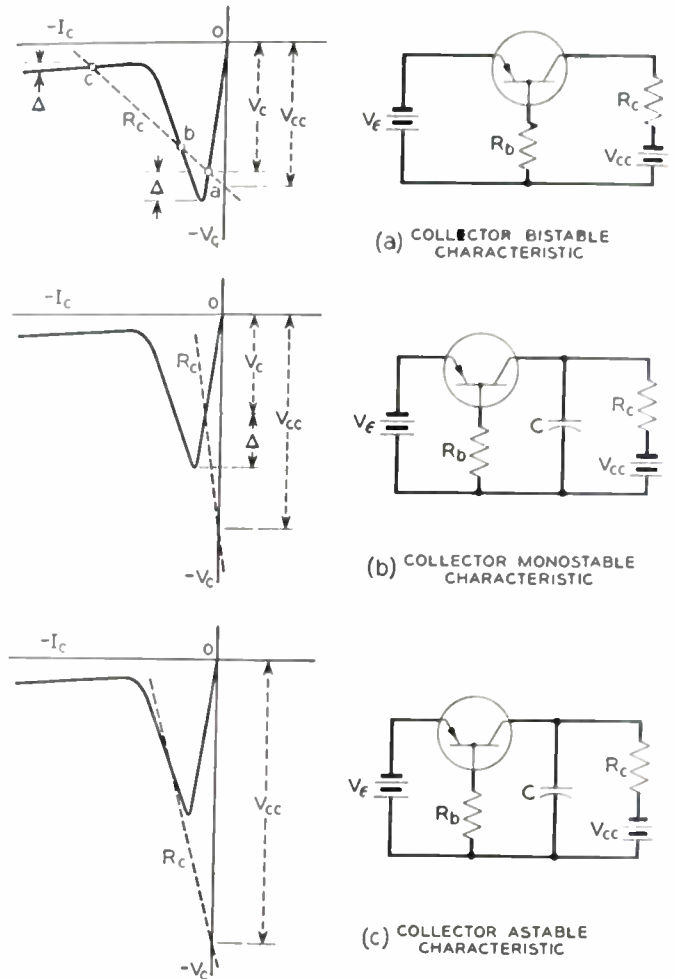
Fig. 9—Collector connection switching circuits.

$$R_{IN} = r_b + r_\epsilon + \frac{r_\epsilon (r_m - r_\epsilon)}{R_c + r_\epsilon + r_c - r_m}$$

$$R_{IN} = r_b + r_c - r_m + \frac{r_c (r_m - r_c)}{R_\epsilon + r_\epsilon + r_c - r_m}$$
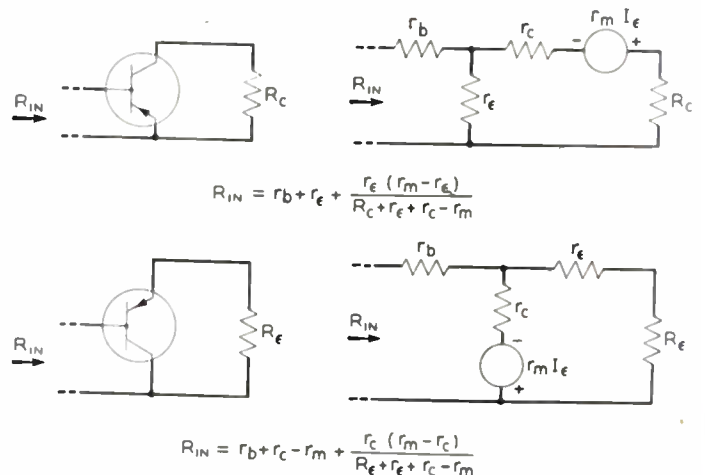
Fig. 10—Base driving-point resistances.

The resistances seen looking into the base are given in Fig. 10. These circuits are short-circuit stable. That is, high values of R_b result in instability. Bistable, monostable, and astable circuits can be constructed also, but use is made of an inductor instead of a capacitor. The

reactance of the inductor affords a quasi-open-circuit in the same manner as the capacitor afforded a quasi-short-circuit in the previous cases. Circuit examples are shown in Fig. 11.
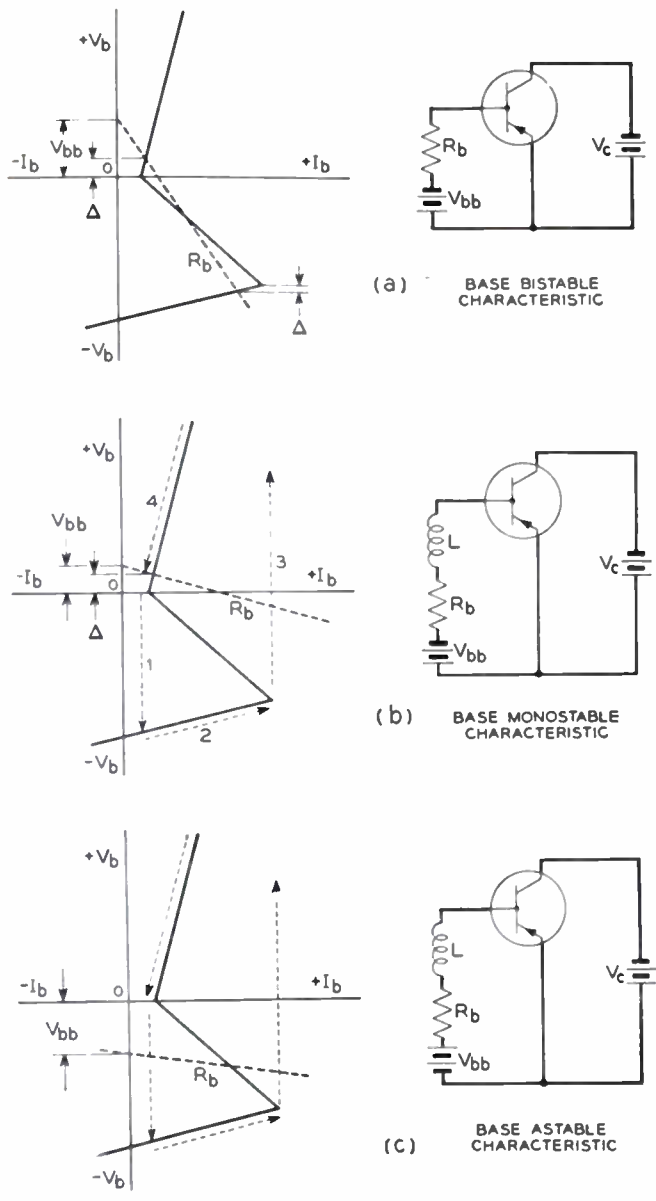
Fig. 11—Base connection switching circuits.

Summary

These simple circuits by no means exhaust the switching-circuit possibilities of the transistor; rather, they are the simplest. The simple circuit is often satisfactory, and may sometimes be employed with little more understanding than that given. More often, however, problems relating to the sensitivity, constancy of sensitivity, operating currents and voltages, interchangeability and the like require a much more quantitative understanding in order to create circuit designs having specific properties.[7] An equal need also exists in transistor de-

[7] See, for example, J. R. Harris, "A transistor shift register and serial adder," Proc. I.R.E.; vol. 40, pp. 1597–1603; this issue; R. L. Trent, "A transistor reversible binary counter," Proc. I.R.E., vol. 40, pp. 1562–1573; this issue; H. G. Follingstad, J. N. Shive, R. E. Yaeger, "An optical position encoder and data transmitter," Proc. I.R.E., vol. 40, pp. 1573–1584; this issue.

sign for analytic circuit relationships. Such information is useful, first, in the creation of optimized designs and, second, in the maintainance of proper parameter controls in manufacture. Finally, the more detailed the understanding the more likely will be the creation of new circuits and new devices.

A complete analytical treatment will not be attempted here; consideration will be limited to the equilibrium case and, in particular, to the simple circuits described.

II. Analysis

In order to deal analytically with circuits and devices it is necessary to have analytic expressions for the device characteristics. For small-signal analysis this is relatively easy. In large-signal applications, as in switching, the situation is not so simple. The problems arise because of the high degree of nonlinearity, wherein the simplifying assumptions employed in small-signal analysis are by no means valid. Furthermore, it is desirable to retain dc terms in many cases.

The method to be employed here is the so-called broken-line method which involves approximating the negative-resistance characteristic by three intersecting straight lines. The assumption is made that there are three distinct regions of operation, in each of which the device is separately linear but involving different parameter values for each region.
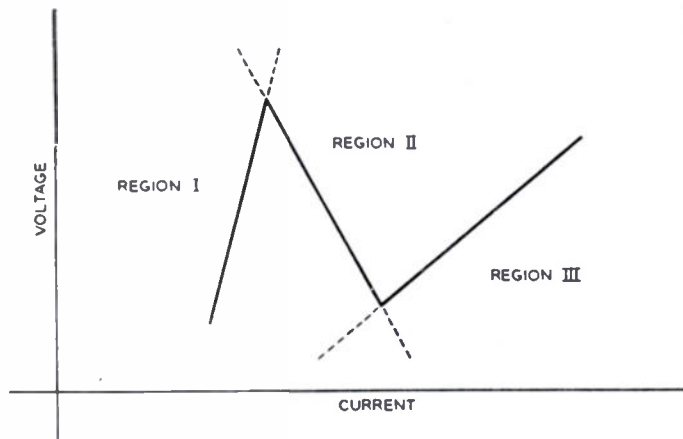
Fig. 12—Broken-line idealization of negative-resistance characteristic—division into regions.

The approximation is shown in Fig. 12. The assumption that the negative-resistance characteristic can be simulated by three straight lines is reasonably valid for gross considerations; for fine detail near the turning points the approximation is by no means accurate although affording zero-order information.

Preparatory to the analysis of the negative-resistance characteristics, it is necessary to obtain analytic expressions for the transistor currents and voltages. This in turn involves the following steps:

1. identification of the three regions in terms of the device characteristics,
2. idealization of the device characteristics to obtain simple, linear relations, and

3. evaluation of the device parameters in each of the three regions.

Fig. 13 is a family of open-circuit characteristics for a typical switching-type transistor. Specifically, in small-signal terms,

TABLE I

SMALL-SIGNAL EQUIVALENT CIRCUIT DEFINITIONS AND RELATIONS

Parameter	Equivalent Tee
$R_{11} = \dfrac{\partial V_e}{\partial I_e}\bigg]_{I_c}$	$R_{11} = r_e + r_b$
$R_{12} = \dfrac{\partial V_e}{\partial I_c}\bigg]_{I_e}$	$R_{12} = r_b$
$R_{21} = \dfrac{\partial V_c}{\partial I_e}\bigg]_{I_c}$	$R_{21} = r_m + r_b$
$R_{22} = \dfrac{\partial V_c}{\partial I_c}\bigg]_{I_e}$	$R_{22} = r_c + r_b.$

Also.

$$\alpha = -\frac{\partial I_c}{\partial I_e}\bigg]_{V_c} = \frac{R_{21}}{R_{22}} = \frac{r_m + r_b}{r_c + r_b}.$$

The above set, normally employed for small-signal analysis, will be assumed to be constant within a given region but changing in value from region to region.

Identification of the Three Regions

It may be recalled with the aid of Fig. 12 that the negative-resistance characteristic consists of a negative-resistance region bounded on each side by a region of positive resistance. Thus the device is first passive in nature with little or no gain, then very abruptly exhibits considerable gain with the resultant negative resistance, and finally becomes very abruptly passive again with little or no gain.

It would seem quite clear that abrupt changes in the transmission properties of a device should be associated with equally abrupt changes in the forward transfer characteristic. In the case of the transistor, the behavior of the forward transfer properties is given by the forward transfer impedance, R_{21}.

Examining the R_{21} family in Fig. 13, it is seen that in the normal, positive emitter current region the slope, R_{21}, is high, indicating the possibility of high forward gain. When I_e is negative, however, the slope is zero or nearly so, changing very abruptly at $I_e = 0$. Furthermore, it is to be noted that as I_e is made negative, the collector voltage is unaffected, remaining constant for further change in I_e. Thus it may be said that the collector voltage is saturated.[8]

If, on the other hand, the emitter current is increased, at constant collector current, it is found that at a critical emitter current the slope again becomes zero or nearly so. There are also two further observations: First, the

collector voltage is reduced to a very small value and, second, that the critical emitter current is related to the collector current. From the small-signal relation,

$$V_c = R_{21}I_e + R_{22}I_c \qquad (2)$$

or

$$V_c = R_{22}\alpha I_e + R_{22}I_c, \qquad (3)$$

the critical emitter current for collector-voltage cutoff may be obtained by setting $V_c = 0$, as,

$$I_e = -\frac{I_c}{\alpha}. \qquad (4)$$

This relationship is dual to the grid voltage-plate voltage relation in tubes for plate-current cutoff, as $V_g = -(V_p/\mu)$. The criteria for defining the three regions are thus established as

Region I (collector-voltage saturation): $I_e < 0$,   (5)

Region II (active): $0 < I_e < -\dfrac{I_c}{\alpha}$,   (6)

Region III (collector-voltage cutoff): $I_e < -\dfrac{I_c}{\alpha}$.   (7)

The identification of device parameters will be made for the several regions by a single prime for Region I as r_e', none for Region II as r_e, and three primes for Region III as r_e'''.

Linearization of the Characteristics and Approximations

The next step is to linearize the characteristics and to make suitable approximations in order to obtain simple linear equations of the terminal currents and voltages. The relations which require linearization are the device parameters R_{11}, R_{12}, R_{21}, and R_{22}, which are in general functions of the currents as $R_{ij} = f(I_1, I_2)$.

Linearization of R_{11} and R_{12}

In terms of the equivalent tee circuit, which has been and will be employed, R_{11} is given as $R_{11} = r_e + r_b$. Also, $R_{12} = r_b$. It is convenient to separate r_e and r_b and discuss each separately since r_b is fairly constant and r_e will have widely different regional values.

In the R_{12} family of Fig. 13, it may be seen that R_{12} or r_b is fairly constant in all three regions, and will be so taken here. Furthermore, in the simple circuits under consideration, external base resistance R_b has been inserted so that minor variations in r_b in the total of $r_b + R_b$ are inconsequential since usually $R_b \gg r_b$. The approximation that r_b is constant is subject to review where finer detail is necessary, particularly at low emitter currents where the rate of change of r_b is at a maximum.

The emitter resistance r_e is approximately the resistance of a diode in the forward direction. As such, r_e is high when the emitter current is negative and low when the emitter current is positive. Experimentally, it is found convenient to give three values to r_e and hence to

---

[8] It is tacitly assumed that in the relation $y = f(x)$ there are extremes at which y becomes essentially constant and independent of further change in the independent variable x. The point farthest removed from the origin at which the dependent variable becomes constant is termed "saturation". The point closest to the origin at which the dependent variable becomes constant is termed "cutoff".
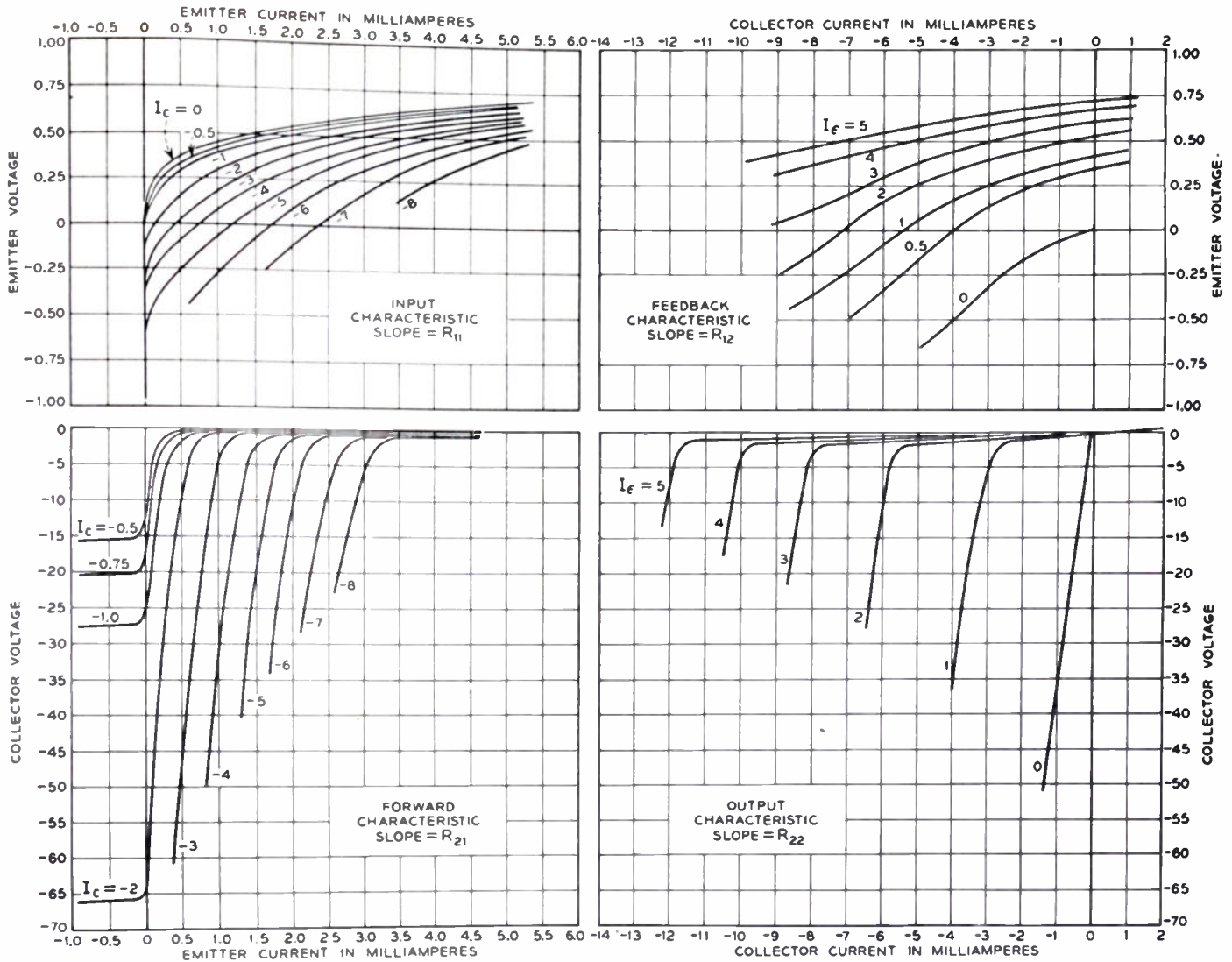
Fig. 13—Static characteristics of the M1689 developmental switching-type transistor.

R_{11}, one for each region as shown in Fig. 14. This recognizes the nonlinearity with I_e in the forward direction and assumes that a single value in the reverse direction is sufficient. As the circuitry becomes more sophisticated, a more precise approximation will undoubtedly be required, particularly near $I_e=0$.

It may be noted that in the functional relation $r_{11}=f(I_e, I_c)$ that R_{11} is taken to be a function of I_e only. The contribution of I_c is to shift the characteristic in voltage by $r_i\Delta I_c$ increments. Thus the relationship of $V_e=f(I_e, I_c)$ can be written very simply as

$$V_e = R_{11}I_e + R_{12}I_c. \qquad (8)$$

Since the problem has been linearized to first-order terms only, the currents and voltages are total instantaneous or dc values as indicated by the capital letters.

Idealization of R_{21}

As indicated previously, R_{21} will be small in Regions I and III and large in Region II. Since $R_{21}=r_m+r_b$, R_{21} can be no less than r_b; the defining approximations will be applied to r_m. In Region I, when the emitter current is negative, r_m is taken to be zero and reflects the device approximation that the emitter current, under this con-
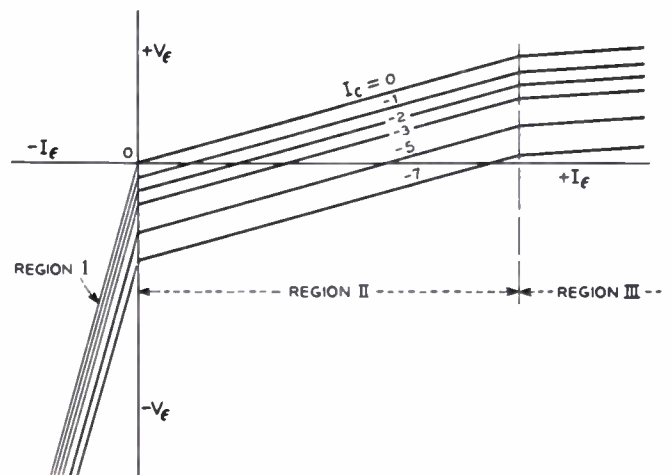
Fig. 14—Idealization and regional division of input characteristic (R_{11}).

dition, is entirely electron current. This is not always a true approximation, particularly near $I_e=0$, and limiting tests are employed in transistor testing.

In Region II, r_m has, of course, high values and, in general, $r_m \gg r_b$. Pending further investigation, r_m will be assumed finite but very small in Region III.

Idealization of R_{22}

In the output family of Fig. 13 it may be noted that R_{22} has two values, a high value for $I_c > -\alpha I_e$ and a low value for $I_c < -\alpha I_e$. The high value corresponds to Regions I and II and the low value to Region III. To a first order the two values are separately constant, which was not true of earlier transistors in which R_{22} underwent extensive degradation in magnitude as I_c and I_e increased.

The lower limit to which R_{22} can fall in Region III is r_b since $R_{22} = r_c + r_b$, implying that r_c is zero in Region III. This is approximately, but not accurately, true. As αI_e approaches $-I_c$ in magnitude, the voltage across the collector barrier becomes nearly zero so that r_c has a low, but finite, value. Under this condition, the hole current is very high and heavy conductivity modulation of the collector barrier resistance occurs. Thus the collector resistance in Region III is indeed quite low and may be neglected for many circuit computations.

In the functional relation $R_{22} = f(I_e, I_c)$ it has been assumed that R_{22} is a function of I_c alone. Also, the approximation involves first-order terms only, and hence the functional relation $V_c = f(I_e, I_c)$ may be written as

$$V_c = R_{21}I_e + R_{22}I_c. \tag{9}$$

Here again, as in the input case, the currents and voltages are total instantaneous or dc values as indicated by the capital letters.

It is believed desirable, however, to give one more consideration to the output relations. When $I_e = 0$, the collector characteristic is approximately that of a diode in the reverse direction. A diode has low reverse resistance until the voltage across the barrier exceeds a few tenths of a volt, and then has quite high resistance, approaching infinite slope in the case of junction diodes.[9,10] This effect is shown exaggerated in the idealized output family of Fig. 15. The current and voltage at the break in the $I_e = 0$ curve have been termed I_{c0} and V_{c0}, respectively. I_{c0} and V_{c0} are quite evident in junction devices; in point-contact devices they are not nearly so evident due to the lower value of R_{22} and the higher voltages and currents normally employed. Where currents and voltages are of the order of several milliamperes and a few volts, I_{c0} and V_{c0} may normally be neglected. I_{c0} and V_{c0} do have considerable interest to the device designer, however. The net circuit interpretation of I_{c0} and V_{c0} is to effectively transfer the current-voltage axis from 0, 0 to I_{c0}, V_{c0}. Therefore,

$$V_c - V_{c0} = R_{21}I_e + (I_c - I_{c0})R_{22} \tag{10}$$

or

$$V_c - V_{c0} = (r_m + r_b)I_e + (I_c - I_{c0})(r_c + r_b). \tag{11}$$

Making the approximation that $V_{c0} = I_{c0}R_{22}'''$ and rearranging, (10) becomes

[9] See footnote reference 2.
[10] W. Shockley, "Holes and Electrons," D. Van Nostrand Co., Inc., New York, N. Y., p. 91; 1950.

$$V_c - I_{c0}R_{22}''' + I_{c0}R_{22} = R_{21}I_e + I_cR_{22} \tag{12}$$

or

$$V_c + I_{c0}(R_{22} - R_{22}''') = R_{21}I_e + I_cR_{22}, \tag{13}$$

which is of the usual form except that a small dc generator of magnitude $I_{c0}(R_{22} - R_{22}''')$ has been added in series with the collector. Since $R_{22} = r_c + r_b$ and $R_{22}''' = r_c''' + r_b$,

$$I_{c0}(R_{22} + R_{22}''') = I_{c0}(r_c - r_c'''). \tag{14}$$

The output family equation with equivalent circuit parameters is then

$$V_c - I_{c0}(r_c - r_c''') = (r_m + r_b)I_e + (r_c + r_b)I_c. \tag{15}$$
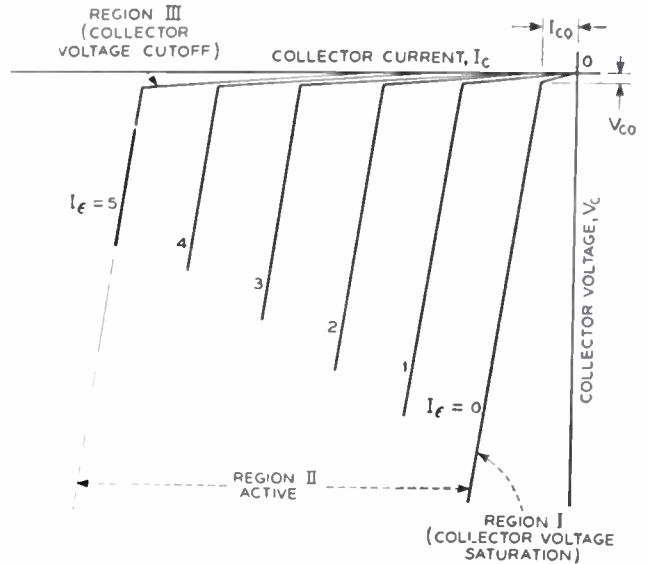
Fig. 15—Idealization and regional division of output characteristic (R_{22}).

Summary of Idealization of Characteristics

The results of the idealization of the device characteristics are summarized in Fig. 16. Here are given analytic expressions for the input and output voltages in terms of the input and output currents; the regions are defined symbolically and by typical values; and an equivalent circuit is given. It may be noted that the equivalent circuit is identical to the small-signal equivalent tee, excepting the small dc generator $I_{c0}(r_c - r_c''')$ which usually may be neglected when dealing with contemporary point-contact transistors.

To obtain any of the negative-resistance characteristics it is only necessary, first, to solve the two equations simultaneously for the appropriate voltage in terms of the appropriate current, and then, second, to insert into the resultant equation the proper parameter values, region by region, to obtain three equations. These equations, when plotted, result in an idealized characteristic similar in form to that of Fig. 12. A detailed example plus synopsis of the properties of the several connections will be given in the following subsection.

The Alpha or Current Gain Factor[11]

The derivation just given has been in terms of the equivalent circuit parameters, r_e, r_b, r_c, and r_m. Another circuit factor, alpha or the short-circuit current gain, is also quite useful. Alpha has been defined in Table 1 as the negative ratio of the incremental change in output current to the incremental change in input current *under the condition of short-circuit output terminals.*

Thus alpha is restricted in interpretation to a specific device termination, and care should be taken in the employment of alpha when other terminations are involved. For example, the circuit current gain under general conditions is given by R_{21}'/R_{22}'. The ratio R_{21}'/R_{22}' has been sometimes termed α_c. Thus,

$$\alpha_c = \frac{R_{21}'}{R_{22}'} = \frac{r_c + R_b + r_m}{r_b + R_b + r_c + R_c}. \tag{16}$$

Depending upon the magnitudes of R_b and R_c, the two current-gain ratios may be markedly different. In Region II, where r_m and r_c are very large, the effects of R_b and R_c in (16) often may be neglected. The circuit current gain, α_c, may then be taken as the device alpha. In Region I, r_m has been taken as zero; hence, the current gain will be somewhat less than unity, given by $(r_b+R_b)/(r_b+R_b+r_c+R_c)$, and is definitely not zero. Equally, in Region III, the circuit current gain is not zero, but rather approaches the ratio $(r_b+R_b/(r_b+R_b+R_c)$. If $R_b \gg R_c$, the ratio is nearly unity.

Analysis of Negative-Resistance Characteristic

The objectives of the circuit analysis, as stated previously, are

1. to determine operating conditions, such as proper values of loads, biases, trigger sensitivities, and operating currents and voltages,
2. to determine relationships of device parameters to the circuit behavior in order that these parameters may be optimized, properly characterized and controlled for required circuit performance.

For example, the trigger sensitivity may be given by the voltage difference between the load-line intersection with the Region I portion of the characteristic and the upper peak or turning point of the characteristic as shown in Figs. 6, 7, and 9. The sensitivity Δ is thus determined by the nearness of the bias point to the peak of the characteristic. Since the bias is normally fixed, variations in the sensitivity will arise from variations in the peak point. Thus it is necessary to know the relationships which determine the currents and voltages of the peak and valley points in order, first, to achieve a design and, second, to establish controls on the proper device parameters.

In this example the emitter negative-resistance characteristic will be solved and analyzed. The solutions for the other characteristics follow in the same manner and will be summarized.

[11] This section is inserted parenthetically as clarifying material due to the use of the α-factor in the subsequent discussion.

Evaluation of Emitter Characteristic as an Example

To obtain the emitter characteristic, it is necessary to solve the two equations of Fig. 16 for V_e in terms of I_e. The equations as given are for the short-circuit case. Since the general case will involve external parameters as R_e and R_c, the equations will be modified to include these parameters.
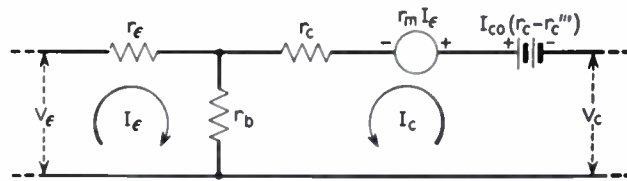
$$V_e = (r_e + r_b) I_e + r_b I_c$$
$$V_c + (r_c - r_c''') I_{c0} = (r_m + r_b) I_e + (r_c + r_b) I_c$$

REGION	r_e		r_b		r_c		r_m	
	SYMBOL	TYPICAL	SYMBOL	TYPICAL	SYMBOL	TYPICAL	SYMBOL	TYPICAL
I	r_e'	100 K	r_b	160	r_c	20 K	r_m'	0
II	r_e	100	r_b	160	r_c	20 K	r_m	50 K
III	r_e'''	25	r_b	50	r_c'''	70	r_m'''	30

$I_{c0} \approx -50\mu A$

Fig. 16—Broken-line transistor equations, regional parameter values, and equivalent tee circuit.

The effects of external parameters may be applied very easily since

$$V_e = V_{ee} - I_e R_e \tag{17}$$

and

$$V_c = V_{cc} - I_c R_c, \tag{18}$$

where V_{ee} and V_{cc} are supply voltages; V_e and V_c are measured from the appropriate terminal to the far end of the external base resistor. External R_b adds directly to r_b. Thus the two equations become

$$V_{ee} = (r_e + R_e + r_b + R_b)I_e + (r_b + R_b)I_c \tag{19}$$
$$V_{cc} + (r_c - r_c''')I_{c0}$$
$$= (r_m + r_b + R_b)I_e + (r_b + R_b + r_c + R_c)I_c. \tag{20}$$

In manipulation of (19) and (20) it is often more easy to do so in the functional form,

$$V_1 = R_{11}'I_1 + R_{12}'I_2 \tag{21}$$
$$V_2 = R_{21}'I_1 + R_{22}'I_2, \tag{22}$$

with substitution at the evaluation stage. The R''s here include both device and circuit parameters.[12]

Solving (19) and (20) simultaneously, the following relationship between V_e and I_e is obtained:

$$V_e = \left[r_e + R_e + r_b + R_b - \frac{(r_b + R_b)(r_b + R_b + r_m)}{r_b + R_b + r_c + R_c} \right] I_e$$
$$+ \frac{(V_{cc} + I_{c0}(r_c - r_c'''))}{r_b + R_b + r_c + R_c} (r_b + R_b). \tag{23}$$

[12] Here the primes indicate generalized open-circuit driving-point resistances rather than reference to Region I. The duplication of symbols is regretted.

Equation (23) is general for the given circuit; it suffers, however, in difficulty in interpretation due to the numerous terms. In the regional evaluation to follow, approximations will be made which bring out the significant factors although decreasing the accuracy somewhat. The $(r_e - r_e''')I_{c0}$ term will be neglected. It is assumed also that large external base resistance R_b is employed.

Evaluation in Region I

In Region I, from Fig. 16, r_m is zero and r_e' is large so that $r_e' \gg (r_b + R_b)$. Also, by assumption, $r_b \ll R_b$. Applying these approximations, (23) becomes

$$V_{eI} \approx r_e'I_e + \frac{V_{cc}R_b}{R_b + r_c + R_c} . \qquad (24)$$

Equation (24) is the equation of a straight line, having slope r_e' and an intercept on the voltage axis at $(V_{cc}R_b)/(R_b + r_c + R_c)$. The small-signal input impedance is just the slope value r_e'.

The short-circuit case where R_c is zero is the most adverse device condition in the sense that the dc term will then be most dependent upon device parameters. When $R_c = 0$, (24) becomes

$$V_{eI} \approx r_e'I_e + \frac{V_cR_b}{R_b + r_c} . \qquad (25)$$

Evaluation in Region II

In Region II all parameters are finite and the only approximations which may be made are $r_b \ll R_b$ and $r_e \ll R_c$. Thus,

$$V_{eII} \approx \left[R_b - \frac{R_b(R_b + r_m)}{R_b + r_m} \right]I_e + \frac{V_{cc}R_b}{R_b + r_c + R_c} \qquad (26)$$

If R_b is not too large, it may be approximated that $(R_b + r_m)/(R_b + r_c) = \alpha$. Taking $R_c = 0$, thus,

$$V_{eII} \approx R_b(1 - \alpha) + \frac{V_cR_b}{R_b + r_c} . \qquad (27)$$

Equation (27) is also the equation of a straight line having the voltage axis intercept of $(V_cR_b)/(R_b + r_c)$ the same value as in Region I. The slope, $R_b(1 - \alpha)$, is negative provided $\alpha > 1$.

Evaluation in Region III

In Region III it may be assumed that $r_b \ll R_b$, $r_c''' \ll R_b$, and $r_m''' \ll R_b$. Other suitable approximations will depend largely upon the magnitude of R_c. From (23)

$$V_{eIII} \approx \left[r_e''' + R_b - \frac{R_b(R_b + r_m''')}{R_b + r_c''' + R_c} \right]I_e$$
$$+ \frac{V_{cc}R_b}{R_b + R_c} . \qquad (28)$$

If R_c is large, that is, large compared to r_e''' but small compared to Region II r_c, then (28) becomes

$$V_{eIII} \approx \frac{R_bR_c}{R_b + R_c}I_e + \frac{V_{cc}R_b}{R_b + R_c} . \qquad (29)$$

Under these conditions, the circuit is essentially independent of device parameters. This is useful where a high independence of device parameters is required, but does not focus the attention upon the device parameters as does the $R_c \rightarrow 0$ case. This is the condition under which the transistor might be operated when it is desired to obtain the maximum ON current, or conversely, the minimum internal switch resistance.

Where $R_c = 0$, (28) becomes,

$$V_{eIII} \approx [r_e''' + r_c''' - r_m''']I_e + V_c. \qquad (30)$$

Since r_e''' and $(r_c''' - r_m''')$ are quite small, the short-circuits currents may be very high. Where the transistor is considered as a switch between emitter and collector circuits, the "switch" voltage drop, as V_e, is given by the first term of (30).

Evaluation of Region I—Region II Transition

Earlier trigger sensitivities were mentioned as being the small voltage and current differences between the turning points of the negative-resistance characteristic and the stable operating points. The determination of the turning points and their stability is of great importance since it is usually desired to obtain maximum stable sensitivity. The voltage and current at the two turning points[13] have been given the subscript p and v for the low- and high-current conditions, respectively, as shown in the synopses (Figs. 17, 18, and 19). V_{ep} and I_{ep} in the short-circuit or $R_e = R_c = 0$ case, for example, may be obtained by a simultaneous solution of (24) for Region I and (27) for Region II. Thus,

$$V_{ep} \approx \frac{V_cR_b}{R_b + r_c} \qquad (31)$$

and

$$I_{ep} = 0. \qquad (32)$$

That the low-current turning point falls on the emitter current axis, i.e., $I_{ep} = 0$, is a consequence of the original assumption that $r_m = 0$ for $I_e < 0$ and $r_m > 0$ for $I_e > 0$. This is not a precise assumption, and the turning point will usually lie slightly in the positive emitter-current region. For very small triggers or more accurate calculations, consideration must be given to closer approximations of $r_m = f_1(I_e)$ and $R_{11} = f_2(I_e)$.

The consequences of (31) can be quite serious. V_c and R_b are, of course, fixed, but r_c is variable from unit to unit, with temperature and perhaps with life. The variability of V_{ep} can result in failure to trigger, self-triggering, or lock-up at high current.

Evaluation of Region II—Region III Transition

The high-current turning point for the short-circuit case is determined from a simultaneous solution of the pertinent equations for Regions II and III, (27) and

---

[13] Sometimes termed "peak" and "valley."

(30). Thus,

$$I_{ev} \approx \frac{V_c r_c}{R_b(1-\alpha)(r_c + R_b)} \quad (33)$$

$$V_{ev} \approx V_c\left[1 + \frac{r_c(r_e''' + r_c''' - r_m''')}{(r_c + R_b)R_b(1-\alpha)}\right]. \quad (34)$$

Where it may be approximated that $r_c \gg R_b$, as has already been done in bringing in α, (33) and (34) become

$$I_{ev} \approx \frac{V_c}{R_b(1-\alpha)} \quad (35)$$

$$V_{ev} \approx V_c\left[1 + \frac{r_e''' + r_c''' - r_m'''}{R_b(1-\alpha)}\right]. \quad (36)$$

In this order of approximation, V_{ev} is nearly equal to V_c. Any variation in the lower trigger point is primarily with I_{ev}, due chiefly to any change in α. It is interesting to note that the trigger point will move along the Region III curve given by (30).

$$\frac{V_{ev}}{V_{ep}} = \frac{[R_b(1-\alpha) + r_e''' + r_c''' - r_m'''](R_b + r_c)}{R_b{}^2(1-\alpha)} \cdot \quad (37)$$

If $r''' + r_c''' - r_m''' \ll R_b(1-\alpha)$, then (37) becomes

$$\frac{V_{ev}}{V_{ep}} = \frac{R_b + r_c}{R_b} \cdot \quad (38)$$

If R_b is very large, the ratio approaches unity with the implication of the existence of only two regions. This is equivalent to saying that the negative resistance becomes infinite over an infinitely short range. The proper choice of R_b in terms of (38) may well be a design problem where it is desired to have a high ratio of V_{ep} to V_{ev}, as in lockout circuits.

GENERALIZED COLLECTOR CIRCUIT

GENERALIZED EMITTER CIRCUIT

SIMPLIFIED COLLECTOR CIRCUIT

SIMPLIFIED EMITTER CIRCUIT

Fig. 18—Synopsis of collector negative-resistance characteristic and properties.

Fig. 17—Synopsis of emitter negative-resistance characteristic and properties.

Synopses of Negative-Resistance Characteristics

Synopses for the three negative-resistance characteristics are given in Figs. 17, 18, and 19. The solution and analysis procedures are the same as outlined for the emitter characteristic. It should be noted that the base characteristic is short-circuit stable in distinction with the emitter and collector characteristics which are open-circuit stable. It would have been more appropriate to solve the base circuit in terms of conductances rather than resistances. The magnitudes of negative resistance obtained in this connection are quite low, which may be misleading; the negative conductance is quite high, however, which is desired in short-circuit stable negative-resistance circuits.

The ratio of V_{ev} to V_{ep} is often of interest to estimate voltage swings or perhaps as a design objective in some switching circuits. Thus from (36) and (31),

Care should be taken in the literal employment of the approximate regional relationships in Figs. 17, 18, and 19. They are very definitely approximate, and are intended to illustrate behavior and the limiting condition only to bring out the relative importance of device parameters. It is suggested that calculations be started with the general case and approximations be made as are valid. For example, the conclusion is reached in the collector characteristic that the negative resistance (Region II) is independent of the base resistance or feedback. This is true for only the limited range where $r_e \ll R_b \ll r_c$.
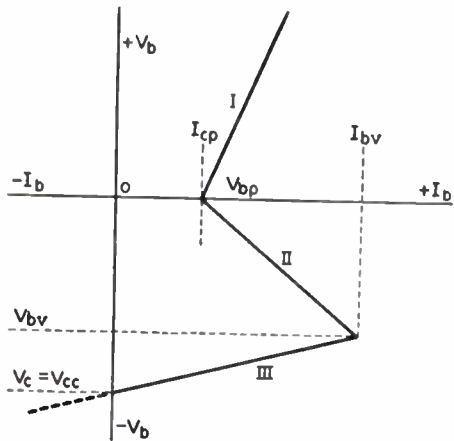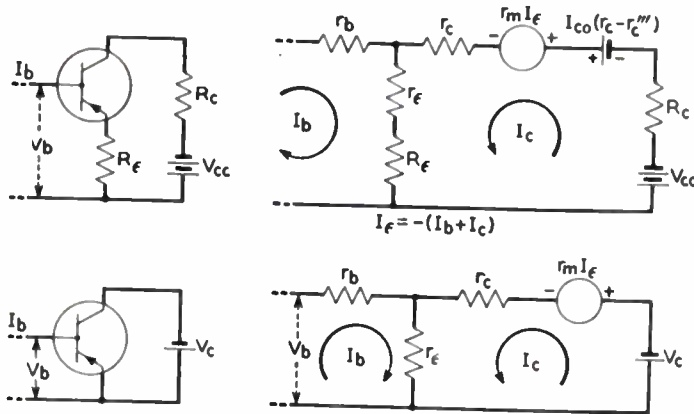
Fig. 19—Synopsis of base negative-resistance characteristic and properties.

Example of Calculated and Experimental Characteristics

An example to illustrate the analysis is shown in Fig 20 where both experimental and calculated characteristics for the emitter circuit are given. In this example there is appreciable load resistance; hence, r_c''', r_e''', and r_m''' are of no consequence since they will all be very small compared to the R_e of 2.2 K ohms. Also, $R_b = 6.8$K ohms is much greater than r_b; hence, r_b can be neglected. Since V_c is -45 volts, the I_{co} term may also be neglected.

Computing V_{ep} first,

$$V_{ep} = \frac{V_c R_b}{R_b + r_e + R_e} = \frac{-45(6.8\text{K})}{(6.8\text{K} + 19\text{K} = 2.2K)}$$
$$= -10.9 \text{ volts.} \tag{39}$$

The calculated value of -10.9 volts compares quite favorably to the measured -11.0 volts.

Region II is given in this case, approximately, by

$$V_e = \left[R_b + \frac{R_b(R_b + r_m)}{R_b + r_e + R_e} \right] I_e + \frac{V_c R_b}{R_b + r_e + R_e}, \tag{40}$$

or

$$V_e = \left(6.8K + \frac{6.8K(6.8K + 50K)}{6.8K + 19K + 2.2K} \right) I_e - 10.9, \tag{41}$$

$$V_e = (-8.9K)I_e - 10.9. \tag{42}$$

The first term is of course the slope in Region II and is the magnitude of the negative resistance. The calculated value is -8900 ohms whereas the measured value was approximately -9200 ohms.

The Region III approximation, derived also from the general relationship, is

$$V_e = \frac{(R_b R_e)}{R_b + R_e} I_e + \frac{V_c R_b}{R_b + R_e}, \tag{43}$$

$$V_e = \left(\frac{(6.8\text{K})(2.2\text{K})}{6.8\text{K} + 2.2\text{K}} \right) I_e = \frac{45(6.8\text{K})}{6.8\text{K} + 2.2\text{K}}, \tag{44}$$

or

$$V_{eIII} = (1785)I_e - 34. \tag{45}$$

The relation for Region III agrees quite well in slope but not in dc value, as may be seen in Fig. 20. Since in this example the Region III behavior is determined essentially by the circuit parameters, it is surmised that the nominal 45-volt battery employed in taking the data was actually 47 volts.
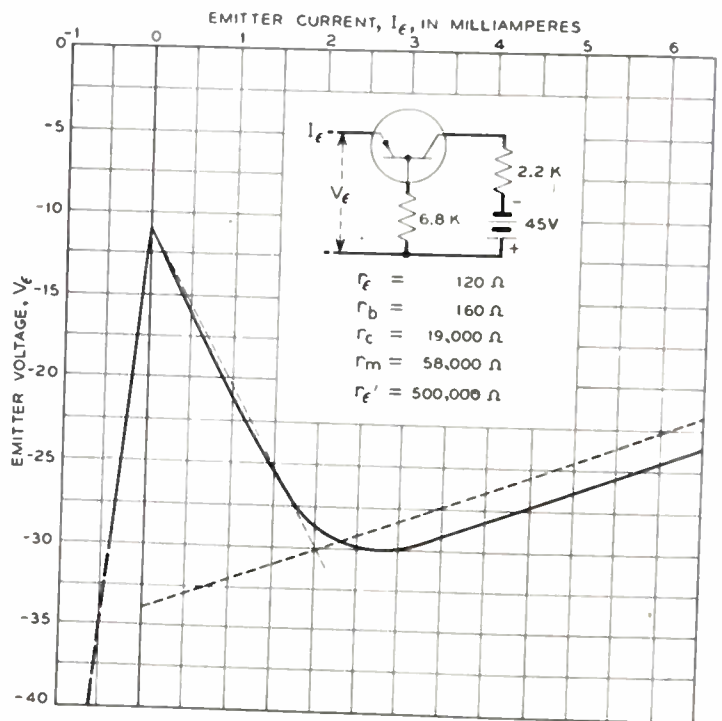
Fig. 20—Experimental and calculated emitter negative-resistance characteristic.

The Region I check is essentially perfect since the approximation given in Fig. 17 is quite good.

Note the error at the intersection of Regions II and

III. The broken-line method predicts a sharp transition whereas the actual case is gradual. The deviation is due to the gradual changes in r_m and r_c as the collector voltage approaches cutoff and is the largest gross error in the approximation.

Summary—Analysis

It is believed that analysis of this sort will reasonably predict circuit behavior and lead to device requirements. There must be a thorough understanding of the approximations involved, and the accuracy will be directly related to the degree to which the original idealized characteristics are approximated. Extended, by means of more than three broken lines, the method will yield fine detail to the degree to which device parameters are known and patience will permit. Transient behavior and analysis have not been discussed and are needed for a more complete understanding, particularly where transitional speeds are of concern.[14]

III. Switching-Type Transistor Properties

An examination of the circuit approximations given in Figs. 17, 18, and 19 will reveal that the transistor and circuit designers will want to know nearly all there is to know about device characteristics. This is not particularly surprising since the device is used over its entire range rather than over a limited portion as in the case of small-signal applications. The same examination of the circuit relations will also show that virtually all of the device parameters should be constant from unit to unit and with ambient conditions.

It can be shown that for small signals a device may be uniquely characterized by five measurements. In terms of the parameters used here these might be R_{11}, R_{12}, R_{22}, R_{21}, and the dc bias point or, equally, r_e, r_b, r_c, r_m, and the bias point. Since the problem was linearized in the approximation, it follows that 15 such measurements, five in each region, are necessary for proper switching-device characterization. The indicated extensive testing required may be reduced somewhat by suitable approximations. It is clear, however, that the switching-device designer and producer must reconcile themselves to making more tests for accurate characterization than when small-signal devices are concerned.

What will be given here is a description of typical developmental switching transistors in terms of the parameters which have evolved as a result of practical approximations. The method will be to discuss device properties and measurements region by region, then to discuss the properties at the transition points. Temperature, frequency, and life behavior will be taken up separately.

Region I Properties

In Region I, the emitter current is negative. Hence the emitter resistance r_e' is large, and is essentially that of a diode in the reverse direction. At present r_e' is measured by a simple dc test of the current which flows at a nominal -10 volts. Both r_e' and r_b will be discussed further under the Region I—Region II transition properties.

The Region I collector resistance is one of the most important parameters in switching. This is because of its determining nature in the turning-point voltages in Figs. 17, 18, and 19. Actually, what is of concern is not the small-signal slope shown as r_c[15] in Fig. 21 but rather the dc current and voltage relationship shown as r_{c0}. For example, in Fig. 17 it may be seen that V_{cp} is given by the voltage drop determined by the product of R_b and the dc collector current.
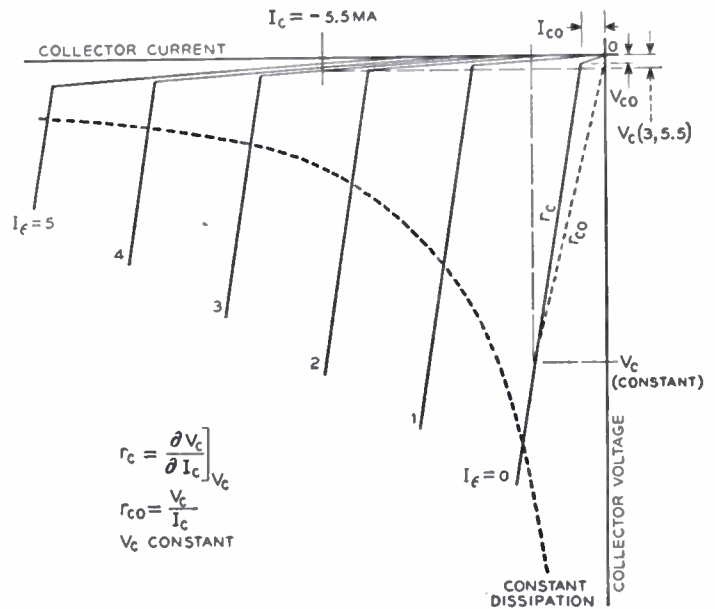
Fig. 21—Idealized output characteristic illustrating parameters.

Fig. 21 is an idealization of the R_{22} characteristic, and has been designed to bring out the diode nature of the collector by emphasizing the saturation current and voltage, I_{c0} and V_{c0}. In junction devices the break in the $I_e = 0$ characteristic at I_{c0} is quite evident, whereas in present point-contact devices the transition is smooth due to the much lower values of r_c. The device significance is the same, however; I_{c0} varies rapidly with temperature whereas r_c varies at a considerably lower rate.

In junction devices the proper measurements would be of I_{c0} and r_c. Since I_{c0} is difficult to define in point-contact devices, r_{c0} has been measured as an approximation. In the idealization, r_c and r_{c0} are related, as

$$r_{c0} \doteq \frac{(I_c - I_{c0})}{I_c} r_c. \tag{46}$$

---

[14] A treatment of the transient behavior between regions is given in B. G. Farley, "Dynamics of transistor negative resistance," Proc. I.R.E., vol. 40, pp. 1497–1508; this issue. Analysis and solution for the periods of the monostable and astable cases, assuming infinite region-to-region transition speed, are given in G. E. McDuffie, Jr., "Pulse duration and repetition rate of a transistor multivibrator," Proc. I.R.E., vol. 40, pp. 1487–1490; this issue.

[15] The actual parameter is of course R_{22}; but since $R_{22} = r_c + r_b$ and $r_b \ll r_c$, R_{22} is taken as r_c.

The measurement of r_{c0} is made at a collector voltage which is typical of the applications in the range of perhaps -10 to -45 volts.

A constant dissipation line has been drawn on Fig. 21, which reveals the desirability of having r_{c0} very large in order to operate at higher voltages and to secure high efficiency through lower dissipation in the OFF or rest condition.

Region II Properties

The Region II low-frequency properties are essentially identical to those of transistors intended for small-signal applications. A possible exception is the somewhat less attention paid to the base resistance, r_b, which is critical to small-signal applications. The characterization consists of a normal small-signal set plus dc bias values. Typical distributions are given later.

Region III Properties

The Region III properties have been defined largely by a figure-of-merit measurement shown as $V_c(3, -5.5)$ in Fig. 21. This measurement is the voltage from collector to base under the condition that $I_e > -I_c/\alpha$. In this instance I_e has been chosen to be 3 ma and I_c to be -5.5 ma. The collector current value is chosen on the basis of the smallest tolerable value of alpha expected so as to place the point of measurement near the R_{22} knee, but definitely in Region III or overload.

The $V_c(3, -5.5)$ measurement is a good measurement for defining the general behavior. $V_c(3, -5.5)$ taken with the r_{c0} measurement constitute a very good defining set for checking the transistor as in remeasuring For design purposes, the $V_c(3, -5.5)$ measurement is not sufficient. It provides an approximate value for r_c''', but does not define r_e''' and r_m'''. A second dc measurement, the collector to emitter voltage drop, V_{ce}, has been employed experimentally also. An improved characterization will undoubtedly involve separate measurements of r_e''', r_m''', and r_c'''.

Region-to-Region Transition Properties

The transition between Regions I and II is accompanied by abrupt changes in r_e and r_m.

The theory assumes that both of these parameters change at an infinite rate at a fixed emitter current, taken as $I_e = 0$. Unfortunately, neither of these assumptions is strictly true; r_e undergoes a gradual change from high to low values, which is only approximated by the three assigned values. In particular, the behavior near $I_e = 0$ is of concern when dealing with small triggers.

The forward transfer impedance changes at a finite rate also. Furthermore, the emitter current at which the maximum rate of change occurs will vary from unit to unit. Present practice also has been to measure α rather than r_m. The rational for doing so is not too good since

r_m is quite likely the better parameter to characterize. Alpha, however, has a strong physical appeal, fits well into the circuit problems, and is easy to measure.

Since $\alpha = (r_b + r_m)/(r_b + r_c)$, it is necessary to assume that r_b and r_c are constant near $I_e = 0$, an only fair approximation. Having made the approximation, the typical α behavior shown in Fig. 22 may be taken as a measure of r_m. Three values are measured, the first of which, α_1, in Region II, is redundant to the Region II small-signal measurements. The two limits, α_2 and α_2, serve to place lower and upper limit on the absolute values of α at the Regions I–II transition. These limits in turn place a lower value on the rate of change in α within the $I_e \pm \Delta$ range shown.
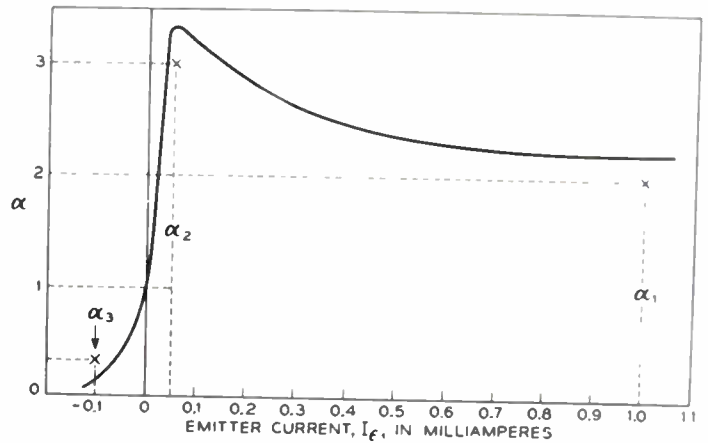
Fig. 22—Alpha characteristic.

It may be noted that α in Region I is finite. There is a lower limit even though r_m is zero since $\alpha_1 \rightarrow (r_b/r_b + r_c)$. The values normally encountered at $I_e = 0 - \Delta$ are usually in excess of this lower limit.

The feedback resistance r_b tends to rise as $I_e \rightarrow 0$, which may be important to some trigger circuits. As the circuitry becomes more sophisticated, it is expected that more attention will need to be paid to the behavior of r_e, r_m, and r_b at and near $I_e = 0$.

The transition from Region II to Region III is determined from the relation $I_e = -(I_c/\alpha)$. The problem is quite similar to the control of the μ factor in tubes where plate-current cutoff is given by $V_g = -(V_p/\mu)$. Present practice has been to depend upon the α_1 values and upon the lower limit placed on alpha in the $V_c(3, -5.5)$ measurement. Further effort is needed here also.

Typical Parameter Values and Distributions

Integrated distribution curves for the parameters of a typical developmental switching transistor are shown in Fig. 23 The unit-to-unit variations are deemed to compare favorably with those of commercial electron tubes. The parameter of most serious variability is r_{c0}, which is unfortunate since r_{c0} is so important to trigger sensitivity stability.

Temperature, Frequency, and Shock Properties

Transistor parameters are reasonably constant with temperatures below room temperature. Above room temperatures some of the parameters are variable: r_e and r_b are fairly constant, changing very little to 70°C; r_c and r_m decrease fairly rapidly, maintaining a ratio so that alpha rises slightly; r_e' and r_{c0} change most rapidly and, while both of these parameters are of little consequence in small-signal applications, they are quite important in switching, particularly r_{c0}.
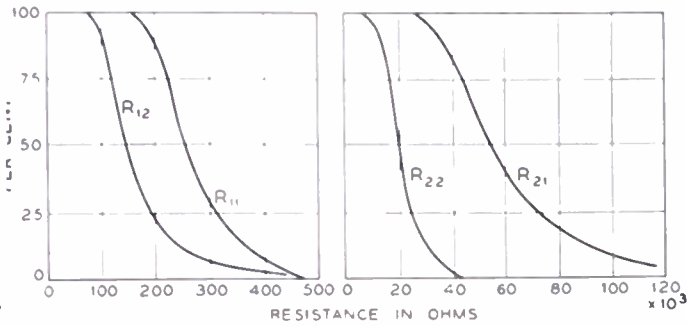
Fig. 23—Variation in parameters of developmental switching-type transistor (M1689).
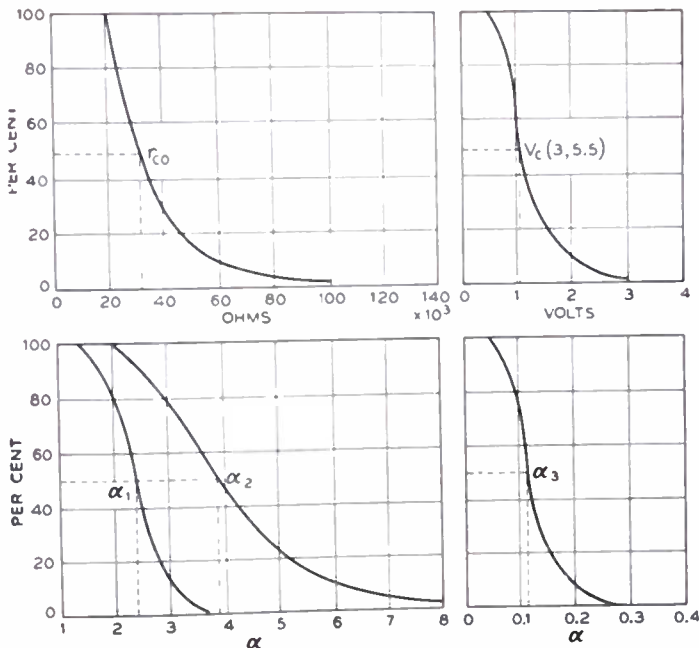
Fig. 23(a)

Early transistors might exhibit a change in r_{c0} at 60°C of 3-to-1 or more from room temperature values. The transistors, of which the data in Figs. 13 and 22 are typical, have an r_{c0} temperature coefficient of about −0.75 per cent/°C; that is, the room temperature value of r_{c0} might be reduced by 30 per cent at 70°C. The improved temperature behavior implies a corresponding reduction in variation in trigger sensitivity. Parameter values, large-signal and small-signal, are shown in Fig. 24 as a function of temperature.

Variation in characteristics will arise from self-engendered heat, that is, dissipation. Transistors may be thermally unstable under constant-voltage conditions. Since the switching properties are exhibited under short-circuit or constant-voltage terminations, thermal properties are of concern. The limitations involved are similar to those of any positive feedback circuit. If the thermal loss through radiation and conduction exceeds the heat input, the system will be stable. The practical significance is to place limitations on dissipation and to employ designs which result in rapid heat loss. Other design criteria, such as miniaturization, may limit the latter.
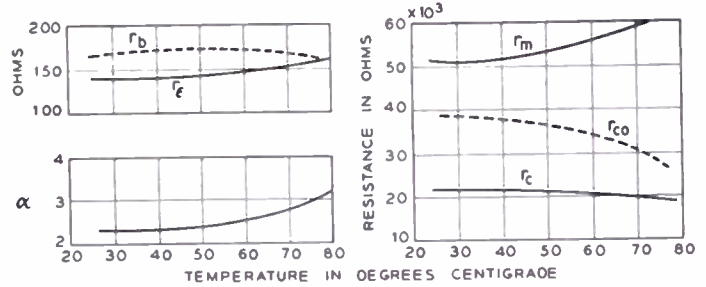
Fig. 24—Temperature behavior of characteristics of developmental switching-type transistor (M1689).

If perfect switching characteristics were obtainable, dissipation would be of little consequence in switching. This is akin to saying that neither a short-circuit nor an open-circuit dissipates any energy. Furthermore, the perfect device has zero transition time and therefore involves no loss. The transistor has finite resistance, both open and closed, and a finite although rapid transition time. There is some advantage, however. A constant dissipation curve shown as a dotted line has been included in Fig. 21. Small-signal operation at mid-range currents and voltages results in fairly low limitations on both current and voltage. The intersection with the R_{22} voltage saturation line ($I_e = 0$) is at fairly high voltage. Similarly, the intersection with the collector voltage cut-off line is at high current. For constant dissipation, approximately,

voltage saturation: $P_d \approx \dfrac{V_c^2}{r_{c0}}$

voltage cutoff: $P_d \approx I_c^2(r_e''' + r_c''' - r_m''').$ [16]

Depending upon the circuit, the assumed dissipation limit may or may not be exceeded during the transitions. Should the limit be exceeded, and substantially so, there are normally no serious consequences due to the very rapid transitions and consequent low thermal energy generated.

Transistors may not be able to tolerate excess dissipation on this basis if the circuits are slow, that is, with transition times in excess of perhaps a few tenths

[16] This includes both emitter and collector dissipation. See (30).

of a microsecond. Such conditions may arise, for example, if loads are inductive. In many such cases, shunting capacitor networks will often permit a rapid transition with consequent transfer of current to the inductive load.

The frequency response of point-contact transistors can be sufficiently good to insure switching-type operation with rise times of the order of 0.1 to 0.01 μs. Fall times may be somewhat longer because of the hole storage effect. In regenerative circuits, operating speeds are faster than might be imagined from the small-signal frequency cutoff. Reliable operation with rise times of 0.1 μs is obtained with only nominal attention to frequency cutoff. Speeds of the order of 0.02 μs require a 10-mc lower limit. Present junction transistors are substantially slower.

Accurate life estimates are difficult to make because of the rapid rate of development, the relative age of the transistor, and the number of parameters involved. A given device is quite likely to be obsolete and forced to give way to an improved version before sufficient models can be obtained for life tests. A small quantity of transistors having properties similar to those of Figs. 20 and 21 have been operated for over 6,000 hours with an indicated life of 30,000 hours. Other similar transistors with longer life histories have indicated lives of better than 70,000 hours. The pattern appears to be similar to that of electron tubes—an early failure and change rate followed by a very slow exponential rate. It is believed that life is extended by low-power operation and is decreased by high-temperature operation.

The relatively high noise level of transistors does not appear to be a significant problem, at present, when considered in terms of automata. Systems employing switching-type circuits in pulse communication will, of course, be concerned. It is suggested that the non-concern for noise in nontransmission type systems is largely a reflection of the ease with which high magnitudes of state changes are obtained. With design trends toward low-power and low-operating levels, noise will undoubtedly set a lower limit of level operation in such systems also.

The extreme resistance of the transistor to shock and vibration with a consequent absence of microphonism may, in some applications, result in effective lower noise. Shocks in excess of 20,000 grams have resulted in no damage. No evidences of current modulation in excess of noise have been detected with vibrational forces of the order of 100 grams at frequencies as high as 1,000 cycles in tests on the transistor of Fig. 1. Transistors have been included in plastic-embedded circuits without change of characteristics.

Summary—Transistor Properties

Transistors have been designed with properties expressly intended for switching applications.[17] The characteristics are acceptable for contemporary switching-type circuits and sufficiently reproducible to permit interchangeability of devices in circuits of normal requirements. The characterization has been sufficiently unique to permit the calculation of first-order circuit performance. The characterization is not sufficiently complete to permit determination of the complete transient behavior.

In terms of the circuits described, the major parameter limitation is concerned with the variability of the dc collector resistance among units and with temperature. It is expected that future circuit development will place additional requirements on the transistor, particularly as related to the transitions between regions. It is also to be expected that future circuit designs may establish new or modify present device requirements.

A major consideration for computer or computer-like systems, reliability, particularly with respect to time and temperature, has not been established, but appears to be favorable.

[17] See footnote reference 3, also.

Graphical Analysis of Some Transistor Switching Circuits*

L. P. HUNTER† AND H. FLEISHER†

Summary—Methods for generating the entire input characteristic for the various terminals of a transistor are given together with procedures for the graphical analysis of three direct-coupled transistor circuits. The circuits analyzed are a base input amplifier, a collector to emitter direct-coupled switching circuit, and a collector to base direct-coupled circuit.

Introduction

SEVERAL PAPERS[1,2,3] have been published dealing with the analysis of transistor circuits using small-signal four-terminal-network theory. Such analysis cannot easily be used for switching circuits because the large signals involved require many oversimplifying assumptions about the linearity and uniformity of the transistor characteristics. The purpose of this paper is to show a method of handling graphically the basic transistor collector and emitter characteristics to analyze some useful switching circuits without having to approximate or idealize the measured characteristics.

The basis of most switching or trigger circuits is regeneration or a negative resistance characteristic. Methods for graphically determining the resistance value of the negative resistance characteristics of some transistor circuits have been given before.[4] Methods of generating the whole characteristic containing the negative resistance region by using only the basic collector and emitter characteristics have not been shown.

In order to illustrate the method of generating base and emitter input characteristics, the performance of the circuit shown in Fig. 1(a) will be analyzed. This circuit represents a base input switching circuit. First, it is necessary to determine the input characteristic measured at the terminals A-A with the terminals B-B shorted, and the voltage of the generator $E_b = 0$. To do this, the conventional load line is drawn on the collector characteristic shown in Fig. 1(b), and this load line is then transferred point by point to the emitter characteristic (at the top of Fig. 1(c) using the emitter and collector currents as co-ordinates). This transferred load line is now the desired emitter input characteristic under the conditions given above.

Next, the input characteristic at the terminals B-B will

be determined with the terminals A-A shorted, and the voltage of the generator $E_b = 0$. Since the collector and emitter currents add algebraically in the base lead and since with A-A shorted the base voltage is also the emitter voltage, the desired base characteristic (shown at the bottom of Fig. 1(c)) may be obtained by adding the collector and emitter currents algebraically for each point of the transferred load line. In doing this the further assumption is made that $|V_e| = |V_b| \ll |E_c|$ which is usually true.
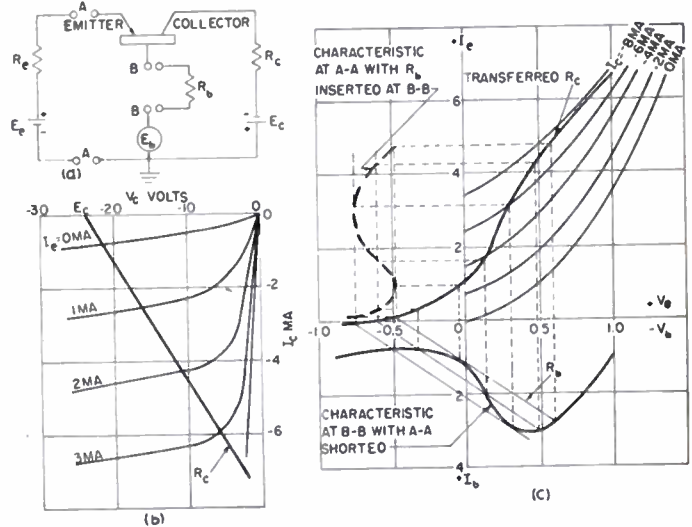
Fig. 1—Generation of special base and emitter input characteristics. (a) Circuit diagram. (b) Collector characteristic. (c) Composite emitter-base characteristic.

It remains to generate the characteristic found at the terminals A-A with the resistance R_b inserted at the terminals B-B. To do this, several load lines representing R_b with different test values of V_b must be drawn. These are shown at the bottom of Fig. 1(c). Since it is assumed that all of these values of $|V_b| \ll |E_c|$, we may consider the test generator producing this displacement of the R_b load line as connected across the terminals A-A. The currents through the terminals A-A corresponding to these test voltages may now be found by tracing the intersection points of the R_b load lines and the base characteristic vertically back to the transferred collector load line. The currents found in this manner may now be plotted vertically above the voltage intercept of the appropriate R_b load line. Such a characteristic is shown as the heavy dashed line at the left of Fig. 1(c).

The method of obtaining the stable dc operating point in the emitter plane is shown in Fig. 2(a), where the emitter resistor is drawn in as a conventional load line. It may be seen that if R_e is sufficiently large, the

[1] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, pp. 367–401; July, 1949.
[2] W. M. Webster, E. Eberhard, and L. E. Barton, "Some novel circuits for the three terminal semiconductor amplifier," *RCA Rev.*, vol. 10, no. 1, pp. 5–17; March, 1949.
[3] R. L. Wallace, Jr. and W. J. Pietenpol, "Some circuit properties and applications of n-p-n transistors," PROC. I.R.E., vol. 39, pp. 753–767; July, 1951.
[4] L. P. Hunter, "Graphical analysis of transistor characteristics," PROC. I.R.E., vol. 38, pp. 1387–1391; December, 1950.

intersection with the emitter input characteristic will be at a single point even if the emitter input characteristic has a negative-resistance region. An emitter load line, $R_e{}^*$, is also drawn in showing that, with a suitable choice of both emitter resistance and battery, bistable operation of the circuit may be obtained.
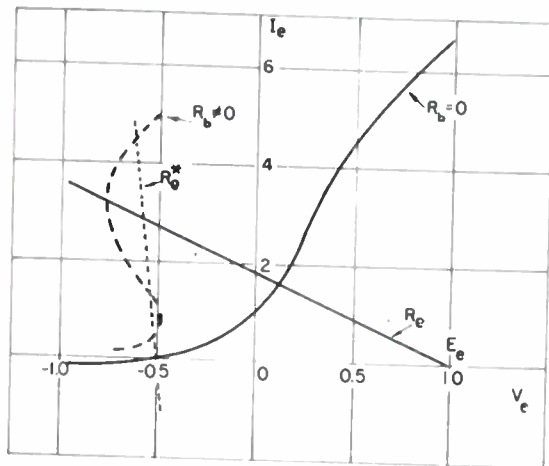
(a)
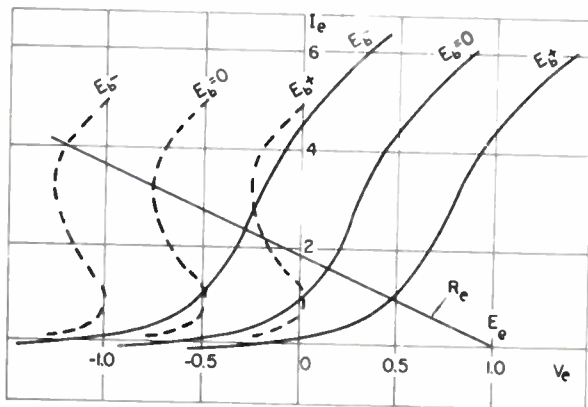
(b)

Fig. 2—Effects of emitter load (a) and base signal (b) on the emitter operating point.

The translation of the emitter input characteristic parallel to itself along the voltage axis as a result of a base signal E_b, of magnitude less than $|E_e|$ is shown in Fig. 2(b). The base signal is assumed to come from a zero impedance generator, and the dashed lines refer to the circuit containing R_b. The translation to the right occurs for a positive base signal (positive relative to ground). The emitter load is the solid straight line.

Inspection of the operating points indicated in Fig. 2(b) shows that without base feedback, and for an input signal excursion from $E_b{}^+$ to $E_b{}^-$, the emitter current will change from a small positive value to a somewhat greater value, resulting in a corresponding change in collector current. However, with the positive feedback obtained by introducing R_b, and for the same excursion of the input signal, it is observed that the change in emitter current is considerably greater, driving the collector to saturation in this particular case.

In the example considered above, both the base characteristic and the emitter input characteristic for R_b inserted in the base lead show negative-resistance

regions. The condition necessary for such behavior will become apparent upon considering first the base characteristic. Since the base current is obtained by the algebraic addition of the emitter and collector currents, it is clear that it is necessary for the collector current to increase in absolute value along the transferred load line faster than does the emitter current in order to produce a negative-resistance region in the base characteristic. Since the transferred load line corresponds point by point (on the current co-ordinates) to the actual load line in the collector characteristic plane, the same condition must be true along the load line. Thus, if a directional derivative is defined as

$$\alpha' = -\left.\frac{\partial I_c}{\partial I_e}\right|_{\text{along the load line}}$$

then the necessary and sufficient condition for a negative resistance region to appear in the base characteristic is $\alpha' > 1$. It follows from this that, in the case of a short circuit load line, a transistor must have $\alpha > 1$ in order to be able to exhibit a negative base-input impedance. (Here α has the usual definition of short circuit current gain.) By inspection of Fig. 1(c), it can be seen that it is possible to choose an R_b sufficiently large to intersect the base characteristics in three points if the base characteristic contains any negative resistance region at all. And, thus, a negative resistance input characteristic at the terminals A-A can be obtained using any transistor of $\alpha > 1$.

COLLECTOR TO EMITTER COUPLING

The graphical representation of the volt-ampere characteristics may also be applied to other circuit components whether linear or nonlinear, and in this way, entire circuits may be analyzed. As an example, the direct coupling between the collector of a transistor and the emitter of another transistor will be analyzed.

Direct coupling may be desirable for several reasons: (1) The state of the system is preserved for an arbitrary period of time; and (2) the number of components required may be materially reduced. A direct coupled circuit is shown in "exploded" form in Fig. 3(a), and the sequential volt-ampere plots are shown in Figs. 3(b) to (e).

The collector characteristics for transistor T_1 are shown in Fig. 3(b). Two of these are labelled "off" and "on," indicating, respectively, the collector characteristic for zero emitter current, and the collector characteristic for sufficient emitter current to drive the collector into saturation. These are observed on looking into terminals A-A from the right, isolating the transistor T_1 from the circuit.

If a battery, E_2, is inserted in the base lead of T_1 with the polarity shown, then the set of characteristics obtained on looking into terminals A-A' from the right is that shown in Fig. 3(c). As can be seen on comparing Fig. 3(b) and (c), this amounts to translating the collector characteristics to the right along the voltage axis by an amount equivalent to E_2.

The branch of the network consisting of the resistor R_1 in series with the battery E_1 is also plotted in Fig. 3(c). This branch is shown across the terminals B-B or A-A'. The combination of the branch volt-ampere characteristic with the translated collector characteristic is obtained by algebraically adding currents at a common voltage since the two are in parallel, and is shown in Fig. 3(d). This is the set of characteristics obtained on looking into terminals B-B from the right.
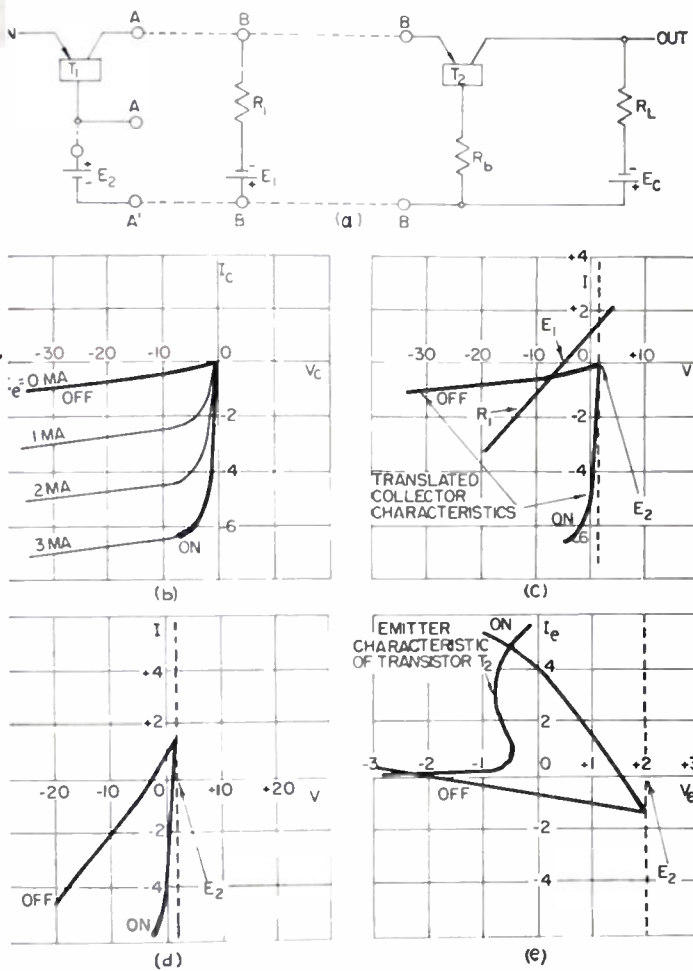
Fig. 3—Graphical analysis of the direct coupling of two transistors. (a) Circuit diagram. (b) Collector characteristic of transistor T_1. (c) Translation due to E_2. (d) Composite with E_1 and R_1. (e) Loading on the emitter of T_2.

At this point, the stage to be driven must be considered. The emitter volt-ampere characteristic (the transferred load line) of transistor T_2 is seen by looking into terminals B-B from the left. This is shown in Fig. 3(e). The modified collector characteristic of T_1, the driving transistor (Fig. 3(d)), may be considered as the load on the driven transistor's emitter. Graphically, this is accomplished simply by reflecting the characteristic plot in the voltage axis, see Fig. 3(d) and (e). Thus, when T_1 is "off," the emitter of T_2 is held at a very small value of current, and when T_1 is "on," T_2's emitter is held at a value of high forward emitter current. By using a high α transistor as the driver, several emitters may be directly coupled in this manner.

COLLECTOR TO BASE COUPLING

As a final illustration of the graphical approach to the design and analysis of transistor switching circuits, consider a "regenerative" current amplifier using two transistors in a collector to base coupling arrangement. This circuit is shown in Fig. 4(a), and the base input characteristic is shown in Fig. 4(b). The polarity sense
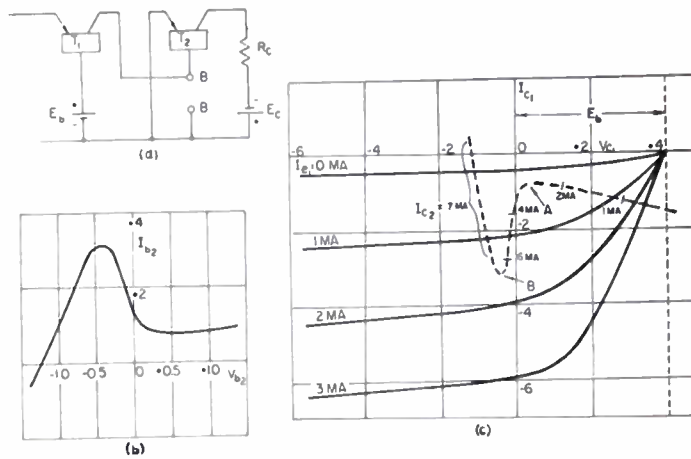
Fig. 4—Graphical analysis of a regenerate current amplifier. (a) Circuit diagram. (b) Base characteristic of T_2. (c) Loading of the collector of T_1.

of both the base voltage and base current is preserved from the plot of Fig. 1, the axes having been redrawn to correspond to the conventional orientation. This base input characteristic is obtained on looking into terminals B-B from the left.

The transposed[5] collector characteristic of T_1 is observed on looking into terminals B-B from the right. This is shown in Fig. 4(c). The base input characteristic of T_2 may be considered as the load line for T_1 by reflection in the voltage axis. This is shown as a dotted line in Fig. 4(c).

It is to be noted that the maximum point of this base input characteristic, labelled "A" in Fig. 4(c), corresponds to minimum collector current for transistor T_2, while the minimum point, labelled "B," corresponds to saturation collector current for T_2, that is, the collector current will not increase in magnitude appreciably beyond this value even if the base voltage is made more negative than its value at "B." Thus, if the characteristics of transistor T_2 are assumed to be those shown in Fig. 1, with the same load R_c, as sketched in Fig. 1(b), it may be seen that the saturation current for T_2 will be of the order of 7 ma, and the minimum current will be of the order of 1 ma.

As shown in Fig. 4(c), the controlling emitter current T_1 needs to swing through an excursion of about 1.5 ma to change the collector current of T_2 by 6 ma. Thus, at $I_{e_1} = 0$ ma, $I_{c_2} \approx 7$ ma, and at $I_{e_1} \approx 1.5$ ma, $I_{c_2} \approx 1$ ma, yielding an over-all current ratio

$$\left| \frac{\Delta I_{c_2}}{\Delta I_{e_1}} \right| = 4.$$

[5] As discussed for the circuit shown in Fig. 3.

With appropriate circuit constants, including the choice of E_b, this over-all current ratio may be readily increased to 10 or more.

If the quiescent operating condition of the circuit occurs at $I_{e_1} = 1$ ma, for this example, then the circuit may be used in a bistable manner, as a trigger. Thus, assume that the circuit is "off," for example, for $I_{e_1} = 1$ ma, $I_{e_2} = 1$ ma. Increasing the current, I_{e_1} to about 1.5 ma, will keep the second collector "off," but decreasing I_{e_1} to zero current will trigger the second transistor to the "on" state, that is, $I_{e_2} = 7$ ma. The circuit remains on when returned to the quiescent state (by allowing I_{e_1} to increase to its quiescent value of 1 ma).

To turn "off" the circuit, the emitter current of T_1 must be increased to about 1.5 ma, and when the circuit is returned to its quiescent state, it will remain "off."

Since the voltage E_b serves to translate the volt-ampere characteristic of the base of T_2 parallel to itself along the voltage axis, it is clear that the base of T_1 may

also be used as an input signal point, provided the proper current bias is applied to the emitter of T_1. In the example considered above in connection with Fig. 4, such an emitter bias would have to be at least 1.5 ma for proper triggering in either direction by a voltage signal in place of E_b.

CONCLUSION

Three examples of a graphical approach to switching circuit analysis have been discussed. This method utilizes the volt-ampere characteristics of both linear and nonlinear devices, and, it is felt, is capable of handling a wide variety of coupling conditions. Furthermore, the results of this analysis are obtained by a method which deals with quantities which are directly observable. Such an approach is felt to be preferable to conventional small-signal network analysis when extrapolated to large signal circuits.

A Transistor Reversible Binary Counter*

ROBERT L. TRENT†, MEMBER, IRE

Summary—The feasibility of performing a fairly complex switching function using a few elementary transistor circuits is illustrated and experimentally verified. The specific function discussed is reversible binary counting. The mechanism used to achieve reversibility and the circuitry within each building block is described. Operating margins and suggestions for design improvements for systems application are given.

INTRODUCTION

THE PURPOSE of this paper is to illustrate the feasibility of accomplishing a fairly complex switching function by the use of a few elementary transistor circuits. This work was part of an investigation sponsored by the Joint Services to determine the feasibility of the development of packaged, functional, miniaturized switching-type circuits employing transistors. Specifically, the function to be described is that of a reversible binary counter.

A binary counter may be defined as a counting mechanism which is capable of translating a series of input pulses into a binary code corresponding exactly to the sum of the number of impressed pulses. A reversible binary counter (RBC), in addition, should be capable of performing the algebraic function of subtraction as well as addition, under the influence of some control function. If the control function indicates that addition to the binary sum present in the counter stages is required, the usual forward cycling will occur with the incidence of the next triggering pulse. However, if the control function indicates that subtraction is required, the next impressed pulse will cause the counter to cycle backwards, and a binary digit unit will be subtracted from the stored code.

METHOD OF OPERATION OF A RBC

Definitions and Conventions

An elementary block diagram illustrating a method of achieving reversibility[1] is shown in Fig. 1. For simplicity, only four stages are shown. The following conventions are assumed in describing this figure:
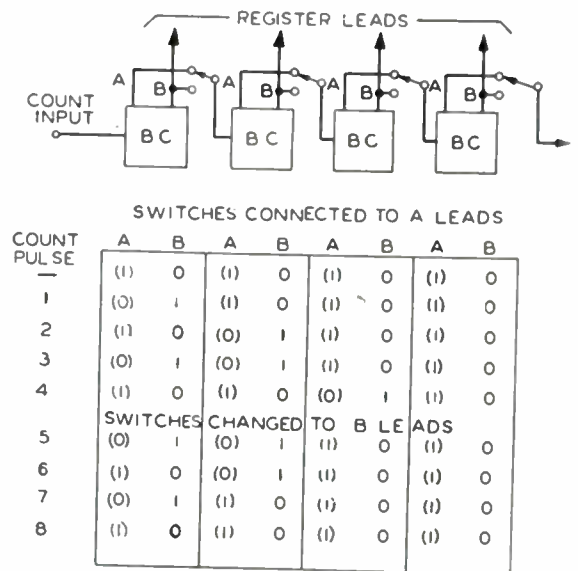
Fig. 1—Functional diagram of reversible binary counter.

1. The output leads of a stage are designated A and B; the states between which a lead will change are defined as 1 and 0, where 1 is more positive than 0.

[1] H. J. Gray, "Logical description of some digital-computer adders and counters," PROC. I.R.E., vol. 40, pp. 29–33; January, 1952.

2. The counter stages are arranged so that they re-pond only to *positive* pulses impressed on the count input leads.

3. The gating functions represented by simple switches on the figure act in unison, making simultaneous connections to all *A* or all *B* output leads under some control stimulus.

Operation

It is assumed that the state present on the registry leads *B* of all stages is 0 and that the control function has caused all of the interconnecting switches to operate to connect to the *A* output leads of the preceding stages to perform addition. A positive triggering pulse impressed on the input lead will cause stage 1 to assume the state 1 on its registry lead *B*. Since the *A* lead changed from 1 to 0, no positive pulse is transmitted to the following stage, and no further action follows.

Count-pulse 2 causes stage 1 to change state again, reverting to 0 on its *B* lead. Since the *A* output lead changes from 0 to 1, a positive pulse is passed to stage 2 causing it to assume the condition 1 on its registry lead. A 1 on the *B* lead of stage 2 has a weight of 2, corresponding to the number of pulses impressed on the count input lead.

Count-pulses 3 and 4 similarly cause further changes in state of the multistage binary counter, resulting in the codes being set up as illustrated in Fig. 1.

At this time assume that the control function receives information to the effect that the succeeding pulses should cause subtraction from the binary sum present on the counter stages. All of the interconnecting switches then make simultaneous connections to the *B* output leads of the preceding stages.

Count-pulse 5 causes stage 1 to change state so that a 1 appears on its *B* output lead, impressing a positive pulse on the following stage. Stage 2 on triggering assumes the state 1 on its *B* output lead, resulting in a positive pulse being passed to the third stage. The third stage then assumes the condition 0 on its *B* registry lead, and no further action follows. The binary code present on the registry leads of the counter stages corresponds to the code set up by the third count pulse, and the fifth count pulse has resulted in subtraction. Similarly count pulses 6, 7, and 8 cause the binary-counter stages to operate to revert to the starting condition.

Thus, a multistage binary counter, provided with switches so connected that the count-input lead of a following stage may be switched from one output lead to another, may be made reversible to perform subtraction as well as addition.

CIRCUIT ELEMENTS OF THE RBC

Switching or Gating Function

An equivalent functional representation of the RBC is shown in Fig. 2. One method of performing the required switching uses threshold-two gates, in which one of the enabling inputs is a step voltage from a control function and the other is the positive pulse obtained from the preceding binary-counter output lead. A transistor regenerative gated amplifier which is capable of performing this gating function has been developed and is described in Appendix A.

Binary-Counting Function

The requirements placed upon the individual counting stages for operation in a circuit as outlined in Fig. 2 are as follows:
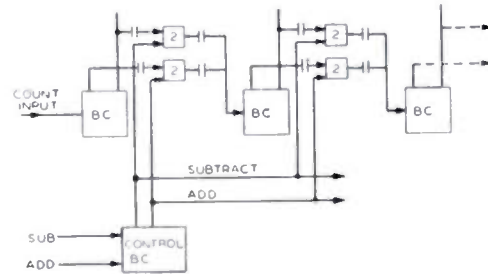
Fig. 2—Functional representation of reversible binary counter.

1. The stages must have two dc stable equilibrium points, and must be capable of being triggered from one dc stable point to the other by single polarity pulses applied at a common input.

2. The voltage-step function present upon each output lead must be capable, when differentiated, of providing a positive pulse of sufficient amplitude to trigger a gated amplifier when the latter is enabled, and must not be large enough to trigger it when inhibited.

The two-transistor stabilized binary-counter stage as covered in Appendix B meets these requirements. The margins of operation in meeting requirement (2) are discussed in Appendix C.

Control Function

The control function for the threshold-two gated amplifier must provide the proper biasing potentials in the form of two step voltages. One of these is applied to all the *A*-gates in parallel, the other to all the *B*-gates in parallel. It must be capable of changing state, reversing the step voltages, under the control of operate and clear information, usually designated Set 1 and Set 0. The operate or Set 1 information pulse will then impress voltages to one control of the gates so that the *A*-gates will be enabled if a positive pulse is impressed on the other input connection and will impress a positive pulse on the next binary-counter stage. The *B*-gates, however, will be inhibited. In a similar fashion, the clear or Set 0 signal will result in inhibition of the *A*-gates, and excitation of the control of the *B*-gates so that a positive pulse impressed on the other input connection will cause a positive pulse to be passed to the next binary-counter stage. The two-transistor stabilized binary-counter stage as covered in Appendix B will meet these requirements.

REVERSIBLE BINARY COUNTERS

Experimental Model

The circuit elements in the form of laboratory experimental circuits have been interconnected, as shown

in the block diagram of Fig. 3, to provide experimental verification of the feasibility of application of the particular transistor circuits. Decoupling networks have been incorporated in the step-voltage leads from the control function to prevent the sharp leading and trailing edges of the step-voltage waveform from causing spurious operations.
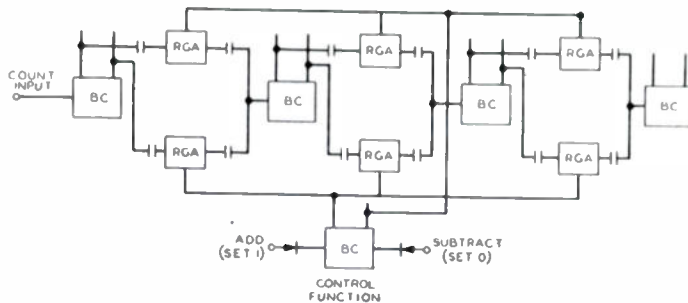
Fig. 3—Reversible binary counter schematic.

Successful operation of the RBC at repetition rates up to approximately 30 kc was obtained. This frequency limitation was caused primarily by the time-constant of the decoupling networks between the control outputs and the gated amplifiers. In addition, the cross-coupling networks between collectors and bases of the transistors in the binary-counter stages were optimized for low-frequency operation in the 600-cps range, and could be modified for operation at higher rates in the order of 100 kc.

Bias supply variations in the order of ±15 per cent were tolerable. The total input power for a 4-stage reversible binary counter using these circuits is in the order of 3.5 watts.

Packaged Model

The transistor circuits, packaged in the forms described in the appendices, have been interconnected to perform the function of reversible binary counting. In order to provide a more satisfactory display of the reversibility function, the circuit of Fig. 3 has been modified to provide the following additional features:

(a) A packaged gated-amplifier circuit (M1731-1) has been used as a pulse generator to provide a triggering-pulse source for the first binary-counter stage. The mode of operation of this packaged circuit when used as an astable multivibrator or pulse generator has been described in the literature.[2] An external capacitor is added from emitter circuit to ground, and the step-voltage input terminal is provided with a positive biasing potential. These constants have been adjusted to provide a pulse train with a repetition rate of approximately 3.3 kc.

(b) Two packaged amplifier circuits (M1733-1) have been used to obtain a delayed-pulse train, whose repetition rate is set by the afore-mentioned pulse generator. The circuit design is somewhat similar

to the gated amplifier except that the dc base stabilization feature has been omitted, and a simple form of stabilization of the triggering level used.[3] These circuits are triggerable by negative pulses, therefore the trailing edge of the output waveform of the pulse generator is used to trigger the first amplifier stage. The pulse length of the amplifier stage is adjusted to derive a pulse in the order of 100 μsec long. The trailing edge of this pulse is used to trigger the second amplifier stage, thereby achieving a pulse train, delayed by an interval in the order of 110 μsec locked in with the pulses impressed upon the first binary counter stage.

(c) Two diode matrix gates are used to sample the states of the four binary-counter stages. These matrix gates are designed to be inoperable except when the binary-counter outputs to which they are connected are all in the ON condition, at approximately −10 volts. The delayed-pulse train from the second amplifier (M1733-1) stage is impressed upon the two matrix gates. One of the matrix gates is connected to all of the *A* leads to the binary counters, and thus will be enabled to pass the delayed pulse when the binary counters are all in the 0 condition. The other is connected to all of the *B* leads, and will be enabled to pass the delayed pulse when all of the binary counters are in the 1 condition. The output of the first gate is then passed to the setting input lead of the control function to cause the control to activate the gated amplifiers causing addition, and the output of the second gate is passed to the setting input lead of the control function to cause it to activate the gated amplifiers causing subtraction. At all other conditions of the binary-counter stages, both of the matrix gates are disabled.

By this means, the add-subtract control function is automatically controlled by the previous condition of the binary-counter stages. If the counter has been adding and attains the condition 1, 1, 1, 1, the control function is caused to change state so that the next triggering pulse will find the binary-counter stage interconnections arranged for subtraction, and vice versa.

(d) In order to permit an oscillographic display of the action of the binary-counter stages, a simple diode registration means is employed. A branch-load circuit is connected to all of the *B* registry leads of the binary counter stages to provide a weighted current increment to be drawn from each stage when it is ON. These currents are passed through a milliameter for static display purposes and an oscilloscope may be connected across the milliameter terminals for visual display.

[2] A. E. Anderson, "Transistors in switching circuits," Proc. I.R.E., vol. 40, fig. 7(c), p. 1545; this issue.

[3] J. R. Harris, "A transistor shift register and serial adder," Proc. I.R.E., vol. 40, fig. 6, p. 1600; this issue.

A typical output waveform, existing at a collector of a binary-counter stage operating in a reversible binary counter incorporating the above features, is shown in Fig. 4. The loading effects of the matrix gate, branch diode registration circuit, and gated amplifier are seen to affect the initial rise and fall characteristics. (This waveform may be compared with Fig. 17.) However, as long as the delay introduced before the application of the pulses to the matrix gates is sufficient to insure that the binary-counter stages will have stabilized, no recognition ambiguity will result.
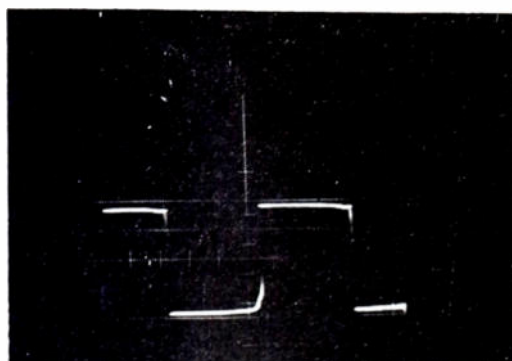
Fig. 4—Loaded binary-counter stage output waveform. Time scale, 100 μsec/division; voltage scale, 5 volts/division.

The analogue output current waveform provided by the branch diode registration circuits is illustrated in Fig. 5. The 16 discrete information bit levels are clearly shown. Operation of the control function at maximum and minimum levels to provide successive addition and subtraction is illustrated.
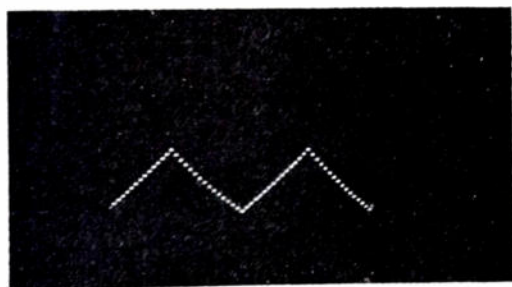
Fig. 5—RBC—binary to analogue output current waveform.

Figs. 6 and 7 show the analogue output current waveforms which result when the matrix gates enabling the passage of resetting pulses to the control function are removed, in turn. In Fig. 6, the matrix gate enabling the passage of the pulse controlling subtraction has been removed from the circuit. Accordingly, the control function remains in the "ADD" enabling condition, and the binary-counter stages continue to register additional current increments with the impression of triggering pulses until the limit of storage capacity (16 bits) is attained. The next triggering input pulse causes all the stages to revert simultaneously to the 0 condition, and then addition reoccurs. The waveshape illustrates this registration of additional increments until

the maximum is attained, then the reversion to 0 condition, followed by addition. In Fig. 7, the matrix gate enabling the passage of the pulse controlling addition has been removed from the circuit. The control function remains in the "Subtract" enabling condition, and

Fig. 6—RBC—binary to analogue output current waveform subtract control disabled.

Fig. 7—RBC—binary to analogue output current waveform add control disabled.

the binary-counter stages continue to register subtractional current increments with the impression of triggering pulses. This process continues until the minimum storage capacity is attained, and the next impressed triggering pulse causes all the stages to revert simultaneously to the 1 condition.
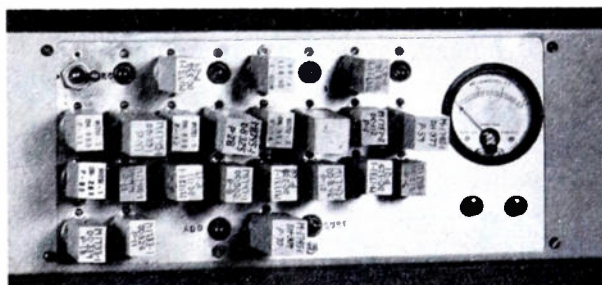
Fig. 8—Packaged transistor reversible binary counter.

A photograph of the packaged transistor reversible binary counter incorporating the features discussed is shown in Fig. 8. Approximately 8 watts of input power are required by the transistor stages used.

Conclusions

The switching function of reversible binary counting may be performed using two elementary transistor packaged circuits. The use of a binary counter and a regenerative gated amplifier for this particular operation has been illustrated and experimentally verified. Margins of operation are satisfactory for laboratory type experimentation and satisfactory performance for extended periods of operation has been obtained. For

application to system usage, the margins of operation may be improved as discussed in Appendix C to achieve greater stability and consistent operation over wide variations in power supply voltages and ambient temperatures.

Additional features which may be generally useful as switching functions, such as delay, recognition of a discrete binary code, and analogue transformation of the binary code, have been applied successfully to a packaged transistor reversible binary counter.

Acknowledgment

The above report comprises a portion of the development work performed under contractual arrangements with the Joint Services.[4] Acknowledgment is also given Messrs. A. E. Anderson, J. R. Harris, and L. W. Hussey for their contributions in conceptual design and analysis.

APPENDIX A

The purpose of this appendix is to describe the circuit and characteristics of a transistor gated amplifier which may be used as a threshold-two gate. Such a gate requires the simultaneous presence of two signals to produce an output, and since the gate employs an active device which is triggered, the output pulse is not a replica of either input, but is rather a discrete pulse.

Past practice in threshold-two gating is exemplified by diode configurations, involving transmission loss and having impedance-matching requirements at input and output.

The design considerations of the gating circuit to be described are based upon the active properties of the point-contact type transistor. The gating properties are achieved by basing the threshold-two feature upon the arithmetic summation of two signal amplitudes. One of these signals is in effect a dc signal and the other a positive pulse. The negative-resistance characteristics of the transistor are used to obtain triggered regenerative amplification. This method has the disadvantage of being sensitive to amplitude variations in the two signals employed, and it is necessary to insure sufficient margins of operation under the most adverse conditions of signal-amplitude variations to be expected.

Basic Circuit Configuration

The elementary regenerative gated-amplifier circuit and a typical input (emitter) negative characteristic are shown in Fig. 9. The circuit incorporates a biased diode in the base circuit to clamp the peak turning point[5] of the characteristic close to a definite potential, in this case zero volts. This feature has been used to offset variations between transistors and between the operation of a given transistor at different temperatures. The changes have been caused by variations in the amount

of collector current which flows when the emitter electrodes are biased negatively or are open-circuited, with constant collector voltage. These variable collector currents, when flowing through the relatively large external base circuit resistor, cause wide variations in the peak turning point in the characteristic. These effects have been treated by Anderson.
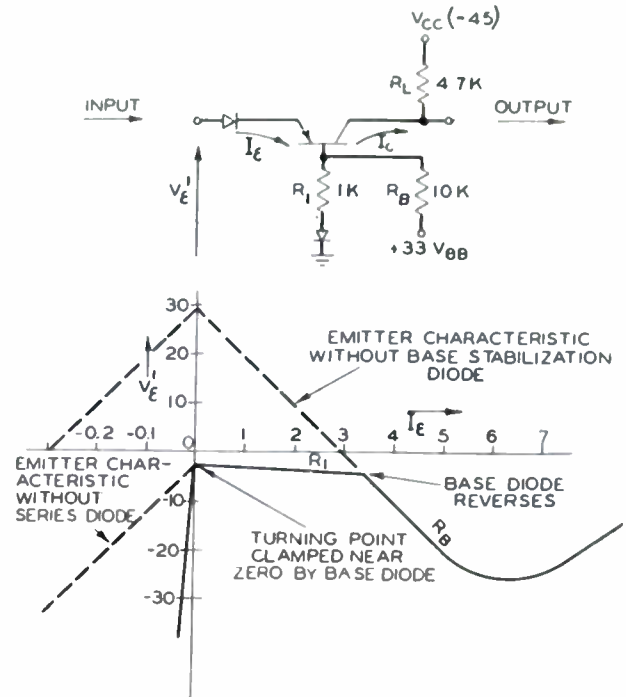
Fig. 9—Stabilized input characteristic.

The biased diode provides a low impedance which shunts the large external base circuit resistor at negative- and low positive-emitter currents, preventing the variable collector currents from affecting the peak turning point. At higher positive-emitter currents, because of the current-gain properties of the transistor, this diode assumes a high-impedance condition. The resultant high-base feedback then results in the development of the normal negative-resistance characteristic.

A small resistance has been added in series with the diode to insure obtaining a negative resistance as soon as the emitter current becomes slightly positive. This is to obtain ease of triggering at a small cost in stability of the peak turning point with transistor-parameter variations.

A diode is inserted in series with the emitter to insure a slope in the negative-emitter-current region greater than 0.5 megohm for additional discrimination. The circuit constants as illustrated have been so adjusted that variations[5,6] in the r_{c0} of the transistor from approximately 70,000 ohms to 12,000 ohms may be tolerated. By this means, initial variations occurring between

[4] "Engineering Services on Transistors," Signal Corps Contract No. W36-039-sc-44497, Dept. of the Army Project No. -3-19-03-031, Signal Corps Project No. -27-323A-1.

[5] A. E. Anderson, *op. cit.*, fig. 17, p. 1553.
[6] The r_{c0} of a point-contact transistor may be defined as the equivalent-collector dc resistance obtained by impressing a constant collector voltage V_c (i.e. -35 volts) with open-circuited emitter ($I_e = 0$) and measuring the collector current.

units, and variations with temperature do not materially affect the operation of the circuit.

With this basic regenerative-amplifier circuit, the function of threshold-two gating, using a step voltage and coincident positive pulse, may be performed as illustrated in Fig. 10. The emitter load line corresponding to resistance R_ϵ is superimposed upon the input characteristic of the regenerative amplifier resulting in dc stable equilibrium points at either A or B in the negative-emitter-current region. The point of intersection (A or B) is determined primarily, for the choice of
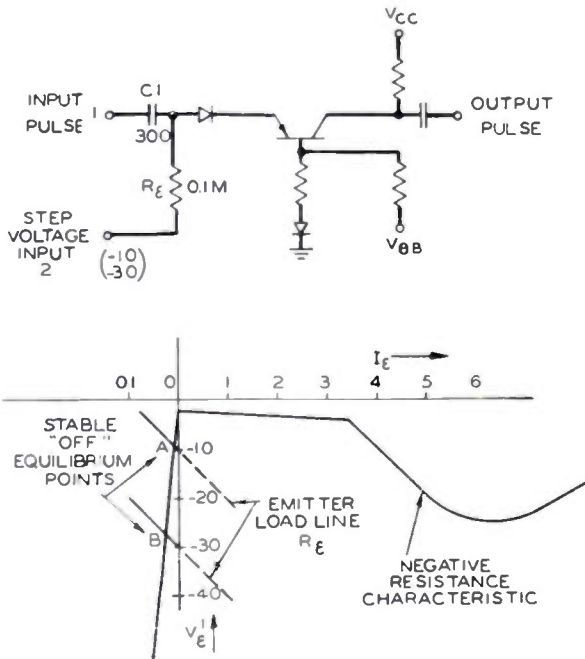
Fig. 10—Gated-amplifier circuit and input characteristic.

components depicted, by the value of the step voltage applied to Input 2.

We may assume that the pulse to be gated has an amplitude of 15 ± 5 volts. If the step-voltage input is at the -10 volt level, and the 15 volt pulse impressed upon Input 1, the circuit will be triggered and a positive pulse of the order of 30 volts, 1 to 7 μsec long, will appear at the output. If the step-voltage input is at the -30 volt level, however, the emitter load line will intersect the input characteristic at point B, and the input pulse will be incapable of triggering the circuit.

The margins of operation depend primarily upon the amplitude variations of the input pulse and step-voltage input, and the slopes of the characteristic and emitter-load resistance.

The step-voltage input biasing potentials of -10 and -30 volts used in the discussion are for illustration purposes only, but represent the potentials present at the collector of a transistor bistable flip-flop when in ON and OFF conditions, respectively.

A typical output pulse waveform at the collector of a transistor gated amplifier is shown on Fig. 11. This waveform was taken on a packaged gated-amplifier

stage used in a laboratory model of a reversible binary counter. The step-voltage input, derived from a collector of a binary-counter stage used as a control, was measured as being approximately -11.0 volts. The triggering pulses, of about 20-volts amplitude, were derived from a preceding binary-counter stage and were impressed at a rate of 1,500 pps. The output pulse of the gated amplifier was passed to the following binary-counter stage, and therefore represents a typical output waveform under operational load conditions.
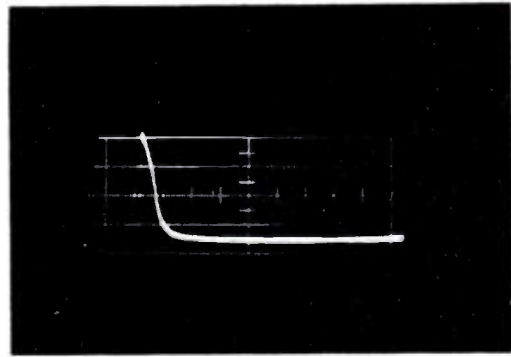
Fig. 11—Typical output pulse of gated amplifier; time scale: 10 μsec/division, voltage scale: 5 volts/division.

The output pulse has an amplitude of $+22$ volts, duration (to 3 db down point) of 6 μsec and rise time of about 0.1 μsec.

A photograph of a packaged transistor gated amplifier is shown on Fig. 12. The package is designed to plug into a standard 7-pin miniature socket, and occupies a
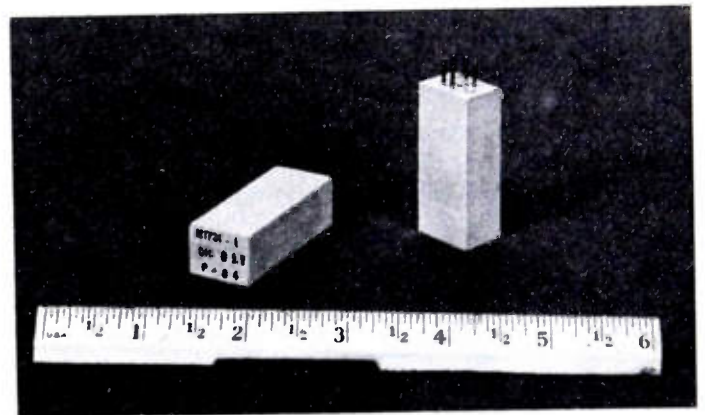
Fig. 12—Packaged transistor gated amplifier.

volume of $\frac{1}{2}$ inch by $\frac{3}{4}$ inch by 2 inches, including the base pins. The circuit elements, comprising standard resistors, diodes, and transistors are mounted on a circuit board using the "auto-sembly" techniques evolved by the Signal Corps. The entire circuit-component assembly is then cast in a mold, using a thermo-setting plastic.

The power dissipated by the circuit components is of the order of 0.1 watt, engendering an internal rise in temperature of approximately 7 degrees C above ambient, under usual operating conditions.

Summarizing, the characteristics of this gated regenerative amplifier are as follows.

(a) The source of step-voltage biasing potential always sees a high impedance of at least the emitter-load resistance R_t; for the illustrated plot, 0.1 megohm. During the interval when no triggering pulse is present at the pulse input terminal, the source sees at least 0.7 megohm. This permits a multiplicity of these gated amplifiers to be operated in parallel by the same controlling potential source.

(b) The triggering-pulse source also sees a high impedance except during a triggering interval at which time the source sees approximately 3,000 to 5,000 ohms, depending upon the gate-circuit constants and external-load impedance.

(c) The output pulse obtained upon triggering is essentially coincident in time sequence with the leading edge of the pulse impressed at the input terminal. The size of coupling capacitor and pulse-generator impedance affects the rise and fall time and duration of the output pulse. Generator impedances less than 10,000 ohms and coupling capacitors larger than 100 to 200 mmfd generally result in rise times less than 0.3 and durations in the order of 1 to 7 μsec. Amplitudes of output pulses in the order of 30 to 40 volts are usual.

(d) The gate has the disadvantages of being sensitive to variations in the amplitude of the gating signals. The margins of operation of this circuit are discussed in Appendix C.

(e) The clamping feature used to obtain constancy of the peak turning point at zero volts results in a power dissipation in the order of 0.1 watt in the clamp circuit during the OFF condition. It also results in a decrease in triggering speed since only a nominal value of negative resistance is attained at the offset of triggering action.

APPENDIX B

The purpose of this appendix is to describe briefly the operation of a general-purpose binary counter[7] employing two transistors and to discuss its characteristics and performance.

CIRCUIT DESCRIPTION

The binary counter consists of two symmetrical and cross-coupled point-contact type transistor stages somewhat analogous in appearance to the Eccles-Jordan electron-tube circuit. The principles involved are, however, different. It is possible to make a bistable circuit with a single point-contact type transistor, since the device has a large signal current gain or alpha greater than unity, giving rise to an n-type negative-resistance characteristic.[5]

However, the symmetrical arrangement has the advantages of providing the two outputs necessary to achieve reversibility and uniform power-supply drain independent of the states of the circuit.

[7] R. L. Trent, "Two transistor binary counter," *Electronics*, vol. 25, pp. 100–101; July, 1952.

Since each of the two transistors can be individually stable at both high- and low-current states, it is possible to have four stable states in a two-transistor circuit of this type. In the binary counter of discussion, the only useful combinations of these four states are the two in which one unit is in its high-current state and the other is in its low-current state. The other two conditions where the devices are simultaneously in either high- or low-current states must be avoided. The circuit design so constrains the behavior that these spurious states cannot exist.

Fig. 13 is a schematic of a complete binary-counter stage. DC coupling between the balanced stages is afforded by the common emitter-load resistor ($R12$) and ac coupling between emitters by capacitor ($C3$). In addition, there are cross-couplings between collectors and bases consisting of resistor-capacitor networks ($C1$
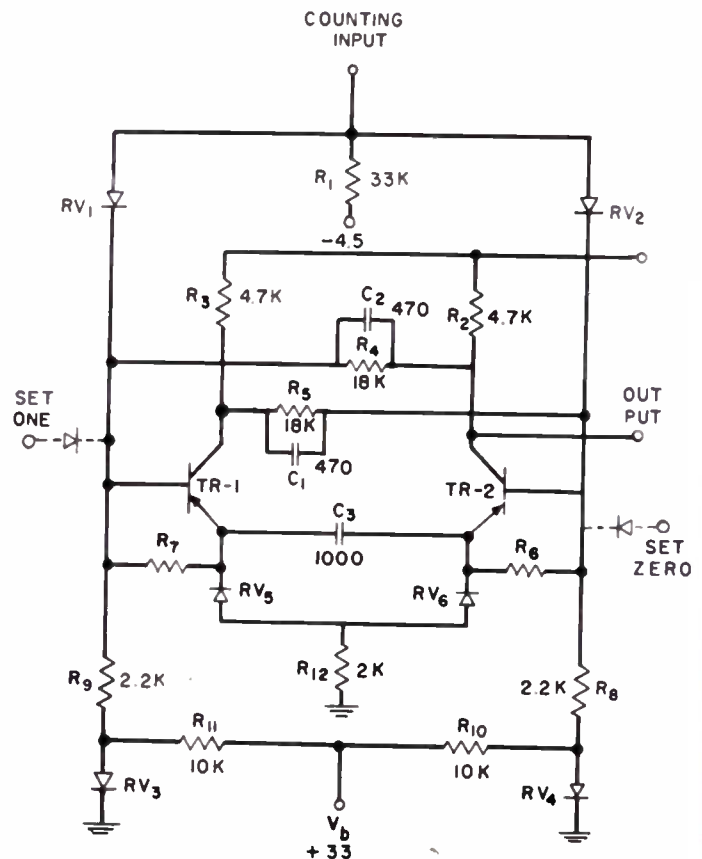
Fig. 13—Stabilized binary-counter circuit.

and $R5$), ($C2$ and $R4$). These couplings provide regenerative paths for the ac signals and dc paths to maintain bistable operation once the transient effects have subsided.

Triggering pulses may be impressed upon the bases under the control of a diode steering circuit, which is given direction by the difference of potential existing between the bases when one unit is ON and the other is OFF. The common emitter resistance ($R12$) eliminates the necessity for emitter biasing supplies. At the two stable equilibrium points it is always passing the emitter current for the ON stage which provides the

biasing potential for the OFF stage. The two transistors are provided with a base-stablization feature, as discussed in Appendix A, consisting of the addition of a biased diode in the base lead. The diode provides a low shunting impedance in the negative-emitter-current region and at low positive-emitter currents, and a high impedance at large positive-emitter currents. In its low-impedance condition, the diode prevents variations in the collector current from causing a wide range of voltage to be developed across the external-base resistance. In its high-impedance condition, the diode allows the remainder of the usual negative-resistance characteristic to be obtained over a range of positive-emitter currents.

Description of the various circuit features to insure low triggering requirements may be facilitated by discussing each in terms of a skeleton circuit.

Triggering Sensitivity Feature

Satisfaction of the bistable criterion that a counter stage must maintain a given condition indefinitely until the next triggering pulse is impressed has usually resulted in poor triggering sensitivity. Circuital means are employed in the circuit under discussion to secure both stability and sensitivity. Fig. 14 illustrates the dc lockup performed by the common emitter resistor ($R1$) and the cross-coupling resistors shown as ($R2$). The
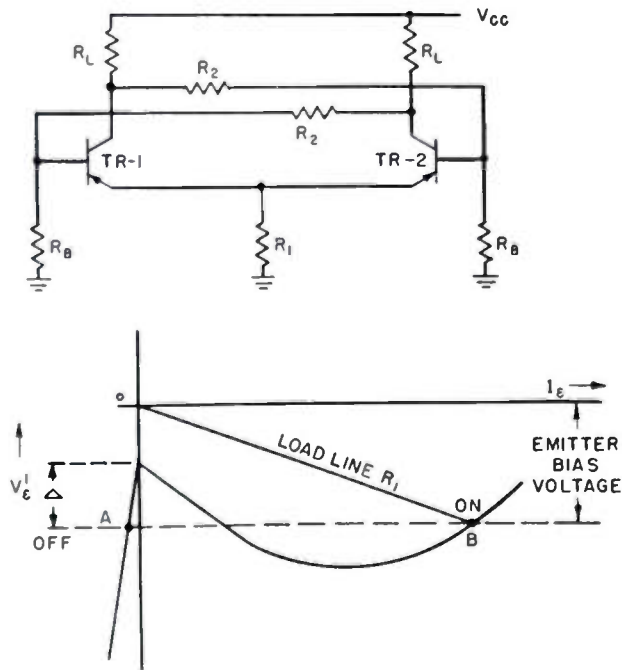
Fig. 14—Trigger circuit lockup feature and characteristic.

simplified characteristics illustrated assumes both transistors to have identical characteristics and that each transistor acts as a two-terminal device. These assumptions are not necessarily true, but will serve to illustrate the point under discussion. The ON unit is assumed stable at B, the intersection of the load line with the characteristic. Because of the coupling action of ($R1$), the OFF unit must be stable at point A at the same emitter

potential. An amplitude of positive triggering pulse greater than Δ would be required for triggering the OFF unit toward the ON region.

Fig. 15 illustrates the effect on the load line and negative-resistance characteristic of adding varistor ($RV1$) and shunt resistor ($R3$). The ON unit has the same
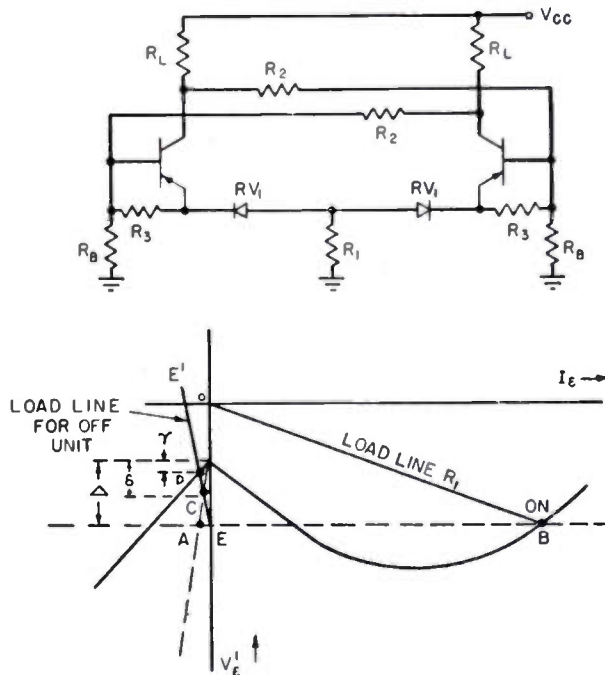
Fig. 15—Trigger circuit for increased sensitivity and characteriistc

operating point B as before. Instead of the OFF unit being stable at point A, however, the diode in series with its emitter is biased in the reverse direction, resulting in the new load line $E-E'$ for the OFF unit. This load line $E-E'$ would intersect the normal characteristic at the point C were it not for the shunting effect of resistor ($R3$). An input pulse of amplitude greater than δ would then be sufficient for triggering action. The negative-resistance characteristic has been modified by shunt resistor ($R3$) to decrease the impedance in the negative-emitter-current region. Load line $E-E'$ therefore intersects the modified characteristic at point D instead of at point C. Triggering action now occurs if the input pulse is greater than γ. While the sensitivity has been increased materially, no concession has been made to the stability of the states of the circuit.

Composite Characteristic of Counter Stage

An approximate composite emitter characteristic is shown on Fig. 16. The base stabilization assures that the peak turning point of the characteristic will be very close to 0 volts. Since the stabilizing current is larger than that required to balance the OFF collector current of the transistor, there will be a region where for positive-emitter currents the negative resistance is essentially determined by the series base resistance and the forward resistance of the base diode ($R8$, RV_4 and $R9$, RV_2 of Fig. 13). These are approximately 2,500

ohms. When the base diode switches to its high-resistance state, the negative resistance is determined by the total series base resistance (R8 and R10, R9 and R11, Fig. 13), approximately 12,200 ohms, neglecting the shunting effect of the base diode reverse resistance.

If the assumption is made that the transistor TR-2 is in the ON condition, the diode in series with the emitter (RV6) will be in its low forward resistance
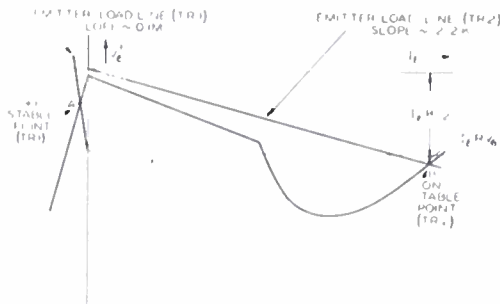
Fig. 16 —Composite input characteristic.

state, and the TR-2 emitter load line is then (R12 +RV6) about 2,200 ohms, intersecting the origin. The stable ON point for TR-2 is the intersection of this load line with the characteristic at B. Since transistor TR-1 is in the OFF condition, its series diode (RV5) will be in its high-resistance state. This determines the slope of the emitter load line for TR-1, and its biasing potential is set by the positive-emitter current of TR-2 passing through (R12). The load line drawn through the biasing potential intersects the characteristic at point A, the OFF stable point for TR-1, very close to the peak turning point of the negative-resistance characteristic.

Mechanism of Triggering

The binary-counter circuit is triggerable with either positive or negative pulses, provided the proper steering diode polarity is observed. With the circuit as shown in Fig. 13, the steering may be accomplished by the difference of potential existing between ON and OFF unit base potentials. The base of an ON unit is always 5 to 10 volts more negative than the base of an OFF unit. Therefore, a positive pulse appearing at the count input terminal will be passed to the base of the ON unit. Because of the low impedance from base to emitter in the ON condition, this pulse is passed by the coupling capacitor (C3) to the emitter of the OFF unit, which is at high impedance. Since the OFF unit is stable at a point very close to the peak turning point of its characteristic, it will be triggered towards its ON stable point. This triggering action results in a very sharp positive pulse of about 20-volts amplitude appearing at the collector of TR-1. This positiveward change is passed by means of the cross-coupling capacitor (C1) to the base of the ON unit. The characteristic of the ON unit is thereby shifted in a positive direction. If the characteristic is shifted sufficiently so that the emitter load line (TR2) in Fig. 16 no longer intersects the characteristic,

the only stable point remaining is in the OFF stable region. Thus the collector voltage becomes more negative, following the time constant of the cross-coupling capacitor (C2) and its effective discharging resistance. It is this time constant which places an upper limit upon the repetition rate of the counter circuit as illustrated at approximately 40 kc.
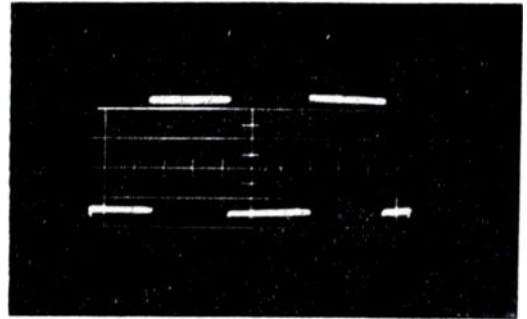
Fig. 17—Typical collector output waveform; rate: 470 pps, time scale; 700 μsec/division, voltage scale; 5 volts/division.

CHARACTERISTICS

A typical output pulse waveform at a collector or output terminal of a binary-counter stage is shown on Fig. 17. This waveform was taken on a packaged binary-counter stage operated unloaded with external circuitry. The triggering pulse, derived from a pulse generator, was 1 μsec long, 6 volts in amplitude, with a repetition rate of approximately 470 cps. The square wave output is about 20 volts in amplitude.
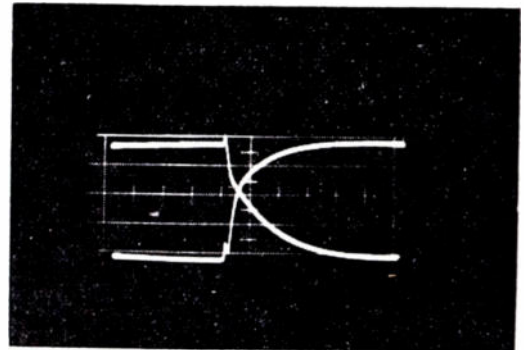
Fig. 18—Typical collector output waveform. Transition detail, rate: 470 pps, time scale; 1 μsec/division, voltage scale; 5 volts division.

The detailed waveforms during transition from OFF to ON and from ON to OFF are shown in Fig. 18, taken under the same operating conditions as discussed above. The time scale is enlarged to 1 μsec/division. The small positive pip on each curve represents the transmission of the leading edge of the triggering pulse. The major portion of the translation from OFF to ON occurs within 1 to 2 μsec after the impression of the triggering pulse, whereas switching from ON to OFF takes a somewhat longer period of from 3 to 4 μsec, as discussed previously.

A photograph of a packaged transistor binary-counter stage is shown on Fig. 19. The package is designed to

plug into a standard 7-pin miniature socket, and occupies a volume of $\frac{3}{4}$ inch by $\frac{3}{4}$ inch by $2\frac{1}{2}$ inches, including the base pins. The same techniques of "auto-sembly"
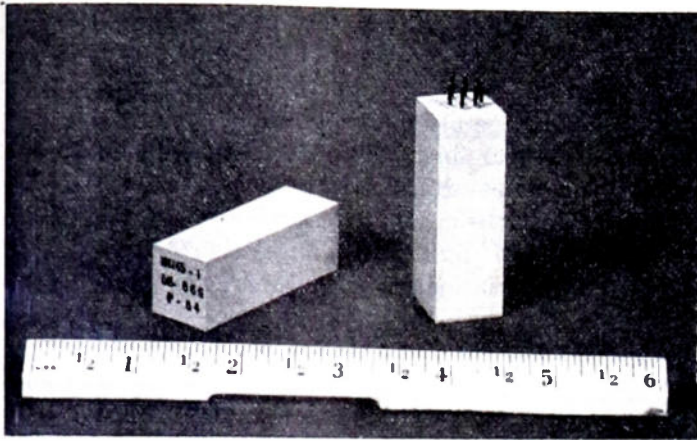
Fig. 19—Packaged transistor binary-counter stage.

and plastic casting were employed as described for the gated amplifier. The power dissipated by the circuit components within the package is of the order of 0.5 watts, engendering an internal rise in temperature of approximately 12 degrees C above ambient, under usual operating conditions.

Fig. 20—Operating margins.

APPENDIX C

The purpose of this appendix is to describe and illustrate the margins of operation of the two-transistor binary-counter stages and the transistor gated amplifier when the latter is used as a gated interstage between two stages of binary counters. The operating conditions for the transistor gated amplifier are depicted in Fig. 20, in which the negative-emitter-current scale is en-

larged for illustrative purposes. If we assume that the transistors used in the binary-counter stages and the control function will have r_{c0} values between 40,000 and 12,000 ohms under the worst temperature conditions to be experienced, this assumption, together with a knowledge of the circuit constants, enables the variations in ON and OFF collector voltages to be calculated. This in turn gives a direct indication of the step-voltage output of the control which acts as the sensitizing-inhibiting means for the gated amplifier, and also of the voltage amplitude of the positive pulses from the binary-counter stages.

Fig. 21—OFF collector voltage calculations.

CALCULATION OF COLLECTOR VOLTAGE VARIATIONS

OFF Condition

The calculation of OFF condition collector voltages may be made as shown in Fig. 21. The following simplifying assumptions have been made.

1. The diodes $(RV1)$ and $(RV2)$ are perfect switches, $(RV1)$ open, $(RV2)$ closed.
2. The ON condition base voltage is constant at -10 volts.
3. r_{c0} variations are between limits of 40,000 and 12,000 ohms.

The computations show that the OFF condition voltage varies between -30 and -34.8 volts. Experimental verification of these calculated values have been made with excellent agreement. We may therefore assume that the collector voltage during the OFF condition will be 32.5 ± 2.5 volts.

ON Condition

The calculation of the ON condition collector voltage variations may be made fairly accurately because the circuit constants are much larger than the transistor parameters. The ON currents and voltages are therefore determined essentially by the circuit parameters. The tolerances on the resistive elements must be considered in these calculations. Experimental verification of the values obtained by computation agree within approximately 0.5 volt, and the variations are between -11.2 and -15.6 volts, or -13.4 ± 2.2 volts.

Output Pulse Amplitude

The maximum and minimum positive-pulse amplitudes obtainable at a collector may be calculated, using the above results.

Maximum: $+23.8$ volts
Minimum: $+14.4$ volts

MARGINS OF OPERATION

Requirements

The conditions which must be fulfilled to prevent ambiguity and consequent failure of the reversibility feature are as follows:

1. The minimum positive output-pulse voltage amplitude obtainable from a collector of a bindary-counter stage (calculated as $+14.4$ volts) must be capable of triggering the threshold-two regenerative gate under all conditions of biasing set up by the control-step voltage when transmission is desired.

2. The maximum positive output pulse from a binary-counter stage (calculated as $+23.8$ volts) must not be capable of triggering the threshold-two regenerative gate under all conditions of biasing set up by the control-step voltage when it is desired to inhibit the gate.

Operational Conditions

Fig. 20 illustrates the range of variation between the ON and OFF condition biasing voltages on the voltage axis. These represent the step-voltage outputs from a binary-counter stage used as a control function, and impressed upon the step-voltage input of a regenerative gated amplifier. When the ON condition voltage

$(-13.4 \pm 2.2$ volts) is impressed, the amplifier should be triggerable by the positive pulse output from a preceding binary-counter stage. When the OFF condition voltage $(-32.5 \pm 2.5$ volts) is impressed, the gated amplifier should be inhibited, and any value of positive pulse output from a preceding binary-counter stage should be incapable of triggering the gated amplifier. From the voltage intercept points on the vertical axis, the emitter load lines may be plotted. The intersections of load lines covering the range of variation in each condition with the negative-resistance characteristic are designated $A-A'$ and $B-B'$. These represent the OFF stable-point variations of the regenerative gated amplifier, and represent biased conditions of -11.5 ± 2.0 volts for the enabled condition and -27.5 ± 2.5 volts for the inhibited condition. Comparison of these biasing conditions with the minimum and maximum pulse amplitudes from a binary counter show that operational margins of about 1 volt under the most adverse conditions to be expected are obtained.

DISCUSSION OF RESULTS

This appendix has illustrated the use of a binary-counter output voltage as the controlling step-voltage source used to enable or inhibit a regenerative gated amplifier. The margins of operation of a gated amplifier when controlled by such a step-voltage source have been shown to be suitable for experimental work. For system usage, however, improvements in the margins of operation would be desirable, and may be obtained if required by several expedients.

(a) Wider margins of operation in meeting condition (1) of the requirements would be obtained if a step-voltage source having somewhat lower ON condition voltages and greatly increased OFF condition voltages were employed. A redesigned bistable circuit using different supply voltages and circuit elements could be designed for this specific job.

(b) Wider margins of operation in meeting condition (2) of the requirements also may be achieved as above. In addition, insertion of a simple limiting network to clip the output-pulse amplitude from a preceding binary-counter stage at some level, for example $+20$ volts, would result in substantial improvement.

An Optical Position Encoder and Digit Register*

H. G. FOLLINGSTAD†, ASSOCIATE, IRE, J. N. SHIVE†, AND R. E. YAEGER†

Summary—The usefulness of transistors in systems has been given a feasibility proof through the construction and operation of a six-digit position encoder and serial-output digit register. This system performs the functions of photoelectric encoding, pulse regeneration, digit storage, reflected-to-natural binary translation, and digit shifting by means of circuits using transistors and other semiconductor devices. The model occupies a volume of about ¼ cubic foot, weighs seven pounds, and consumes 16 watts of power.

I. INTRODUCTION

THE PROBLEM DESCRIBED in this paper was undertaken to demonstrate the ability of circuits using transistors and other semiconductor devices to perform completely all of the functions ordinarily required of vacuum-tube circuits in a data-handling system of some complexity. As a vehicle for such a demonstration a system was chosen in which analog-information is presented in the form of the linear displacement of a shaft along its axis. It is desired that this shaft position be encoded into a binary-digit number and delivered to an output as a time sequence of pulses representing this number. The entire cycle of operations is designed to repeat approximately 10 times per second. Fig. 1 shows a functional diagram of such a system.
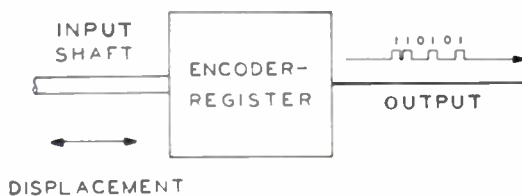
Fig. 1—Functional diagram of the encoder-register.

A. Specific Requirements on the System

The linear-encoding accuracy of this system was chosen so as to be favorably competitive with the accuracy of other encoding methods, such as those employing potentiometers or multidigit commutators and pickup brushes. A desired accuracy of ±0.0005 inch was thus chosen, corresponding to an encoding quantum distance of 0.001 inch. With a fixed quantum distance the maximum linear displacement which the system can encode is determined by the number of binary digits it can handle. For example, a 0.001 inch, 10-digit encoder can encode displacements up to $0.001 \times 2^{10} = 1.024$ inches. These developments were looking toward the feasibility proof of an eventual 32-inch encoder, corresponding to a 15-digit system.

The shaft is assumed driven by a servo system or delicate gear train of such a nature that frictional drag must be kept to an absolute minimum. This assumption rules out the possibility, for example, of using the shaft to

drive a pickup brush along a potentiometer. Accordingly, a no-drag optical method of encoding is used.

The 10-cps repetition rate is not related to the speed with which the shaft is expected to move. The motion may, in fact, be many quantum distances between successive encodings. It is merely desired to acquire new position information every tenth of a second. It is, however, necessary that the complete act of encoding take place within the time required for a displacement over a quantum distance at the highest speed the shaft will move. In the present system a maximum shaft speed of about 8 inches per second is assumed. Consequently, the encoding act may not consume more than 125 μsec.

The time interval during which a single-digit pulse may be present at the output is related to the repetition rate and the number of digits in a word. Provision for a 15-digit word every tenth of a second requires 150-digit time slots per second. However, because it was envisioned that transistor-data circuits would sooner or later be called upon to handle information at much faster rates, it was decided to test their potentialities in this direction, and provision was made in this system for 600 time slots per second. The duration of each time slot is accordingly 1.67 msec.

Other features of the system, such as the details of the circuits involved, power and impedance levels, and the ways and means of accomplishing the required functions, were left to the designers' choice.

B. Scope of the Present Objective

While the tentative requirements have been enumerated for a system which was in mind at the time this work was undertaken, the immediate objective of the present endeavor was not actually to build and operate such a system, but rather to *demonstrate the feasibility* of such a system, using transistors and semiconductor devices. It was decided that all of the systems, circuit, and device problems occurring in the complete system could be given a satisfactory feasibility proof by devising, building, and demonstrating a 6-digit model. This model was to be identical with the 15-digit encoder in every respect except digit capacity.

II. FUNCTIONAL ANALYSIS OF THE SYSTEM

The accomplishment of the system objective of Fig. 1 requires the performance of a number of related functions. The first of these is the encoding of the shaft position into a binary-digit number. Fig. 2 shows schematically how this is done. Attached to the shaft and moving with it is a glass plate having a 6-digit binary-code pattern of apertures in an opaque coating on one side. This pattern is illuminated by a flashing discharge lamp whose 25-μsec flashes are programmed to

correspond with the 10-cps instants at which the position of the code plate is to be encoded. The code is read by a line of six photocells, one for each digit row on the code plate. These photocells are placed behind a reading slit in a stationary plate, which clears the code surface by a small distance just sufficient to insure that the code plate moves completely free. The purpose of the reading slit is to define the resolution of the optical system and to insure that the photocells do not straddle several of the quantum distances on the code plate.
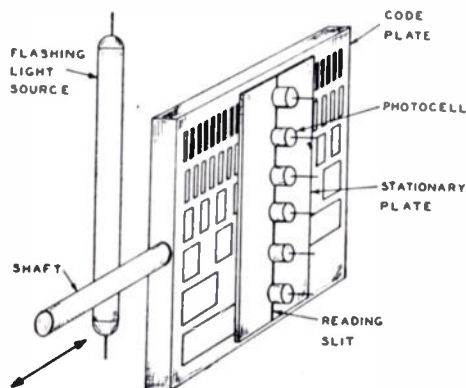
Fig. 2—Code plate and photocell encoder.

At each flash of the lamp the photocells simultaneously give output indications of their respective digits: an illuminated photocell puts out a pulse, corresponding to a 1, while a photocell which is in darkness behind an opaque portion of its digit code pattern at the instant of the flash puts out nothing, corresponding to a 0 in a binary code. The resulting 6-digit presentation of 1's and 0's at the photocell outputs then comprises the binary number defining uniquely the position of the code plate at that instant.

In order to secure the desired accuracy of encoding to a quantum distance of 0.001 inch, the width of the stationary reading slit must be no greater than 0.0005 inch. The small amount of light passing through this slit to the photocells produces photocell-output pulses of limited amplitude and power. This necessitates the insertion of a triggered pulse regenerator following the photocell in each digit channel. This pulse regenerator consists of a transistor monostable-multivibrator circuit capable of being triggered by the output pulse of a flash-illuminated photocell. At each such triggering the pulse regenerator puts out a single pulse whose characteristics depend upon the circuit and not upon the nature of the trigger. The pulse generator performs another important function, that of discrimination. If a photocell is reading the edge of a code aperture which is partway between a zero position and a one position, the photocell puts out a pulse having less than the full amplitude corresponding to an unobstructed aperture. As the code plate moves between two such positions, a gradual change occurs in the amplitude of the output pulses of the photocell. It is necessary that the system be able to discriminate between pulses which have an amplitude greater than that corresponding to the exact quantum transition point and those whose amplitude is

less. The transition ambiguity is minimized to a small fraction of a quantum step by the trigger threshold of the pulse regenerator which is set so that the circuit triggers only on pulses exceeding a reference amplitude corresponding to the exact transition point and remains unresponsive to those below it.

Following the pulse regenerators is a register in which the number is stored until it is to be delivered at the output. This register must have the property of accepting simultaneously all six digits of the number to be stored, and subsequently of shifting these digits, in pulse form and in the proper time sequence, to the output. The register consists of a number of transistor flip-flop circuits, each capable of being set-to-1 by an output pulse of the preceding pulse regenerator. The register stages are interconnected by other circuits so that the function of progressive shift-out can be performed.

The provision of an ordinary binary-code pattern on the code plate could lead to serious errors of reading. Examination of the four-digit code plate example in the left-hand drawing of Fig. 3 shows that there are a number of abscissae, such as *A-A*, at which several of the digits change from 1 to 0 or from 0 to 1 as the code plate advances from one quantum position to the next. Should the reading slit straddle such an abscissa at the instant of a flash of light, then the photocells corresponding to those digits which are in transition at this

NATURAL BINARY   REFLECTED BINARY
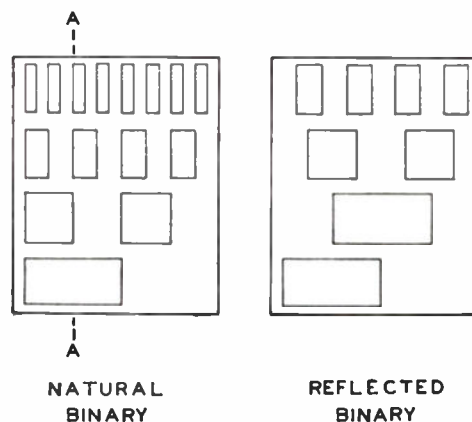
Fig. 3—Left, 4-digit natural-binary code pattern; right, 4-digit reflected-binary code pattern.

abscissa may read either 1 or 0, depending on the exact adjustment of the discrimination of each photocell and regenerator circuit. In such a case the output pulses may yield a number having little relationship to the position of the code plate at the instant of the flash. In order to reduce the potential errors of such transitional readings, the code plate is provided with a special "reflected binary" code pattern.[1] This scheme is illustrated in the right-hand drawing of Fig. 3. The reflected binary-code pattern has the property that at any transition between any quantum position of the code plate and the next, only one of the digits changes from 1 to 0 or from 0 to 1. This means that only one photo-

[1] A more thorough discussion of the properties of the reflected binary code is given by W. M. Goodall, *Bell Sys. Tech. Jour.*, vol. 30, no. 1, pp. 33; 1951.

cell of the reading array may misread and that only one digit of the resulting number may be uncertain. Furthermore, regardless of whether the uncertain digit is the
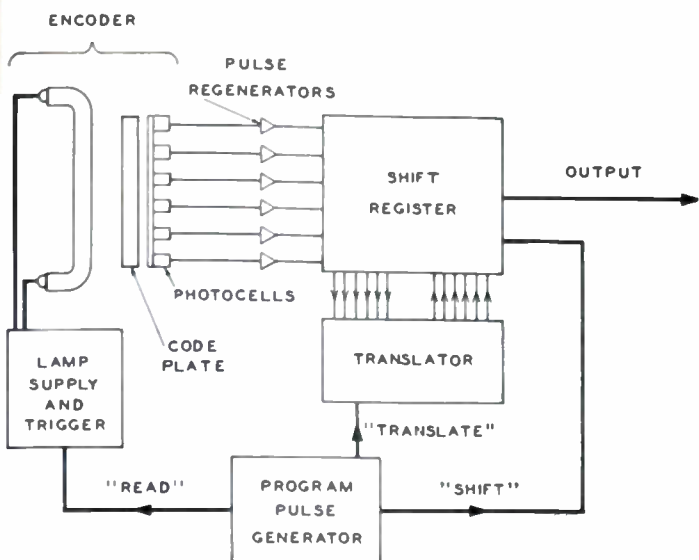
Fig. 4—Breakdown of the system into functional units.

least significant, the most significant, or any digit in between, the greatest position uncertainty will be plus or minus one half quantum distance.

The use of a reflected-binary encoder presents one difficulty to the prospective user of the system. If the encoder-output information is to be fed, for example, into a conventional binary computer, then the number must be translated back from reflected binary to natural binary, or else the computer must be capable of handling reflected-binary numbers. Although reflected-binary computing circuits can be devised, such circuits require an auxiliary function similar to the translation process. Accordingly, a translation function must be included somewhere in such a tandem arrangement, and it was chosen, in the undertaking here reported, to provide this translation among the functions to be performed by the model system constructed.

Fig. 4 shows an expanded functional block diagram of the system as it has been described so far. While the shift register and translator blocks are shown separately in the figure, it will be shown later how it was found possible to so combine these functions as to cause the translation of the reflected-binary number into its natural-binary equivalent as the number sits in the stages of the register.

The system is under the timing control of a program generator which puts out pulses to the various parts of the system, causing them to perform their respective functions at the proper instants in the sequence of system events. The program sequence is illustrated in Fig. 5. The relation of the program pulses to the operation of the encoder-transmitter is as follows:
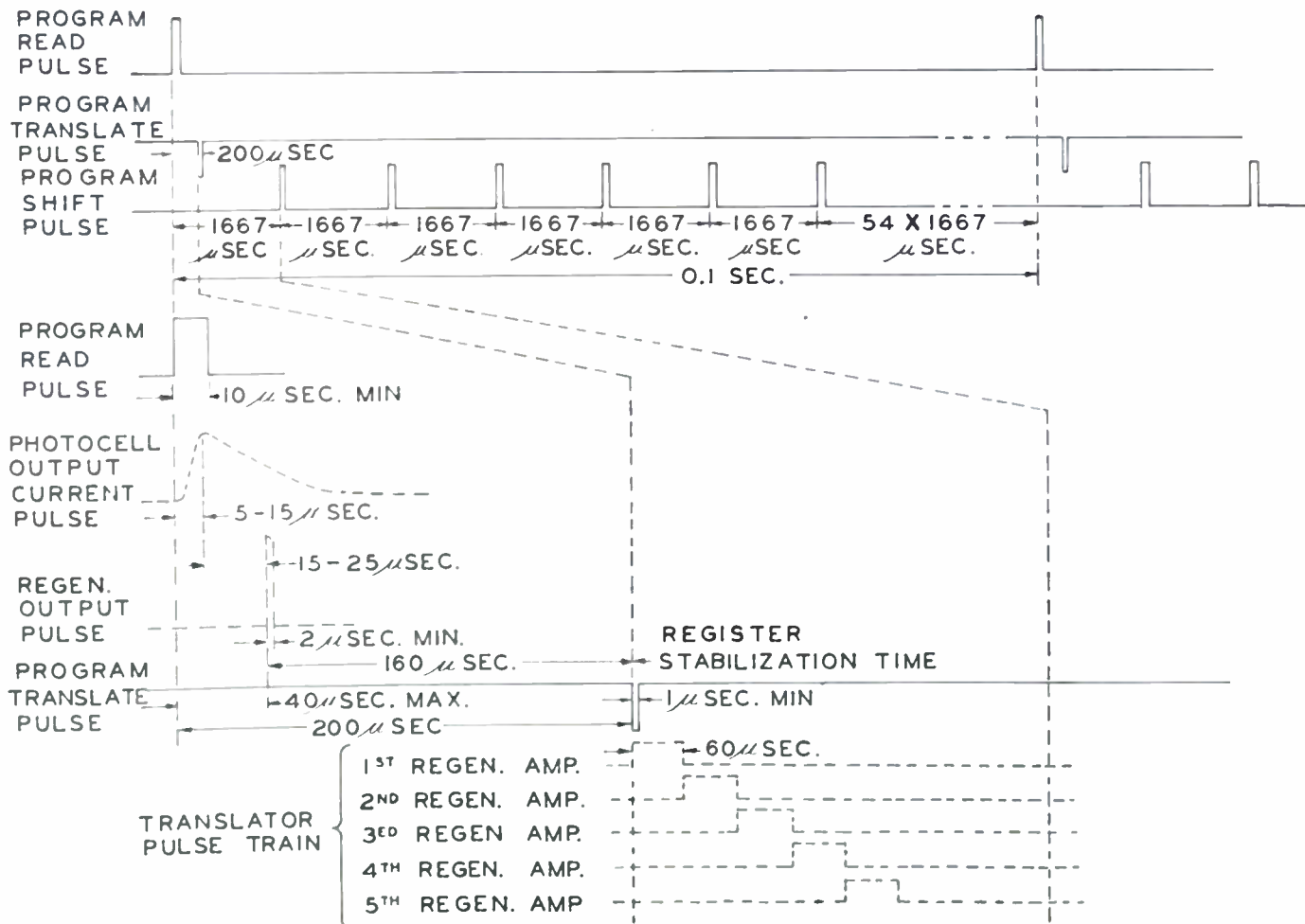
Fig. 5—System program sequence.

A READ pulse applied to the light source causes simultaneous photoelectric sampling of the reflected-binary code digits on the code plate. These digits appear at the pulse-regenerator outputs in the form of parallel reflected-binary digits and are stored simultaneously in the register. After stabilization of the register is complete, a negative TRANSLATE pulse applied to the translator permits a particular NOT AND comparison of digits with the result that the stored reflected-binary digits are translated into stored natural-binary digits. After this process is complete, a series of shift pulses applied to the register causes a shift of the natural-binary digits out of the register in time sequence.

The time interval between READ and TRANSLATE pulses is expanded to show the timing delays in the encoder, regenerator, and register which must elapse before the TRANSLATE pulse is applied. Photocell output-current pulses reach their peak value practically simultaneously with the light flash about 5–15 μsec after the READ pulse begins. Storage delay in the inputs of the pulse regenerators adds about 15–25 μsec to the light-flash delay with a resulting maximum delay of 40 μsec between the rise of the READ pulse and the rise of the regenerator output pulses. A stabilization time of about 160 μsec is then allowed for the registers to store the parallel code digits before translation is initiated. The negative TRANSLATE pulse is then applied in the translator-register to a regenerative pulse amplifier whose trailing edge triggers a second amplifier. The triggering process is duplicated through five amplifiers in sequence, with a resulting train of 60 μsec pulses whose positive leading edges initiate the translation process in the successive register stages, beginning with that of the most significant digit. Next a series of 6 SHIFT pulses shifts the stored binary digits out of the register at time intervals of 1,667 μsec.

The bulk of the remainder of this paper will be devoted to a more detailed discussion of the separate function-blocks of Fig. 4. Attention will be called to the use of transistors and other semiconductor devices throughout and to the circuital means by which the various functions are realized.

A. The Optical Encoder

The functional principles of optical position encoding were discussed previously. This section is devoted to a description of the encoder actually constructed. An early choice had to be made between a number of alternate positions of the light source, code plate, reading slit, photocells, and optical components, if any. Fig. 6 shows side-view and top-view section diagrams of the arrangement finally selected. The code is provided on the *rear* surface of the code plate, and the 0.0005-inch reading slit is on the *front* surface of the stationary plate. This, in turn, is secured to the front face of a plastic block in which are embedded the six photocell elements. This arrangement was chosen because it presented what was

deemed the most advantageous combination of the following desiderata: (1) greatest useful light flux actually reaching the photocells; (2) least fuzzing out of the edges of the code shadows because of diffraction, hence consequent maximization of resolving power; (3) least danger of cross talk between adjacent digit rows; and (4) elimination of the need for optical components and elaborate mechanical means for providing their adjustment.
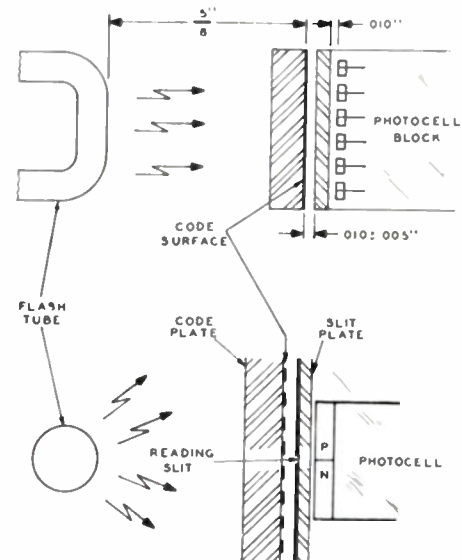
Fig. 6—Diagrams showing the arrangement of the parts of the encoder; upper, side view, lower, top view.

The light source was a Sylvania SA-309 gas flash tube, operated by discharging through it a 2 μF condenser charged to 600 volts, and triggered at the 10-cps repetition rate by a conventional trigger circuit under the control of the READ pulses from the program generator. This light source is indicated in Fig. 7. It is completely housed in a metal shield which also contains the condenser and trigger circuit. The flash tube is about 0.20 inch in diameter and about $\frac{5}{8}$ inch long, so arranged that its length is parallel to the reading slit. The actual core of light has a diameter considerably smaller than 0.2 inch. The tube is about $\frac{5}{8}$ inch away from the code surface.

The code plate was made by photographing, with the desired reduction in size, a master code pattern accurately drawn on paper. The reduction was such that the pattern of the least significant digit consisted of alternating light and dark lines 0.002 inch wide, corresponding to a 0.001-inch quantum distance in reflected binary. Each digit row is 0.100 inch wide so that the entire 6-digit number can be read by a line of photocells 0.600 inch long. Some of the sharpness of the edges of the code pattern is lost in the photographic process because of emulsion grain and aberrations in the photographic lens. Examination of the code plate under the microscope showed a confusion of about 0.0002 inch at the edges of the code lines. The code plates so produced,

FLASH LAMP HOUSING   CODE PLATE HOLDER   PHOTOCELL BLOCK   PULSE REGENERATORS   REGISTER TRANSLATOR
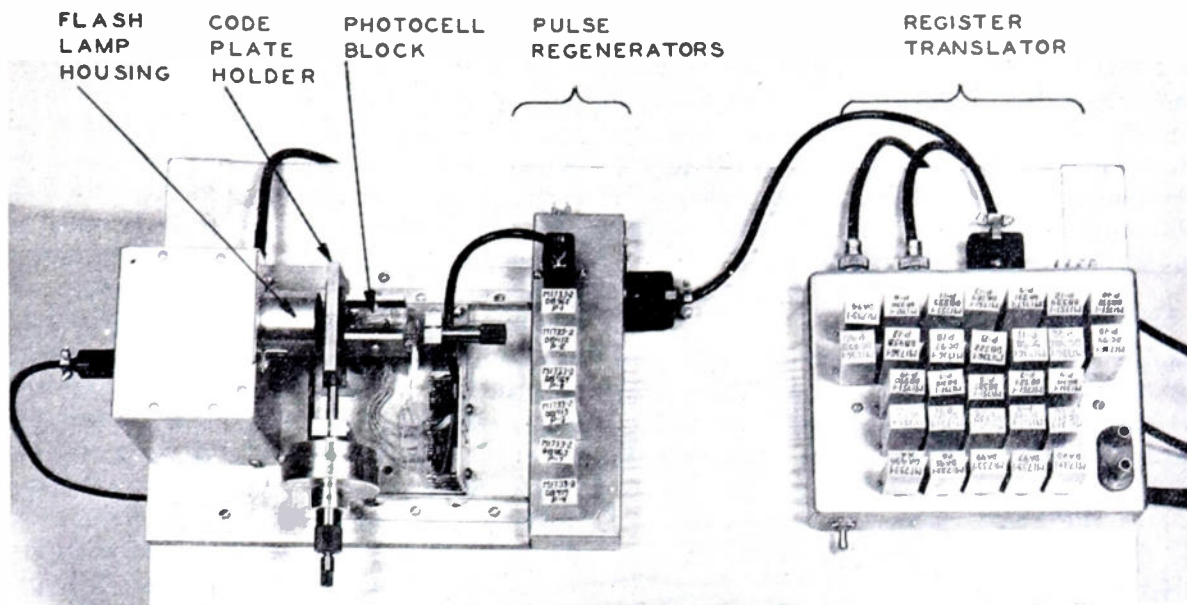
Fig. 7—Photograph of the encoder register.

however, were deemed satisfactory for the purposes of a feasibility demonstration. Fig. 8 shows a reproduction of this code pattern.
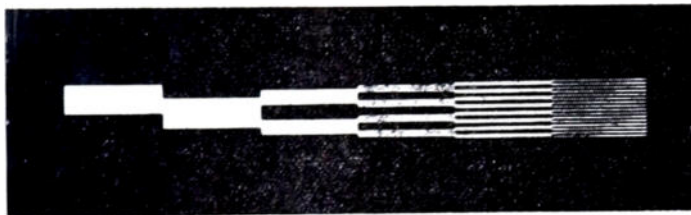
Fig. 8—Reproduction of 6-digit code plate.

The code plate mounting is shown in Fig. 7. It consists of a frame onto which the code plate can be clamped. The frame, being suspended on a spring-leaf mounting, is movable in one dimension parallel to the code plane. The movement of the frame is controlled by a micrometer screw provided with a drum head reading to 0.0001 inch. This head can be rotated manually to provide displacement of the code plate.

The six photocells are embedded in a clear bioplast block which is mounted in a holder having adjustable motion perpendicular to the code plane. This holder is indicated in Fig. 7, with a photocell block in place. Provision is made for clamping the photocell block so that the plane of the reading slit plate is parallel to the code plane, and the clearance between the movable code plate and the stationary reading-slit plate and photocell block can be adjusted. A nominal clearance of 0.010 ±0.005 inch was used in these studies.

The photocells are the p-n junction elements of M-1740 photocells.[2] These elements are mounted 0.100

inch apart, with their p-n junctions in a straight line, and with their front surfaces 0.010 inch inside the front face of the block. Connections are made within the
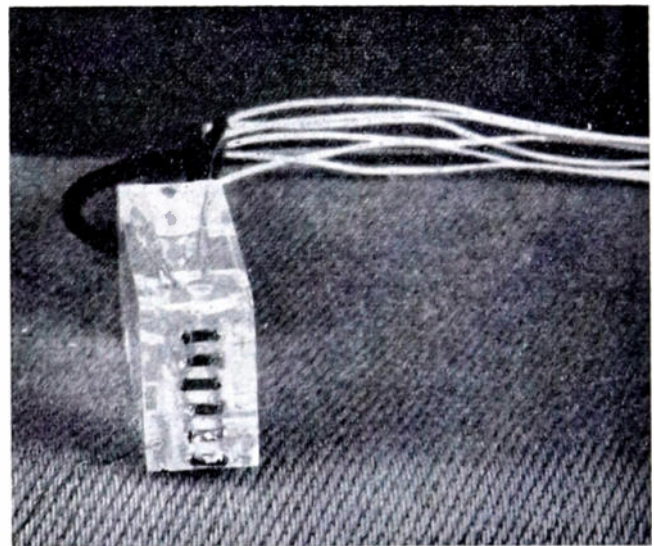
Fig. 9—Photograph of 6-digit photocell block without reading slit.

block to a 7-hole subminiature socket embedded flush with one of the surfaces of the block. The thin glass plate bearing an 0.0005 inch reading slit engraved in an opaque coating on its front surface[3] is attached to the front face of the block, with the slit parallel to and immediately over the line of the p-n junctions. This plate is 0.020 inch thick. The photograph of Fig. 9 shows a 6-cell block before the attachment of the reading slit plate.

[2] J. N. Shive, "Properties of the M-1740 p-n junction photocell," vol. 40, pp. 1410–1414; this issue.

[3] The reading-slit plate was engraved by Lyman Nichols, Cherryhurst, Fort Collins, Colo.

With the geometry of the optical parts of the encoder as described, the width of the band of light incident upon a photocell surface due to the width of the reading slit, the angular width of the source as seen through the reading slit, and the calculated broadening due to diffraction, are about 0.012 inch. It is evident that this width is less than the width of the spatial response curve of the photocell;[4] hence most of the light entering the reading slit is effectively used by the photocell.

B. Pulse Regenerator

A typical code-resolution envelope of peak p-n-junction photocell output current versus relative position of the code plate for any one digit is shown in Fig. 10. The dotted waves indicate the shape of the output-
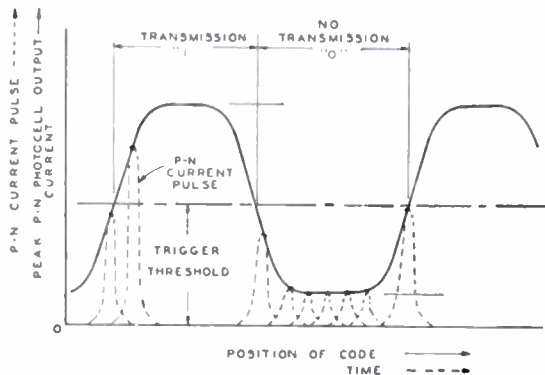
Fig. 10—Code-resolution curve of photocell output.

current pulses as functions of time. The envelope exhibits a trapezoidal form, deviating from the ideal rectangular (full on-full off) form because of the finite width of the reading slit. In addition some light leakage exists in opaque positions due to transmission through adjacent transparent portions of the code to the defining aperture. To reduce transition ambiguity to a small fraction of a quantum step a level discriminator is needed which will ignore output pulses when their current peaks are below a certain threshold amplitude and which will pass those with current peaks above this threshold. Ideally such a threshold should be equal to the peak p-n output-current-pulse amplitude occurring at equal on-off code distance points as indicated on Fig. 10. Regeneration of the p-n current pulses is also necessary because of their limited amplitude and energy. A circuit which conveniently combines the functions of threshold discrimination and pulse regeneration is the transistor monostable multivibrator of Fig. 11.

The typical emitter characteristic shows a negative-resistance portion which permits monostable multivibrator action. With negative-emitter bias the input-impedance is high (~25,000 ohms) and remains high until the triggering point is reached. A zero-impedance voltage source (condenser C) supplies the necessary current during the short but finite interval when the operating point is traversing the zero-impedance region. In order that the critical triggering voltage be independent

[4] J. N. Shive, *op. cit.*, fig. 3, p. 1411.

of the value of C, the latter must be small enough to be almost completely charged to the desired triggering value during the rise time of the p-n output current pulse. For the type of pulse shape encountered in this application that optimum value of C was found to be 300 $\mu\mu f$, but such a small C resulted in a rather short output pulse (~1 μsec). To increase the output-pulse energy somewhat at a slight cost in input sensitivity, C was raised to 500 $\mu\mu f$. The resulting output pulse is shown in Fig. 11. Its amplitude exceeds a minimum amplitude (V_2) for an average duration of 4 μsec. The energy and amplitude of this pulse are adequate for operation of the following register. In cases where a longer output pulse is required, an additional regenerator of the same type but with a much larger value of C may be used in tandem following the former circuit.

The monostable multivibrator in Fig. 11 has a trigger threshold which is reasonably independent of transistor variations. The bias current flowing through R_1 is essentially constant, determined largely by the values of E_b and R_1, since R_1 is much higher than the transistor input impedance to ground in the *off* condition. Trigger threshold is stabilized by shunting between emitter and base a resistance R_2 whose value is low in comparison
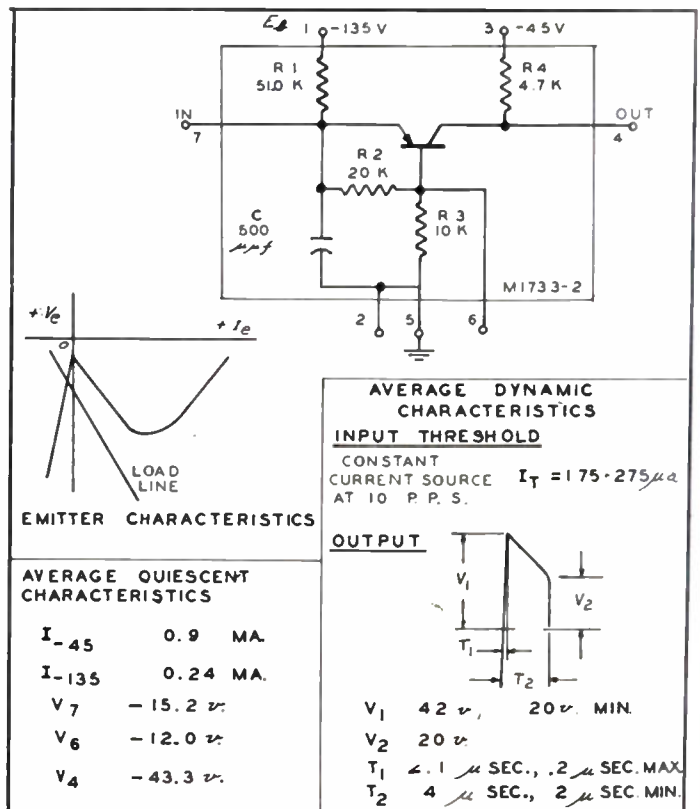
Fig. 11—Transistor pulse regenerator characteristics.

to the back resistance of the emitter. A small variation in threshold is caused mainly by a complex combination of currents flowing through the collector-to-emitter leakage resistance and base resistance. When the circuit is used with trigger threshold suitable for this application, the maximum threshold variation is about ±25 per cent, and is tolerable as a level reading variation.
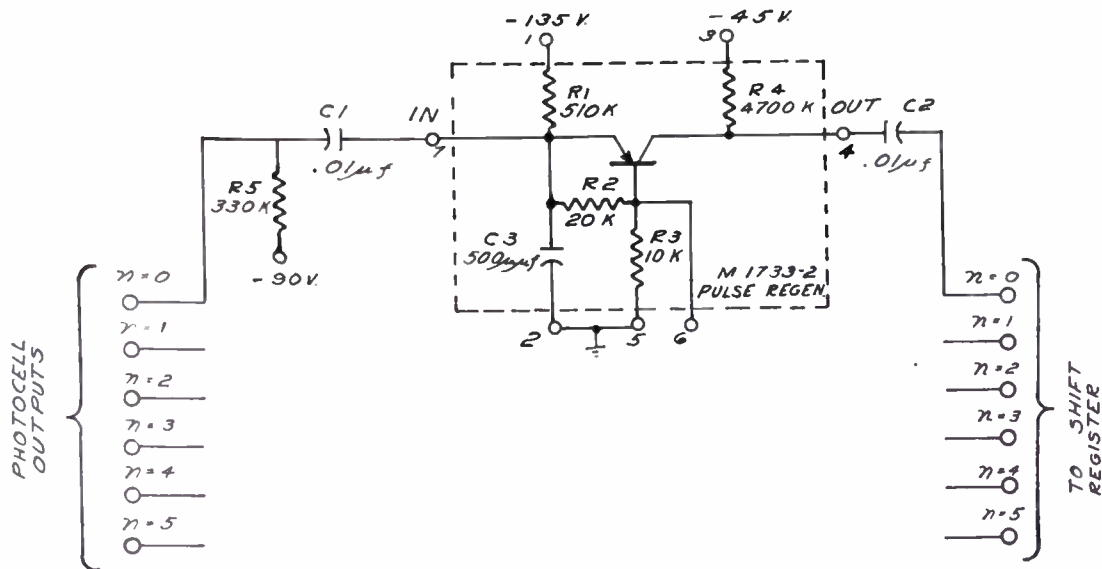
Fig. 12—Pulse regenerator connecting circuits.

The maximum trigger rate for this regenerator is 10,000 times per second. The pulse regenerator was coded M11733-2 and was packaged in bioplast using the auto-sembly technique of wiring. The package was designed to fit into a 7-pin miniature tube socket. The 6 M11733-2 regenerator circuit packages are shown in Fig. 7 following the encoder to the right.

The circuits interconnecting the pulse regenerator photocells and registers are shown in Fig. 12. Negative bias is supplied to each photocell through bias resistors R_5. The bias resistance was chosen to be small in comparison to the very high (several megohms) reverse impedance of the p-n photocell but large enough not to load appreciably the regenerator input impedance of 25,000 ohms.

The photocell output currents are coupled to the regenerators through condensers C_1 to insure that slow dark current increases with temperature will not trigger any regenerator falsely. If this restriction were not present, photocell bias could have been supplied by the negative dc voltage at the regenerator input. An advantage accruing from the use of a pulsating light source is that the ac sensitivity characteristics of the p-n junction photocell, which are fairly constant with temperature, can be utilized.

C. Translator—Register

1. Digit Shifting Register. The purpose of the shift register is to store the parallel code information received from the encoder, and under the control of the program generator to shift the code information serially to the register output.[5] The particular type of shift register employed in this system is shown functionally in Fig. 13. The principal part of the register is the ONE BIT storage circuit, a transistor circuit similar in action to

<hr>

[5] The general problems of information storage and digit shifting are discussed in a companion article by J. R. Harris, "A transistor shift register and serial adder," Proc. I.R.E., vol. 40, pp. 1597–1603; this issue.

the familiar bi-stable multivibrator. This circuit is described in detail in Harris's article. Each of these storage circuits is set to a condition 1 or left in a condition 0 depending on the original code information from its
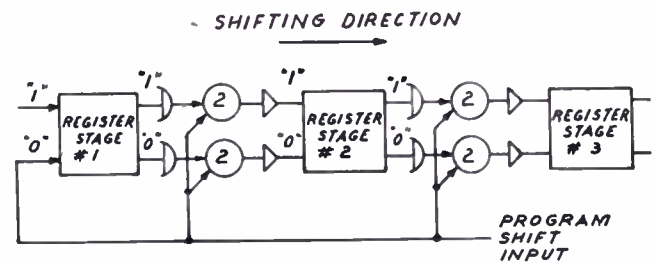
Fig. 13—Functional diagram of digit-shifting register.

associated digit of the encoder. At the time it is required to step the first digit out of the register, the first program SHIFT pulse is applied simultaneously to all the interstage threshold two AND gates and the set-to-0 input of storage circuit no. 1. The first SHIFT pulse, therefore, sets storage circuit no. 1 to 0 and enables one of the two AND gates in all of the interstages. That AND gate which is enabled is determined by the original condition of the storage circuit just ahead of the interstage. The delay provides the memory of this condition even though the controlling storage circuit may already be changing condition. Therefore, if storage circuit no. 1 is in condition 1 just before the first SHIFT pulse is applied, the gate on the set-to-1 lead of storage circuit no. 2 is enabled, causing the first SHIFT pulse to set circuit no. 2 to 1 if it was previously in condition 0, or leaving it in condition 1 if it was initially in condition 1. Likewise, the state of storage circuit no. 2 is transferred to circuit no. 3 and the state of storage circuit no. 3 to a following circuit at the time of the first SHIFT pulse. In effect, the original number stored in the register is shifted one digit to the right. Successive SHIFT pulses perform the same operation, each pulse shifting the

stored number one digit to the right until all digits of the number have been shifted out of the register. In the actual shifting circuit used for the combined translator-register circuit, the first two stages of which are shown on Fig. 14, the delay and gating functions are accomplished with a simple R, C, and diode network, the circuit of which will be described in detail later. These gates are driven from a single transistor regenerative amplifier triggered by the program SHIFT pulse.
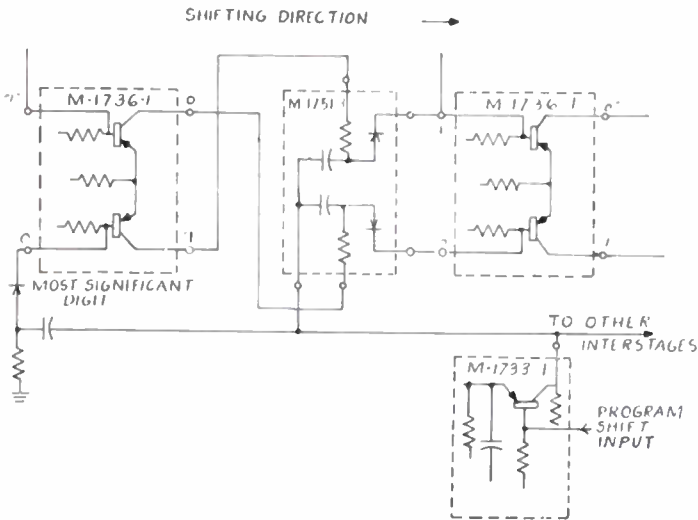
Fig. 14—Schematic of digit-shifting register.

Fig. 15—Comparison of reflected- and natural-binary codes.

An important feature of this type of shift register is that the interstages are entirely under the control of the program generator; hence they will not act on changes in the register which occur because of other processes such as the translation process which is described in the following section.

2. Translator. Although it is possible to describe the arithmetical gymnastics which show the required translator process, it is simpler to describe here the basic process by comparison of the representative groups of reflected-binary and natural-binary numbers shown in Fig. 15. Although only four digits are shown, the general rules apply to any number of digits. In examining the two different codes it is readily seen that the most significant digits of the codes are always the same in equivalent numbers. In the example shown in Fig. 15 the first translation step shows the most significant digit (first) becoming binary without change. The next or second digit of the natural-binary number depends, however, on the comparison of the second reflected-binary digit and the first binary digit. If in this comparison the digits are alike (0, 0, or 1, 1), the resulting binary digit is a 0; if they are different (0, 1, or 1, 0), the resulting second binary digit is 1. In turn, the third binary digit depends on the same comparison of this resulting second binary digit with the third reflected-binary digit. When these comparisons have been completed down to the least significant digit, the result is the equivalent binary number.

A basic functional configuration for translation of parallel reflected-binary code is shown in Fig. 16. Here

the NOT AND gates provide the series of comparisons described above. The NOT AND gate is a device such that if either one, and only one, of the two inputs is
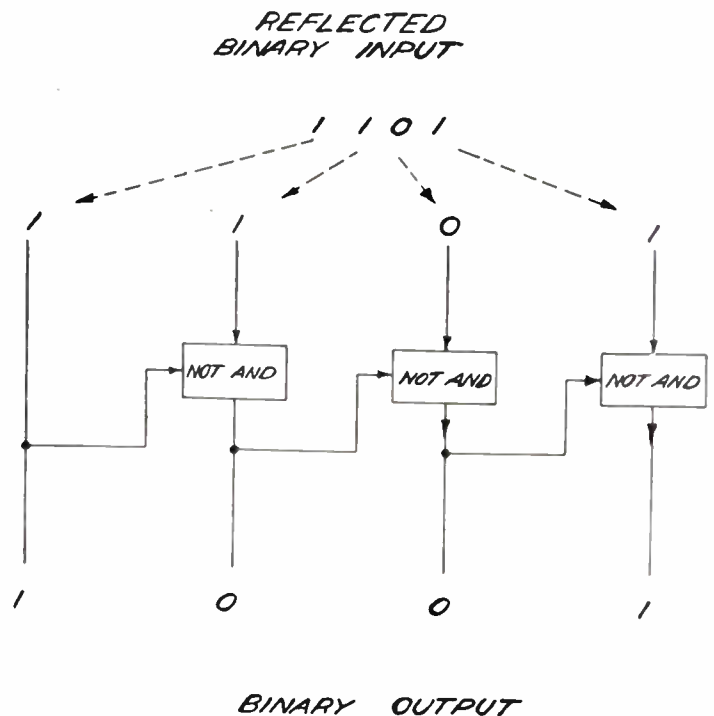
Fig. 16—Parallel translation.

stimulated with 1, the output will be a 1. If both or neither are stimulated, the output is 0. The connections to these gates are such that the nth binary digit is the

result of the NOT AND comparison of the nth reflected-binary digit and the (n+1)th digit resulting from a similar previous comparison. It is apparent that the binary input to any NOT AND gate is delayed with respect to the reflected-binary input from the encoder by the time required for the previous comparisons of the more significant digits. The NOT AND gate must therefore be capable of comparing two input pulses that are not necessarily coincident.

Investigations of the shift-register circuits showed the possibility of combining the functions of translator and shift register, reducing the total number of components required for these two functions. This resulted in a program controlled NOT AND gate configuration shown in Fig. 17. The essential parts of this gate are the ONE BIT register of the shift register, and a supplementary threshold two AND gate. As indicated in this figure, steering control has been added to the second register. The function of the steering control is to control input pulses in such a way that they always change the state of the register. That is, if the register is in a state of 1, a pulse on the input will change it to a 0; if the state is a 0, a pulse will change it to a 1. This action is identical to the count action of a binary-counter stage.
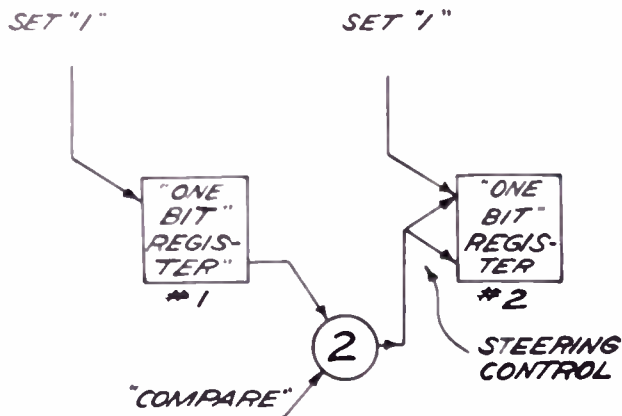
Fig. 17—Program controlled NOT AND gate.

To examine the operation of this gate, assume that both registers have been cleared by a previous operation and that a new pair of digits is read into the register stages by setting those stages to 1 where the digit is a 1 and leaving the other stages in a zero condition. A TRANSLATE pulse is then applied to one of the inputs of the AND gate. This gate is connected to the first register in such a way as to transmit the compare pulse only if that register is in a 1 condition. If the gate does transmit a pulse, it will then change the state of the second register. The final state of the second register is the NOT AND comparison of the original numbers. This can readily be checked by trying the four possible combinations: 0, 0 and 1, 1 which result in a 0; and 1, 0 and 0, 1 which result in a 1. A desirable feature of this gate is that the input signals to be compared do not have to be coincident. The only timing requirement is that the registers be set before the TRANSLATE pulse is applied.

Fig. 18 shows the configuration for a parallel reflected-binary to natural-binary translator incorporating this NOT AND gate. Each register stage is also a part of a complete shift register, the connecting circuits of which are not shown for reasons of simplicity. To examine the translation operation, assume that the shift
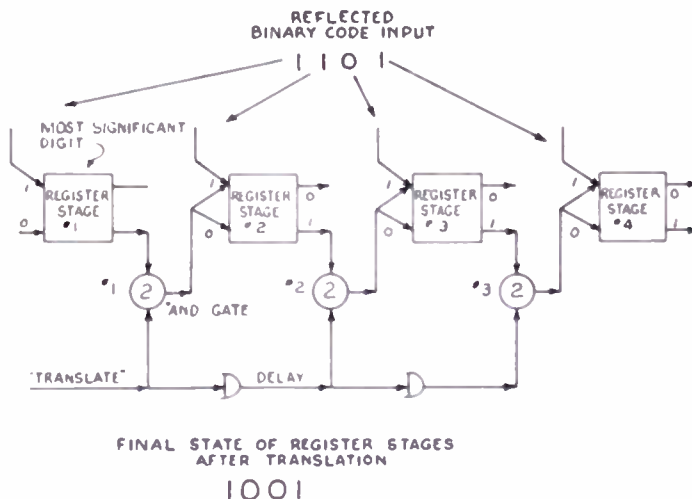
Fig. 18—Functional schematic of reflected-binary to natural-binary translator.

register has been cleared initially so that all ONE BIT registers are in a state of 0. Under the control of the system program READ pulse a reflected-binary code number is "read" in to the encoder and the digits, in short-duration pulse form, are applied to the inputs of the register. When the condition of the register has stabilized, storing this parallel reflected-binary code number, a TRANSLATE program pulse is applied at the first AND gate. The NOT AND comparison described above is then made for the most significant digit in the first register and the next digit in the second register. The resulting condition of the second register is the binary equivalent of the original reflected-binary digit. After a time delay of sufficient length to insure translation and stabilization in the second register, the TRANSLATE pulse is applied to the second gate, comparing the state of the second and third registers. After this comparison, the digit in the third register is in binary form. Again, after an appropriate delay the TRANSLATE pulse is applied to the third gate, with the result that all registers are then storing the binary equivalent of the original reflected-binary number.

Fig. 19 shows in schematic form the details of the translator described above. For simplicity, only three storage registers are shown and the interstages for digit shifting are omitted. The method for steering control for this register is shown in Fig. 20. In this figure assume transistor no. 1 of the register is in the *on* condition, in which case its base and collector are at approximately -10 and -12 volts, respectively. The difference between these voltages, 2 volts, is the effective back voltage on diode no. 1 of the steering circuit. At the same time the base and collector of the *off* transistor are -4
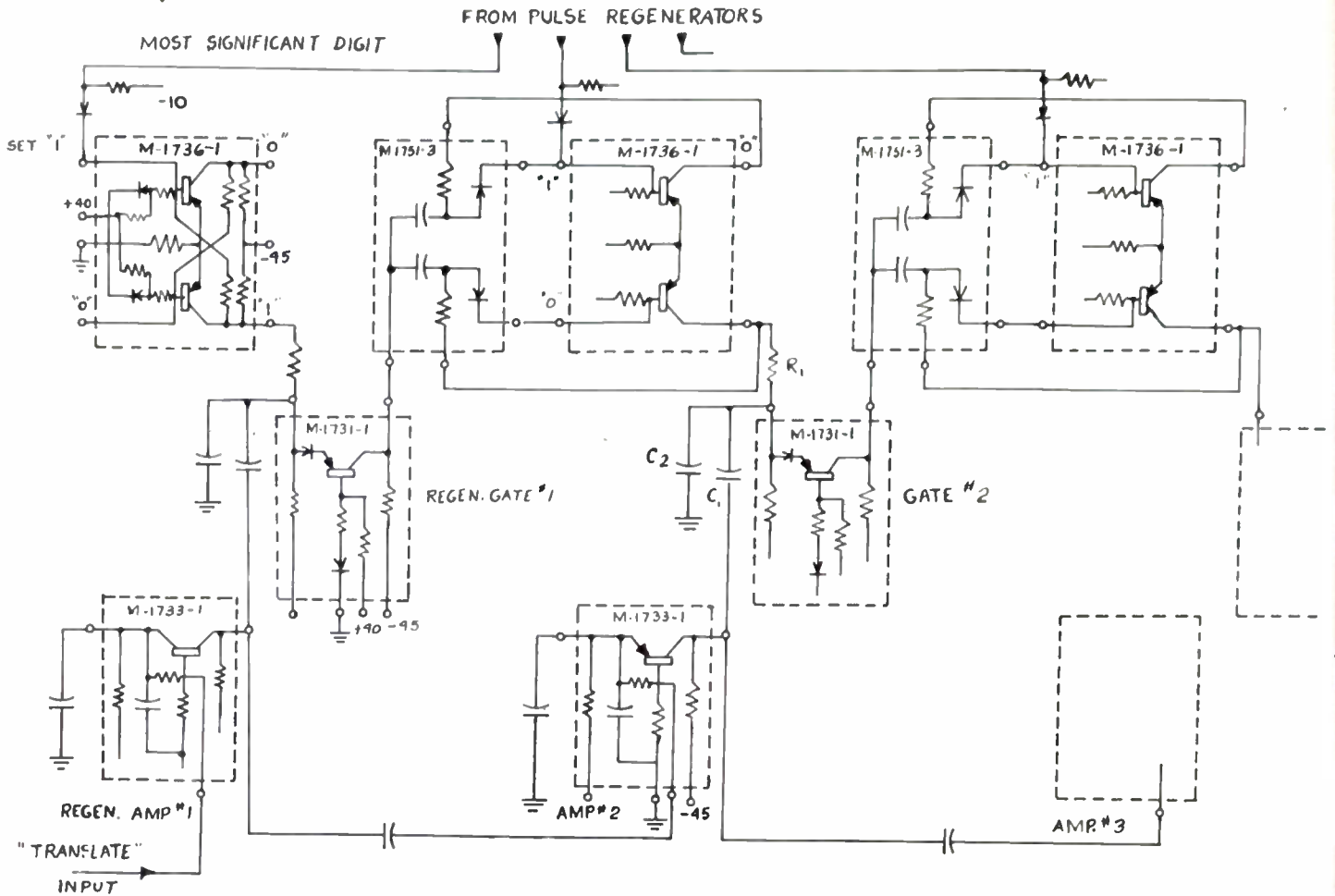
Fig. 19—Partial translator schematic.

and −35 volts, respectively, biasing diode no. 2 approximately 31 volts in the back direction. It is apparent that under these conditions a positive input pulse on the control circuit, having an amplitude between 20 and 35 volts, would be readily steered through diode no. 1 to the base of transistor no. 1, turning it *off* and transistor no. 2 *on*. The storage effect of C_1 and C_2 tends to keep the input pulse correctly steered until the register has completely changed state. After several time constants of R_1C_1 (or R_2C_2) diode no. 2 becomes the readily conducting diode to an input pulse, enabling the next input pulse to turn transistor no. 2 *off* and no. 1 *on*. This circuit is identical to the circuit of the interstage in the shift register. In the interstage, however, the input to a register stage is controlled by the state of the previous stage rather than the same stage. To provide sufficient triggering level for the register stage a transistor regenerative gate[6] is used for the AND gating function as indicated in Fig. 19.

The delays for the TRANSLATE pulses are obtained from the cascaded stages of regenerative pulse amplifiers in the following manner: At the time the negative program TRANSLATE pulse is applied, the first regenerative amplifier (REGEN AMP no. 1) is triggered, providing a positive output pulse of about

60 μsec duration. The positive leading edge of this pulse is applied to the first gate, while 60 μsec later the negative trailing edge triggers AMP no. 2. The positive leading edge of the output pulse of AMP no. 2 is applied to the second gate; while 60 μsec later AMP no. 3 is triggered by the trailing edge. This action continues along the entire chain of amplifiers, providing in each step a delay of approximately 60 μsec.
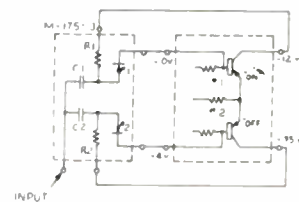
Fig. 20—Steering control.

As seen in the diagram of Fig. 19, the output of the regenerative amplifier is differentiated and attenuated from approximately 25 volts to 20 volts by the condenser divider network, C_1 and C_2, providing the proper critical triggering level for the regenerative gate. When the regenerative gate is triggered, the emitter current maintaining the transition of the gate in the "ON" condition is supplied by the charge stored in C_1 and C_2. C_1 and C_2 in parallel, therefore, determine the charge available in the output pulse of the gate. Two considera-

[6] R. L. Trent, "A transistor reversible binary counter," PROC. I.R.E., vol. 40, pp. 1562–1573; in this issue.

tions determine the value for the parallel combination of C_1 and C_2. If the register controlling the regenerative gate has just changed state because of the translation of the digit it was storing, the change must be applied to the gate before translation of the next digit by charging C_1 and C_2 in parallel through R_1. The higher the value of C_1 plus C_2, the greater the time required per digit for translation. Yet C_1 and C_2 must be sufficiently large for the gate to supply an adequate signal charge for setting its associated register stage. As high-speed translation was not required in this system, it was possible to make C_1 plus C_2 large enough to meet this requirement. Where higher speed translation may be required, a nonstorage type gate might be used in conjunction with regenerated gain.

3. Program Generator. The sequence of timing pulses is produced by a vacuum-tube program generator of conventional general layout. The use of vacuum tubes, while appearing as heresy in a transistor endeavor, was dictated by the following circumstances. (a) The requisite number of transistors was not readily available at the time this project was undertaken. (b) An all-transistor program generator had already been constructed as part of a companion project, so that feasibility proof of such a system function using transistors was not at issue. The pulse program for the present system has already been discussed.

III. Performance

The ultimate proof of performance is the ability of the system to deliver to the output, every tenth of a second, a six-digit natural-binary word whose 1's and 0's are recognizable and which is uniquely representative of the position of the shaft, with resolution of every quantum step. The system as constructed met these requirements. Visual evaluation of the performance was obtained by terminating the register output in a resistance simulating an output load and observing on an oscilloscope screen the appearance of the output word pulses. The entire traverse of the encoder could be scanned without the loss of a single quantum and without failure to encode, register, translate, or shift.

It can be appreciated from the consideration of the foregoing sections that the resolution of every quantum position depends on the proper relative adjustments of light intensity, photocell sensitivity, and trigger threshold of the pulse regenerator. A decrease in light intensity and/or a decrease in photocell sensitivity causes the system to read 0 slightly before the code plate comes to the exact transition point of a digit from 1 to 0. Similarly, decreasing the trigger threshold causes the system to read 1 slightly beyond the exact transition point of a digit from 1 to 0. If the system is to resolve every quantum, these deviations must be kept less than ± 0.0005 inch. Operational tests have shown that the variability of photocell sensitivity from unit to unit and the variability of trigger threshold from channel to channel in the system as constructed are comparatively unimportant. The most serious performance limitation of the present system is the de-

crease with time, over a period of a few hundred hours, of the intensity of the light source.

One of the objectives of this project was to demonstrate the conservation of size, weight, and electric power which the use of transistors and miniature circuitry makes possible. In Fig. 7, which shows the system exclusive of power supply and program generator, the encoder is $6\frac{1}{2}$ by $8\frac{1}{2}$ by $4\frac{1}{2}$ inches; the bank of regenerators is 2 by $7\frac{1}{2}$ by $3\frac{1}{4}$ inches; and the translator-register is $5\frac{1}{2}$ by $7\frac{1}{2}$ by $3\frac{1}{4}$ inches. The combined weight of these three units is 7 pounds. A power consumption of 16 watts is itemizable.

light source	8.0 watts
photocell block	0.36 watt
pulse regenerators	0.45 watt
translator-register	7.2 watts
	16.0 watts

Substracting the power used by the light source, the total power per digit for the performance of the purely information-handling functions of reading, regeneration, storage, translation, and shifting is only 1.3 watts.

When the system was first put into operation power-supply voltage deviations of ± 20 per cent from nominal were tolerable for operation of the system. At the end of a six-month period these tolerances decreased to about ± 5 per cent. This difficulty was not considered irremediable since new techniques in packaging transistor circuits have been developed which avoid these aging effects.

The original model was operated continuously during working hours for a period of three months. During this time one replacement of the discharge lamp and one replacement of the trigger tube in the lamp circuit were made. All photocell, diode, and transistor circuits performed satisfactorily.

IV. Conclusion

An optical position encoder and translating shift register have been described. This system performs the functions of photoelectric encoding, pulse regeneration, digit storage, reflected-to-natural binary translation, and digit shifting by means of circuits using transistors and other semiconductor devices. Although these functions may be performed by vacuum-tube circuits, it has been demonstrated that the use of semiconductor active devices results in considerably less power consumption, size, and weight. The satisfactory performance of this system clearly shows the usefulness of transistors in computing systems of some complexity.

Regenerative Amplifier for Digital Computer Applications*

J. H. FELKER†, ASSOCIATE, IRE

Summary—A description of the negative-resistance properties of the point-contact transistor is presented as an introduction to the description of a regenerative amplifier. The choice of circuit parameters for the amplifier is discussed and a sample design presented. The illustrative amplifier regenerates digital information at a megacycle rate and develops pulses with rise times of less than 0.05 μsec. It operates from supply voltages of +6 and −8 volts, with a battery drain of less than 0.05 watt. A complete set of computer building blocks has been designed around the amplifier. Their use is illustrated in two computer applications.

I. INTRODUCTION

A. Basic Considerations

THE following introductory material has been included in the hope that it will supply some of the basic concepts which, when missing, make transistor circuits difficult to understand. The readers with a background in transistor circuitry may well choose to proceed directly to the next section.

For the circuit designer one of the most useful transistor concepts in nonlinear circuits is based on the properties of semiconductor diodes. Consider the conventional germanium point-contact diode. When this diode is biased in the back direction, it has a high impedance that is generally greater than 10,000 ohms. When the voltage is carried positive, the impedance of the diode is reduced to a value of a few hundred ohms. Consider a piece of germanium having two metallic points on it that are very close together. Connect the germanium (base) to ground. Bias one point (emitter) in the forward direction and bias the other point (collector) in the back direction. Looking into the emitter, a low impedance of the order of several hundred ohms will be seen. Looking into the collector, a much higher impedance will be seen. As the emitter current is varied, the back impedance of the collector diode is modulated. The transistor emitter can be viewed as a terminal which controls the effective impedance between collector and base.

A transistor in which a change of emitter current produces an equal current change in the collector will have voltage gain because of higher collector impedance. In the point-contact transistor there is a current multiplication in addition to voltage gain. These transistors can be built to exhibit consistently current multiplication greater than 1. When the current gain,

$$\alpha = -\left.\frac{\Delta i_c}{\Delta i_e}\right|_{\text{constant collector voltage,}} \tag{1}$$

where i_e flows into the emitter and i_c flows into the collector, is greater than 1, a negative input impedance can be obtained. This negative impedance is not limited to a band of frequencies, but is exhibited at dc as well as high frequencies. Because of this, the equivalent of many conventional two-tube circuits, such as free-running multivibrators, single-shot multivibrators, and bistable flip-flops, can be built with a single transistor. Since the point-contact transistor has a greater bandwidth than the junction type, and also exhibits useful negative-resistance properties in simple circuits, most high-speed switching work is based on its use.

WHERE $\alpha_c = \dfrac{r_m + r_b}{r_c + r_b + R_l}$
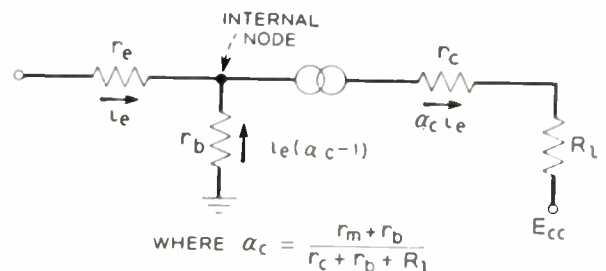
(a)
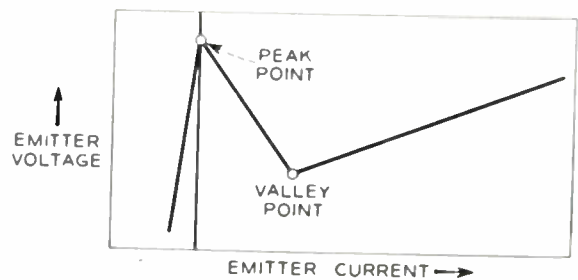
(b)

Fig. 1—Emitter characteristic. (a) Equivalent circuit. (b) Input characteristic for $(\alpha_c - 1) r_b > r_e$.

An equivalent circuit of the transistor is shown in Fig. 1(a).[1] This is the conventional circuit except that, rather than the current gain for zero load impedance (1), the current gain,

$$\alpha_c = \frac{r_m + r_b}{r_c + r_b + R_l}, \tag{2}$$

is used; this represents the actual ratio of the collector current to the emitter current for a specific load impedance (R_l).

[1] R. M. Ryder and R. J. Kircher, "Some circuit aspects of the transistor," *Bell Sys. Tech. Jour.*, vol. 28, pp. 367–401; July, 1949.

B. Cutoff and Overload

Remembering the analogy between a transistor and two diodes, consider the equivalent circuit with the emitter and collector both biased in the negative direction. The emitter and collector currents that flow will each be the back current of a diode and the transistor is cutoff. (It is assumed that the transistor base ma-material is *n*-type.) If the emitter voltage is brought up towards ground, the collector current will remain constant and the emitter current will decrease slowly (Fig. 1(b)) because the impedance seen at the emitter is high. When the emitter current is zero, the collector current flowing through the base impedance will have biased the internal node of the equivalent circuit slightly negative. As soon as the emitter voltage gets positive enough for positive emitter current to flow, r_e becomes quite low and comparable to the forward impedance of a diode. At that point the transistor becomes an active element and exhibits current gain. This point is known as the "peak point."

Suppose, for example, that the α_c of the transistor is 2. For every milliampere of forward current put into the emitter, 2 ma will flow out of the collector and 1 ma will flow into the base terminal carrying the internal node of the equivalent circuit below ground. If the base impedance of the transistor is greater than the emitter impedance, the emitter will be carried negative as the positive emitter current is increased. This is the mechanism whereby the transistor develops a negative input impedance and is fundamental in many circuits. The slope of the characteristic is given by

$$r = -(\alpha_c - 1)r_b + r_e. \tag{3}$$

As the emitter current continues to increase, the internal node goes farther and farther negative. At the same time, the collector current flowing through the load impedance is raising the collector terminal more positive. When the internal node has fallen to the voltage to which the collector terminal has risen, the transistor gain is cutoff. This point is called the "valley point" and occurs because the flow of the emitter current has reduced the effective collector impedance to a minimum value. In one sense this is comparable to the saturation which occurs when a vacuum tube has a series cathode resistor. A positive grid voltage may bring the cathode up and the plate down to the same potential, thereby reducing the ac impedance between plate and cathode to the impedance of a diode. The two cases are different in the important respect that for the transistor the peak (cutoff) point is more positive than the valley (saturation) point, which makes the emitter current a multiple-valued function of the emitter voltage. Another practical difference is that the internal node, unlike the cathode terminal of a vacuum tube, is not accessible.

Once the emitter current is large enough to overload the transistor, the device becomes a nonlinear but pas-sive network and further increases in emitter current will raise emitter voltage. Impedance seen at the emitter will be the impedance of the equivalent emitter diode plus parallel impedance of base and collector circuits. Impedance of the collector circuit will be approximately equal to load impedance because the drop between the internal node and the collector terminal approaches a constant and low value in overload.

The picture given above is greatly idealized. The peak point was described as occurring at the transition point between negative and positive emitter currents. Actually, it cannot occur (see (3)) until α_c is greater than 1. This does not occur at zero emitter current, but generally occurs at a positive current of the order of 50 μa. The characteristics obtained in practice are generally rounded out at the valley point and the peak point. The fact that the transition from positive impedance to negative impedance in the vicinity of the peak point is gradual increases the current required to trigger the transistor because triggering is determined by the magnitude of the negative resistance (see (9)).

It is useful to have a general idea of how the transistor and circuit parameters affect the turning points.[2] This understanding is best obtained by considering the potential at the internal node of the transistor equivalent circuit (Fig. 1(a)). The principal reason that the peak voltage is not at ground is that for zero emitter current and a specified collector voltage V_c, some collector current, $I_c(0, V_c)$, flows through the base resistance r_b and biases the internal node negative. The larger the product $r_b I_c(0, V_c)$ the more negative the peak point, because the transistor becomes an active element when the emitter terminal is positive with respect to the internal node. A high load impedance will make the peak point approach ground because the drop produced across it will make the collector voltage more positive, and therefore lower the collector current. Since the emitter current at the turning point must be sufficiently positive to make (3) equal to zero, the peak voltage measured at the emitter will be slightly more positive than the potential computed for the internal node.

Useful approximations to the valley voltage and current can be computed. At overload the voltage between the internal node and the collector terminal will be reduced to a small and nearly constant value V_s. From the equivalent circuit of Fig. 1(a), the emitter current at the valley (I_v) must satisfy the relationship

$$I_v(\alpha_c - 1)r_b + V_s + I_v \alpha_c R_L = -E_{cc},$$

and therefore

$$I_v = \frac{-E_{cc} - V_s}{(\alpha_c - 1)r_b + \alpha_c R_L}. \tag{4}$$

[2] For a more complete analysis of the transistor turning points, see A. E. Anderson, "Transistors in switching circuits," Proc. I.R.E., vol. 40, pp. 1541–1559; this issue.

This shows that the valley current is increased by a decrease in α_c, r_b, or R_L.

The valley voltage (E_v) will be given by

$$E_v = -I_v(\alpha_c - 1)r_b + I_v r_e,$$

which can be written as

$$E_v = \frac{(E_{cc} + V_s)\left(1 - \dfrac{r_e}{(\alpha_c - 1)r_b}\right)}{1 + \dfrac{\alpha_c R_L}{(\alpha_c - 1)r_b}}. \tag{5}$$

This shows that if $(\alpha_c - 1)r_b$ is large compared to R_L and r_e, then the valley voltage is nearly the collector supply voltage. If the valley voltage is far removed from the collector supply voltage, then any increase in the base impedance or any decrease in the load impedance will deepen the valley.

It should be realized that (4) and (5) are based on assumptions that are not always justified. It was implicitly assumed in the derivations that the peak voltage was zero and that all currents were zero at the peak voltage. It was also assumed that the emitter, base, and collector currents are linearly related for positive emitter current. The assumption that the valley occurs at collector overload may not be justified in particular cases because there is evidence that the high r_b of some transistors drops to a low enough value to make the input resistance (see (3)) positive at a current that is less than the current-producing overload. In spite of these reservations (4) and (5) provide a useful basis for estimating the relative importance of r_e, r_b, R_L, and α_c in determining the valley point.

A numerical example of the use of these equations may be of interest. If for example

$$\left. \begin{array}{ll} r_e = 125 & \alpha_c = 2 \\ r_b = 2000 & E_{cc} = -8 \\ R_L = 500 & V_s = 1 \end{array} \right\}, \tag{6}$$

(3) gives $-1,875$ ohms for the emitter impedance. Equations (4) and (5) give $2\frac{1}{3}$ ma and $-4\frac{3}{8}$ volts for the valley current and voltage.

C. Negative-Resistance Characteristics

The discussion of the equivalent circuit has explained the origin of the negative resistance seen at the emitter. This negative resistance is called the "series type"[3] for reasons which may be explained with reference to Fig. 2(a). Note that at the turning points a and b, the resistance becomes zero. Therefore a generator feeding the negative resistance through a large enough series positive resistance will see, if the circuit does not oscillate, a positive resistance for all values of current and voltage. This is not true of all types of negative resistance. In particular, the shunt type shown in Fig. 2(b) cannot be

[3] G. Crisson, "Negative impedance and the twin 21-type repeater," *Bell Sys. Tech. Jour.*, vol. X, p. 485; July, 1931.

made to look positive by adding series positive resistance because at the turning points a and b the negative resistance has an infinitely large magnitude. However, if a low enough value of positive resistance is put in shunt with this negative resistance, the combination will have a positive resistance even at the turning points. Note that the series type (Fig. 2(a)) goes through zero at the turning point; therefore no matter how low a resistance is put in parallel, in the vicinity of the turning point the parallel combination will be a negative resistance.

The type of negative resistance seen at the base of a transistor when $\alpha_c > 1$ and r_b is small is the shunt type shown in Fig. 2(b). If the base impedance is high, the v-i characteristic will be shown in Fig. 2(c). Note that the characteristic has a negative slope only from a to b and c to d. If positive series resistance is added to a shunt resistance characteristic, the curve of 2(b) will approach the kind shown in 2(c). As the series resistance added approaches ∞, the regions of negative resistance decrease in range but the circuit does not become more
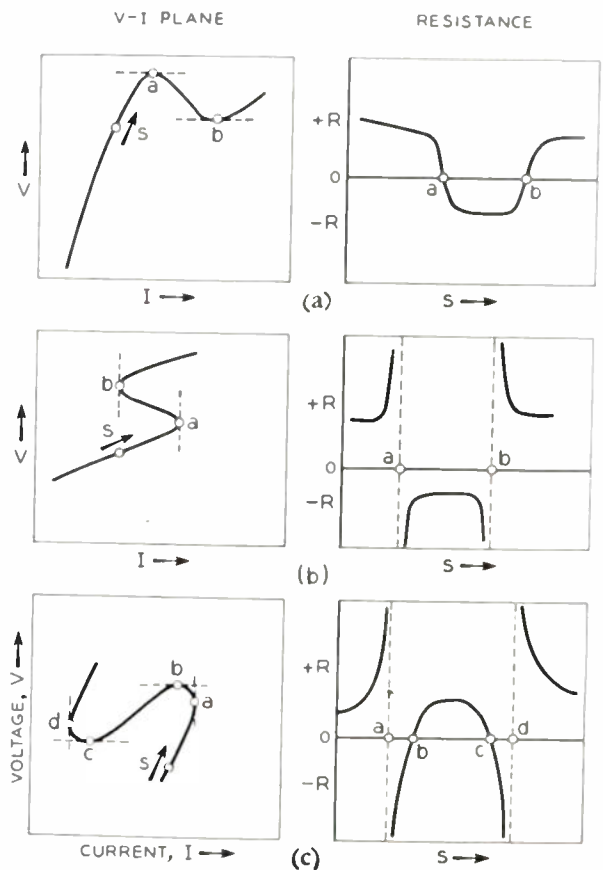
Fig. 2—Negative-resistance characteristics. (a) Series. (b) Shunt. (c) Shunt with series resistance added.

stable. It may be noted that a shunt-type negative resistance is the dual of the series type. The dual of the compound type shown in Fig. 2(c) can be obtained by shunting positive resistance across a series negative resistance.

The question of the stability of a negative resistance[4] is fundamental to transistor pulse circuits. Fig. 3 shows a series negative resistance connected to a variable voltage source through a positive resistance R_1. The negative resistance will always have some capacity shunted across it, stray or otherwise, and this is repre-

(a)

(b)

(c)
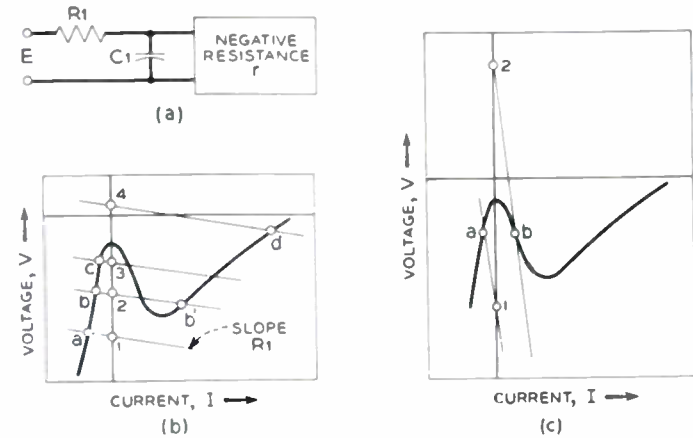
Fig. 3—Emitter load lines. (a) Equivalent circuit. (b) Operation when $R_1 < |r|$. (c) Operation when $R_1 > |r|$.

sented by C_1. The negative resistance itself will also have reactive components because of the finite band-width of the device responsible for the characteristic.

Consider the case 3(b) in which R_1 is much lower than the maximum value of the negative resistance. If the voltage is set at a value below the valley point and successively moved from point 1 to 2 to 3, the circuit will be stable and the operating point will move from point a to b to c. If the voltage is raised to point 4 the circuit will be stable at the operating point d. The circuit is always stable in the sense that it does not oscillate. However, the circuit is unstable in the following sense: If the negative resistance is biased to point b and the voltage E is temporarily displaced from 2 to 4 and then reduced to 2, the circuit may stabilize at the point b' instead of returning to the point b. Whether or not the circuit will stabilize at b' depends upon whether or not the voltage E stays at point 4 long enough for the circuit to switch to the high-current region. This is determined first of all by the time required for C_1 to be charged through R_1 to a voltage greater than the peak point. The second determining factor is the length of time required for the current required at b' to be established through the negative resistance.

If the resistance R_1 is much larger than the maximum negative slope, then the result is as shown on Fig. 3(c). In this case the total resistance seen by the source E is positive, which is illustrated by the fact that the load line never intersects the v-i characteristic in more than one point. Operation at point a is stable, but operation at point b is not necessarily stable and depends upon the reactive components of the negative resistance.

[4] B. G. Farley, "Dynamics of transistor negative-resistance circuits," Proc. I.R.E., vol. 40, pp. 1497–1508; this issue.

The drop in gain of the transistor at high frequencies can be accounted for by assuming that a capacity (c_m) by-passes the mutual impedance (r_m) in the collector circuit. The frequency

$$f_0 = \frac{1}{2\pi r_m c_m} \tag{7}$$

is approximately the same as the frequency at which the short-circuit current gain (1) is down by 3 db. With the assumption that c_m is the only reactive component of the transistor, it can be shown that a transistor operated with its collector and base terminals at ac ground will be unstable with a capacitance C_1 shunted across its emitter. This holds even though

$$R_1 > -r \tag{8}$$

provided that,[4]

$$-2\pi f_0 r C_1 > 1 + \frac{R_2}{R_1}, \tag{9}$$

where r is the dc negative resistance exhibited (3) and

$$R_2 = r_e + \frac{r_b r_c}{r_b + r_c}. \tag{10}$$

(a)

(b)

(c)
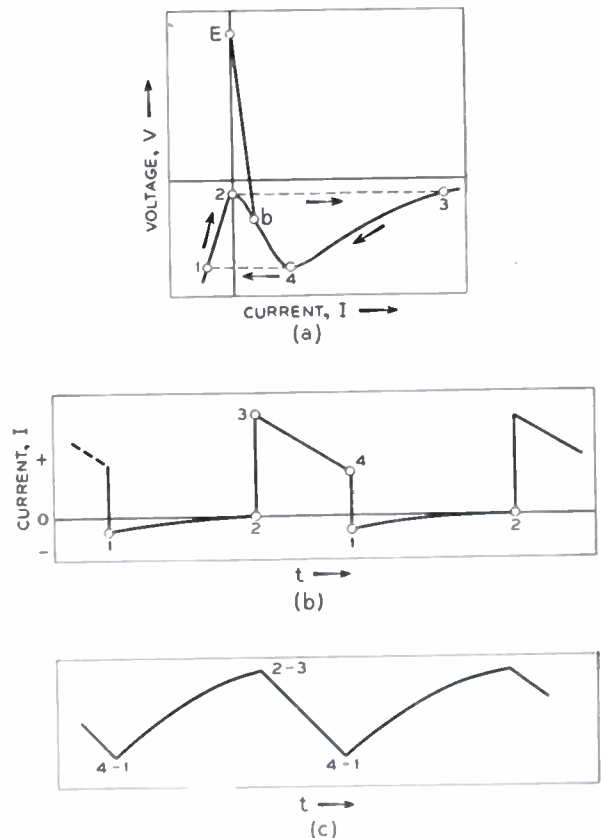
Fig. 4—Free-running oscillator. (a) Locus in V-I plane. (b) Emitter current as function of time. (c) Emitter voltage as a function of time.

If the transistor emitter is biased by a constant current source, then R_2/R_1 is zero. The requirement for instability is then

$$-2\pi f_0 r C_1 > 1,$$

where r is the dc resistance at the emitter for the chosen bias current.

If the product in (9) is very much greater than

$$1 + \frac{R_2}{R_1},$$

the locus of operation in the v-i plane will be as shown in Fig. 4. The capacitor C_1 charges through R_1 and the high back impedance of the emitter diode from point one to the peak point (2). Since the capacitor cannot discharge instantly from the peak voltage, operation will snap out to point 3 as the condenser supplies current to the emitter. The capacitor will then discharge through the low impedance of the saturated transistor to the valley point (4). At the valley point, operation will snap to point 1 and the cycle will repeat. Fig. 4(b) shows the emitter current as a function of time and Fig. 4(c) shows the corresponding voltage waveform. If the product in (9) is gradually decreased, the effect will be first to put a slope on the dotted lines in Fig. 4(a) and finally to cause the locus to collapse to point b when the inequality is no longer satisfied.

II. OPERATION OF REGENERATIVE AMPLIFIER

A. Use of Flip-Flop as Amplifier

The highest speed transistor available for laboratory work has been the Bell Telephone Laboratories developmental type M1734. The current gain, base impedance, and cutoff frequency are sometimes large enough that the inequality of (9) is satisfied even when C_1 is the minimum stray capacity. If, for example, f_0 is 50 mc, C_1 is 4 mmfds, and r is $-1,875$ ohms as per the example (6), then

$$-2\pi f_0 C_1 r = 2.35.$$

Adding impedance in the collector circuit to decrease α_e and the negative resistance, r, does not often stabilize a high-speed transistor because the stray capacity at the collector may by-pass the impedance sufficiently to permit oscillations. By shunting down the emitter with resistance, the circuit can be made dc stable; but it then stabilizes outside the gain region as in Fig. 3(b), and a bistable flip-flop is obtained.
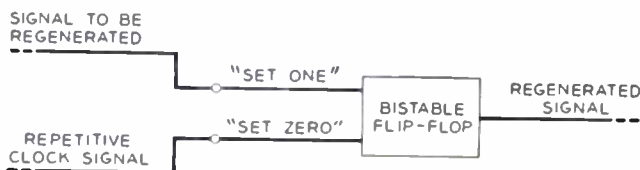
Fig. 5—Flip-flop as a regenerative amplifier.

A bistable flip-flop can be used as a regenerative amplifier (Fig. 5). The signal to be regenerated is connected to the "set-one" terminal. Thus every time an input signal is obtained, the flip-flop sets to its one position and the next clock signal resets it to zero, giving an output pulse from the flip-flop. Some advantages of the flip-flop as a regenerative amplifier are

1. it gives a standard output whose amplitude is independent of the amplitude of the input signal;
2. the duration of the output can be made independent of the duration of the input signal;
3. minimum rise times are obtained because of the regenerative action;
4. transistors now available develop a higher effective current gain in a regenerative circuit.

Fig. 3(b) shows the emitter characteristic of the transistor with a low-impedance load line superimposed. Let the bias voltage E be set to voltage point 2 and assume the circuit has stabilized at b. The emitter current will be very low, as will the collector current. Suppose that a signal is injected into the emitter that temporarily raises the operating point from b to above the peak point. The emitter current and voltage will follow a trajectory in the v-i plane ending up at point b'. The transistor will remain in its stable high-current state at point b', and the collector as well as emitter current will be quite high. With typical M1734 transistors, the time taken by the transistor to go from the peak point to point b' is of the order of a few hundredths of a microsecond.

If the transistor is locked up in its high-current state and a negative pulse is introduced which lowers the emitter below the valley point, the transistor will snap back to its low current state at b, where it will remain until it receives a positive trigger. This mode of operation could be used to develop a transistor flip-flop for use as an amplifier according to the scheme of Fig. 5. The circuit, however, would be unattractive. If low-voltage operation is desired and high-emitter currents are to be obtained, the load line from b to b' must have a slope of the order of a few hundred ohms. The point 2 must be well below the transistor peak point to allow for variations in the peak point from transistor to transistor. If this margin is 1 volt and the load line has a slope of 200 ohms, a triggering current of at least 5 ma would be required to raise the emitter above the peak point and the flip-flop would not have a very large current gain.

B. Broken Load-Line Input Circuit

To avoid high triggering currents, an input circuit has been developed, Fig. 6(a), which looks like a high impedance when the transistor is being triggered, and then changes to a low impedance once the transistor is far enough into the negative-resistance region for the emitter voltage to drop. If current is taken from the junction of the resistor and the diode, the voltage plot in the v-i plane will consist of a broken line. The slope of the characteristic is high and equal to the impedance of R_1 when the voltage is above E_2. The slope is very low and equal to the impedance of the conducting diode X_1 when the voltage is carried below E_2. Fig. 6(b) shows the circuit modified by the addition of another

resistor and diode. Consider the variation of voltage with injected current. If the input current is zero and

$$\frac{E_2 - E_3}{R_2} > \frac{E_1 - E_2}{R_1} , \qquad (11)$$

then the terminal 1 will rest at a voltage that is below E_2. The voltage difference will not be very great because it is equal to the forward drop across the conducting diodes X_1 and X_2.

Fig. 6—Development of input circuit. (a) Broken load line. (b) Input circuit.

Suppose that with no external current, the current flowing through X_1 into X_2 is represented by i_1. If a positive external current is supplied that is less than i_1, the voltage will change very little because the impedance seen at point 1 is the forward impedance of diodes X_1 and X_2. However, as the current is increased and made greater than i_1, diode X_1 will be cut off and the slope of the characteristic will increase and be approximately equal to the parallel impedance of R_1 and R_2, which can be made quite high. If E_1 and R_1 are made large and E_3 and R_2 are also made large, then the point 1 can be regarded as a current sink for a constant current i_1.

C. Triggering to High-Current States

Consider the emitter characteristic shown in Fig. 7(a) with the broken load line of diode X_1 and resistor R_1 superimposed upon it. If there is no input signal, the transistor will be biased below its peak point at point 1 and will be in its low-current state. The current flowing through diode X_1 into X_2 will have the value i_1. Suppose that current slightly less than i_1 is injected into the input. The input voltage will increase slightly, but not very much because the impedance is low for this current. If the current is increased beyond the value of i_1,

diode X_1 will cut off and the input voltage will rise rapidly with further increases in input current because the input impedance is high. If the input current continues to increase, the emitter terminal will rise above the peak point. When it rises above the peak point, positive emitter current will begin to flow. Because of the negative input resistance, the emitter will begin to go negative. When the emitter goes negative, diode X_2 cuts off. This isolates R_2 from the emitter circuit, and the emitter load is then determined by diode X_1 and resistor R_1, as in Fig. 6(a). As the capacity across the emitter discharges into it, the emitter voltage will fall until diode X_1 conducts and provides a continuous source of high-emitter current. The transistor will then lock up at point 2 of Fig. 7(a).

The above discussion of triggering has been based upon the assumption that the negative-resistance characteristic of the transistor lies entirely below the broken load line, as in Fig. 7(a). When the circuit is designed for minimum triggering current, then the load line provided by R_1 intersects the transistor characteristic at a voltage more positive than the voltage E_2 to which diode X_1 is returned. If this intersection were stable, the current gain of the amplifier would be negligible. To insure that the intersection will not be stable, it is only necessary that the inequality of (9) be fulfilled. Under these conditions the emitter trajectory of the v-i plane can be expected to spiral outward around the bias point provided by R_1. Once this trajectory falls below the voltage E_2, the diode X_1 will conduct and provide the low-impedance load line which permits lock-up at point 2 of Fig. 7(a). It can be seen that the only function of the transistor during triggering is to carry the emitter below the voltage E_2 so that the load line is

Fig. 7 (a) and (b).—Regenerative amplifier.

switched from the high impedance R_1 to the low impedance provided by diode X_1 when it conducts. Since the emitter current is high at point 2, the collector current will be correspondingly high and the collector voltage will have risen towards ground because of the drop produced across resistor R_4. Note that besides the amplification of emitter current by α_e, an additional and greater current gain is produced by the

negative resistance carrying the emitter current from the triggering value to the lock-up value.

D. Return to Low-Current State

It has been shown that a small triggering current can be used to produce a very large current at the collector. The large current will maintain itself until the transistor is returned to its low-current state. The return can be obtained by supplying a negative signal at the emitter. However, the source turning off the transistor would have to work into the low impedance of the saturated transistor and considerable power would be required to drag the emitter below the valley. The turnoff can also be obtained by raising the base positive since this has the same effect as lowering the emitter.

To compute the base voltage required to reset the transistor to its low-current state, consider the equivalent circuit of Fig. 1(a) substituted for the transistor in the amplifier circuit, Fig. 7(b). By including the resistance of the conducting diode X_1 in r_e, the circuit can be analyzed as though the emitter terminal were held at the constant potential E_2 (Fig. 7(b)). As explained earlier, overload occurs because the voltage between the internal node and the collector terminal falls to some minimum value V_s. The base signal can only reset the transistor to its low-current state by raising the internal node of the transistor to a voltage greater than

$$E_{cc} + i_c R_L + V_s, \qquad (12)$$

when the transistor will become active. The flow of emitter current establishes the internal node at

$$E_2 - i_e r_e. \qquad (13)$$

When (13) is greater than or equal to (12), then the transistor can be assumed to be in its active region, where

$$\left. \begin{aligned} i_e &= \frac{i_b}{\alpha_c' - 1} \\ i_c &= \frac{\alpha_c'}{\alpha_c' - 1} i_b \end{aligned} \right\}. \qquad (14)$$

The parameter

$$\alpha_c' = \frac{r_m - r_e}{r_c} \qquad (15)$$

is the ratio of collector to emitter current when the base terminal is driven by a generator and is of the same magnitude as α_c. When (12) and (13) are equated and the above relations are substituted, it is found that the base current required to reset the transistor is

$$I_{bs} = (\alpha_c' - 1) \frac{E_2 - E_{cc} - V_s}{r_e + \alpha_c' R_L}. \qquad (16)$$

The base voltage will be more positive than the internal node potential by $i_b r_b$.

From (16), (14), and (13)

$$E_{bs} = E_2 + \frac{(E_2 - E_{cc} - V_s)}{\left(1 + \alpha_c' \dfrac{R_L}{r_e}\right)} \left[\frac{r_b}{r_e}(\alpha_c' - 1) - 1\right], \qquad (17)$$

where E_{bs} is the voltage required to take the transistor out of overload. This equation is subject to all the limitations implied for (5), but it has proven useful. To relate the emitter and base characteristics it may be of interest to consider the impedance seen looking into the base terminal.

$$r' = \frac{r_b(\alpha_c' - 1) - r_e}{\alpha_c' - 1}. \qquad (18)$$

Comparison of (18) with (3) shows that the base impedance may be positive when the emitter impedance is negative. In the gain region, where α_c and α_c' are greater than one, the impedance seen at the emitter terminal with the base grounded will be negative as from point a to point b of Fig. 2(a). The impedance seen at the base with the emitter grounded will be positive as in the region b to c of Fig. 2(c).

The statements made will have to be modified slightly with regard to Fig. 7(b). When applying (3) to this circuit, the resistor R_3 must be added to r_b. This is not so when applying (18) because the resistor R_3 is then in parallel with the base. Correspondingly, when applying (18) the emitter impedance must be increased by the impedance of the conducting x_1 diode since it is a part of the emitter path.

A numerical example based on the values of (6) is of interest. It will be assumed that the value for r_b shown in (6) included an external base resistance of 500 ohms. The base impedance to use in (18) is therefore 1,500 ohms. It will also be assumed that E_2 is −1 volt and that the conducting diode X_1 (Fig. 7(b)) adds 125 ohms to r_e to make the total emitter impedance 250 ohms. With the assumption that the collector impedance is 20,000 ohms, it can be shown that $\alpha_c' = 2.09$. Then (18) gives an impedance at the base terminal (neglecting the parallel resistance R_3) of +1,280 ohms as compared to the −1,875 ohms that would be seen at the emitter. The base voltage required to raise the transistor out of overload will be 5.6 volts, whereas the valley voltage was $-4\frac{3}{8}$ volts.

E. Waveforms in Amplifier

If a regular succession of high-amplitude clock pulses are fed to the base through the diode X_3, the amplifier can only go to its high-current state provided both an input signal is present and in addition the diode X_3 has been cut off by the clock voltage falling below ground. The function of the signal fed to the emitter is merely to set the stage so that when the base voltage falls to ground the circuit will trigger. After the lock up the transistor will be returned to the low-current state

when the clock voltage rises sufficiently positive. Thus both the end of the pulse and the beginning are under the control of the clock. Fig. 8 shows a photograph of actual waveforms in the regenerative amplifier when it is regenerating pulses at a megacycle rate. Note that the rise times of the pulses are a few hundredths of a microsecond. All output pulses have the same duration and amplitude. The large signal at the base has no effect on the transistor unless there has been a signal at the emitter. It may also be noted that the collector voltage during the ON state rises very close to the emitter and base voltages.

(a)

(b)

8 VOLTS

BASE
EMITTER
GROUND
COLLECTOR
- ½ μ SEC - -
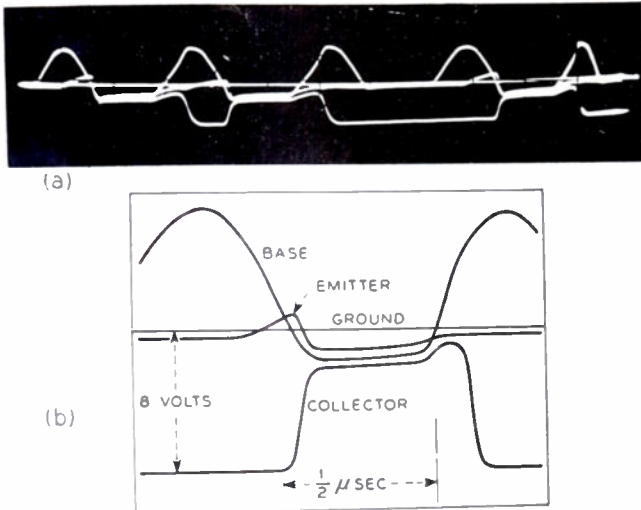
Fig. 8 (a) and (b)—Waveforms in amplifier.

The positive pip on the trailing edge of the collector pulse may be explained in terms of hole-storage in the transistor. When the transistor is locked in the high-current state, it has ceased to be an active element. The emitter diode is biased in the forward direction and is carrying a large current. The collector diode is biased in the back direction, but its impedance has been broken down by the charge carriers (holes) made available to it. The signal at the base of the transistor cannot return the transistor to its low-current state, except by making the transistor an active element again. As long as the collector impedance remains saturated, the transistor remains passive. Therefore, even though the clock voltage may raise the base terminal to the voltage $E_{b_\bullet}$ of (17), the transistor will not return to its low-current state until the stored charge carriers have been pulled out of the germanium. This process may take a tenth of a microsecond or longer, depending upon the length of time the transistor has been saturated. The positive pip on the trailing edge of the collector wave-form represents the attenuated clock signal as coupled to the load through the low impedance of the saturated collector. Note that once the stored carriers have been removed the transistor returns rapidly to its low-current state.

The phase of the clock voltage applied to the base is selected to lag the input signal by one-quarter microsecond in order that the triggering current will have

time to charge the capacitor at the emitter to a voltage above the peak point. The result is that the onset of the output is under control of the clock. The amplifier output is always delayed, but it always occurs at the same time in the clock cycle even though there may be slight variations in the time of arrival of the input signal.

III. ILLUSTRATIVE DESIGN

A. Input Requirements

The following discussion will be based upon the illustrative design shown in Fig. 9, which is the circuit in which the waveforms shown in Fig. 8 were developed. It is hoped that the discussion will be complete enough to permit the reader to produce variations upon the design to meet his own requirements. The usable gain of such a regenerative circuit is set in part by the margins which must be allowed to insure that nonexistent signals will not be regenerated. The output obtained when these margins are overcome is largely a matter of the load impedance into which the circuit works and the magnitude of the collector voltage supply. In the computer operations for which this amplifier was intended there is no inherent value in power output as such. Each amplifier merely drives some logic circuits, which in turn drive other amplifiers. If a reduction in the collector supply voltage reduces the driving power as much as it does the power output, there is no reason why the collector voltage should not be reduced. The first step in designing a specific amplifier is to pick the lowest collector voltage at which efficient operation of the circuit can be obtained.
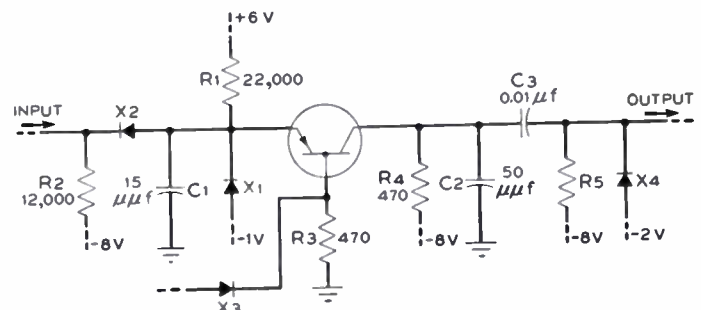
Fig. 9—Illustrative design.

The voltage drive required is determined primarily by the statistical distribution of peak points of the transistors that are to be used in the circuit. The voltage E_2, to which diode X_1 is returned, is chosen so that with zero emitter current, the emitter will rest at a voltage less than the peak point. If this requirement is not met, the circuit will have no stable low-current state. If transistors can be produced to have peak points that are very stable with age and are very close together for different units, E_2 can be set quite close to the average peak. Even if there is no spread in peak points, however, some voltage drive will be required. In the low-current state the emitter does not rest at the voltage E_2, but is

depressed below that voltage by the drop across the diode X_1. This drop will be a variable from circuit to circuit, and the voltage E_2 must be set far enough negative to compensate for these variations. In addition, there will be a voltage drop across diode X_2, so that the input terminal of the amplifier rests farther negative than the emitter. Thus the input signal must not only overcome the margin on the peak point, but must also overcome the drop across diode X_2. These considerations place a premium on low-impedance diodes for use as X_1 and X_2 as well as on a narrow spread of peak points. It will be demonstrated in a later paragraph that the transient back current of diode X_3 also effects the margin which must be allowed on the peak point. With M1734 transistors and Western Electric 400A diodes in the circuit of Fig. 9, it has been practical to use a voltage drive of two volts.

The current required to trigger the circuit will be at least as great as the emitter current at the peak point. This is approximately 50 μa.

When the base voltage falls to the ground, successful triggering to the high-current state requires that the transistor be unstable (See IIc). This can be assured by making the product of the transistor bandwidth, shunt capacity, and negative resistance exceed some minimum value (see (9)). If the product is not large enough, then the emitter current will settle down to the small value supplied by resistor R_1. To produce a high value of $-r$ and f_0, the triggering current may have to carry operation far into the gain region. To attempt to lower the required value of $-r$ or f_0 by increasing C delays the operation of the circuit by requiring a longer time for the triggering current to charge C above the peak point. Another possibility for reducing the required triggering current is to raise the value of the external base resistor R_3. This, however, has the unfortunate effect of increasing the required voltage drive. As explained earlier, the principal variation in peak point is caused by variations in the collector current for zero emitter current. Any increase in the base impedances magnifies the effect of this current.

Because of these considerations, amplifiers which are to regenerate pulses occurring at a high rate must pay the penalty of requiring increased triggering current, and therefore have less gain. With the M1734 transistor in the circuit of Fig. 9, a negative resistance and a bandwidth sufficiently high to cause triggering with a shunt capacity of 15 $\mu\mu f$ can be developed for an emitter current of about 300 μa. This current can produce a voltage change of 2 volts across 15 $\mu\mu f$ in 0.1 μsec.

B. Effects of Diodes on Operation

Another factor which has a major influence on the required triggering current and response time is the dynamic reverse current[5] of diode X_1. A diode exhibits

[5] L. A. Meacham and S. E. Michaels, "Observations of the rapid withdrawal of stored holes from germanium transistors and varistors," *Phys. Rev.*, vol. 38, no. 2, pp. 175–176; April 15, 1950.

a low impedance for a current in one direction and exhibits a high impedance for current in the reverse direction. However, if forward current has been flowing and the voltage is reversed rapidly, the current through the diode reverses rapidly but takes some time to fall to the low back current that is indicated by dc measurements. The transient in the back current that follows a reversal of voltage has been called the "enhanced current." It is a function of the magnitude and the duration of the positive current that has been flowing. The enhanced current appears to result from the storage of charge carriers. In the point-contact diodes the back current 0.1 μsec after a reversal of applied voltage may be several hundred times the back current indicated by dc measurements.

Enhancement affects the triggering of the amplifier in the following manner: As the input to the transistor amplifier rises, and diode X_2 cuts off, diode X_1 will have only a small reverse voltage on it. The enhanced back current after a few tenths of a microsecond may still be several hundred microamperes. Since this current must be supplied through the resistor R_1, the entire triggering current may flow through diode X_1 in the back direction rather than into charging the shunt capacity or into the emitter. This adds a further delay to the triggering of the transistor and increases the triggering current required for high-speed operation. With a current of 300 μa through R_1, it has been found practical to supply the enhanced back current of diode X_1 (when selected Western Electric 400A diodes are used), charge up 15 $\mu\mu f$ shunt capacity, and carry the emitter sufficiently far into the gain region to trigger a M1734 transistor to its high-current state within less than a quarter microsecond of the rise of the input signal.

It may be noted that the enhanced back current of diode X_2 is in a direction which helps to trigger the transistor. However, the enhanced current is a function of the forward current that has been flowing. Since the transistor locks up on the current drawn through diode X_1, diode X_1 carries a much larger forward current than diode X_2. Therefore, the enhanced back current of diode X_2 contributes negligible compensation for the enhanced back current of diode X_1.

It should be mentioned that the diode X_3 can cause false triggering because of its enhanced back current. When the clock goes below ground, the enhanced back current of the diode may carry the base sufficiently below ground to trigger the transistor even though its emitter is below the peak point for zero volts on the base. It has been necessary to allow for this effect in choosing the bias E_2.

C. Design of Input Circuit

Proper triggering is obtained by choosing a triggering current that will clear out the charge stored in diode X_1, charge up the shunt capacity in the time available, and establish emitter current which will supply a value of $-rf_0$ large enough to satisfy the inequality (9). Having

chosen the triggering current, the voltage E_1 and the resistor R_1 can be chosen. E_1 should be chosen so that it is many times larger than the variation in voltage to be developed at the emitter when going from the low-current state to the triggering region. If this is done, then the resistor R_1 will look like a constant current source, and all transistors will get the same triggering current regardless of the variations in their peak voltage points.

In the illustrative design E_1 was set at 6 volts and R_1 was chosen to be 22,000 ohms to give an available triggering current of about 300 μa. R_2 and E_3 should be chosen so that in the low-current state of the transistor, diode X_1 will be required to pass current in the forward direction to bring the emitter terminal up close to the clamping voltage E_2. The current that flows through diode X_1 should be set to be an appreciable fraction of the triggering current. In the illustrative example (Fig. 9), E_3 was set a -8 volts and R_2 was set at 12,000 ohms. With average Western Electric 400A diodes used for X_1 and X_2, the emitter will rest at -1.2 volts, the voltage to the input terminal will be -1.45 volts, the current through R_1 will be about 340 μa, the current through diode X_1 will be about 200 μa, and the current through diode X_2 into R_2 will be about 540 μa. Assuming that all transistors have peak points below ground, the driving voltage required would be 1.45 volts plus the 0.25 drop across diode X_2, to give a total of 1.7 volts. The driving current required is 8.25 volts divided by 12,000 ohms, and is approximately 670 μa.

The current through diode X_1 in the low-current state was set to 200 μa, so that the resistors R_1 and R_2 and the supply voltages could vary over a range of values without producing significant changes in the voltage at the emitter. If the voltages E_1 and E_3 had been made larger, and the resistors readjusted to keep the current through R_1 and the diode X_1 constant, it would have been possible to have produced a design in which the triggering current is only slightly greater than 540 μa. This, however, would have increased the power wasted in resistors R_1 and R_2.

D. Output Circuit

As mentioned earlier, the power output is determined primarily by the collector voltage and the load impedance. In the illustrative design, the collector voltage was set at -8 volts. If it had been made more negative, more power output could have been obtained, but the peak voltage would have gone more negative and a larger drive would have been required. Similarly, the collector voltage could have been made smaller and the voltage drive decreased. However, if the collector voltage is made much smaller than 8 volts for the M1734 transistor, its efficiency in this circuit would have been decreased markedly.

Since the amplifier is to be used to drive diode logic circuits, which in turn drive other amplifiers, the dc level has to be restored at some point or other. This change is generally obtained by means of the condenser C_3 of Fig. 9, which works into the diode X_4 for dc restoration.

If a single positive-going pulse is developed at the collector, the voltage change across the diode X_4 can be made almost as large as the collector pulse by choosing C_3 so big that the output current does not discharge it appreciably during the pulse. There will always be a slight change in the voltage across the condenser; and when the collector falls, it will not fall to the supply voltage but will differ from it by the lost voltage. This small voltage will produce a current through R_4 and the low forward impedance of diode X_4 which, in time, will recharge C_3 and restore the collector voltage to the supply voltage. If C_3 is made large enough, a few successive pulses at the output will all have the same amplitude. However, as the number of successive pulses is increased, the amplitude will drop no matter how big C_3 is. Suppose that during the positive part of the pulse a current of 4 ma flows through C_3 and resistor R_5. If the train of pulses is continuous and symmetrical, that is, the transistor conducts one-half the time and is shut off one-half the time, 4 ma must flow back through the condenser in a reverse direction during the transistor off period. This current produces a voltage drop across the resistor R_4, and when the next pulse comes along and triggers the transistor, the collector voltage does not start at the normal supply voltage but starts from a more positive value because of the drop produced across R_4 by the charging current.

The fact that the voltage at the collector terminal is not fixed but varies according to the duty cycle of the signal complicates the problem of providing adequate margins. The most positive peak voltage will occur, for example, when the input is a continuous chain of pulses while the most negative peak voltage will occur when the signal consists of a single isolated pulse. Thus a defective transistor may operate satisfactorily for one signal but not another. These considerations lead to very low values of R_4. A lower bound for R_4 is brought about by two factors. One of these is the power dissipation in the transistor which goes up as the value of R_4 goes down. It is generally assumed that this dissipation should be kept less than 100 mw. A more limiting factor is the influence of R_4 upon the turnoff characteristics of a transistor.

E. Effect of Load on Return to Low-Current State

Equation (17) gives the base voltage (E_{bs}) required to reset the transistor. With the set of parameter values (6) modified as explained with reference to (17), it was shown that E_{bs} is $+5.6$ volts. If the load impedance, R_L, were reduced from 500 to 125 ohms, (16) shows that E_{bs} would be $+15.8$ volts. The power dissipated in the transistor during the hole-storage period would then be dangerously high.

The clock current required to reset the transistor is set in part by the value of R_3, but mainly by the hole-

storage characteristics of the transistor. In the illustrative circuit the clock may have to supply 12 ma through resistor R_3 to raise the base to $+6$ volts, and may have to supply an additional current of 20 ma to clear out the charge stored in the transistor. This is a serious handicap, though not a disabling one. Since the clock signal contains no information, it can be developed by a high-power steady-state oscillator. In the applications where this circuit has been used, it is customary to supply the clock power for the entire computer from a 1-mc power oscillator.

The choice of appropriate output circuit parameters may be summarized in the following manner: The resistor R_4 is made high enough to permit reasonable clock voltages to shut off the transistor without excessive dissipation in the transistor. Then the output current to be supplied through C_3 is limited to a value small enough for the drop it produces across R_4 to leave sufficient voltage drive for the load when the signal is a continuous chain of pulses. In the illustrative design, R_4 was set to 470 ohms which makes it possible to shut off M1734 transistors by raising the base terminal to $+6$ volts. The loads used, represented by R_5, have been limited in their current drain to less than 4 ma. This current produces a voltage drop of about 2 volts across R_4. Under full load the amplitude of a continuous chain of pulses is therefore 2 volts less than the amplitude of an isolated pulse (see Fig. 15).

Experience has shown that reliable operation can be obtained in the illustrative circuit, provided transistors and diodes are used which have sufficient margins on the peak voltage, required triggering current, negative resistance, bandwidth, base impedance, current gain, and transient back-current characteristics. A year's experimental work with an average of over 50 amplifiers in daily use has shown that transistors and diodes can be made which have the required margins.

IV. APPLICATION OF REGENERATIVE AMPLIFIER TO DIGITAL COMPUTERS

A. Basic Functions

Digital computer operations can be divided into two classes, memory and logic. Memory can be defined as the representation in space of a function of time. In computers, logic operations consist of the recognition of spatial distributions of voltages and currents.

Memory can be obtained by circulating information in a passive delay line which has regenerative amplifiers in series with its sections to maintain the signal level. Logic operations can be mechanized with such passive nonlinear elements as semiconductor diodes.[6] It is possible, therefore, to build complete transistor computers in which the only transistor circuit is the regenerative amplifier that has been described in this paper.

Fig. 10 shows a complete set of building blocks that have been developed for use in serial computers operat-

ing at a megacycle pulse rate.[7] The OR-circuit (Fig. 10(a)) will have an output if there is a signal on any of the input leads. The AND-circuit (Fig. 10(b)) will have an output only if there is an input signal on all of its input leads. The signal on the lower terminal of the INHIBITION circuit (Fig. 10(c)) will always go through the circuit unless there is an inhibition signal, in which case the circuit has no output at all. The inhibiting signal lead is schematically distinguished by a small button.
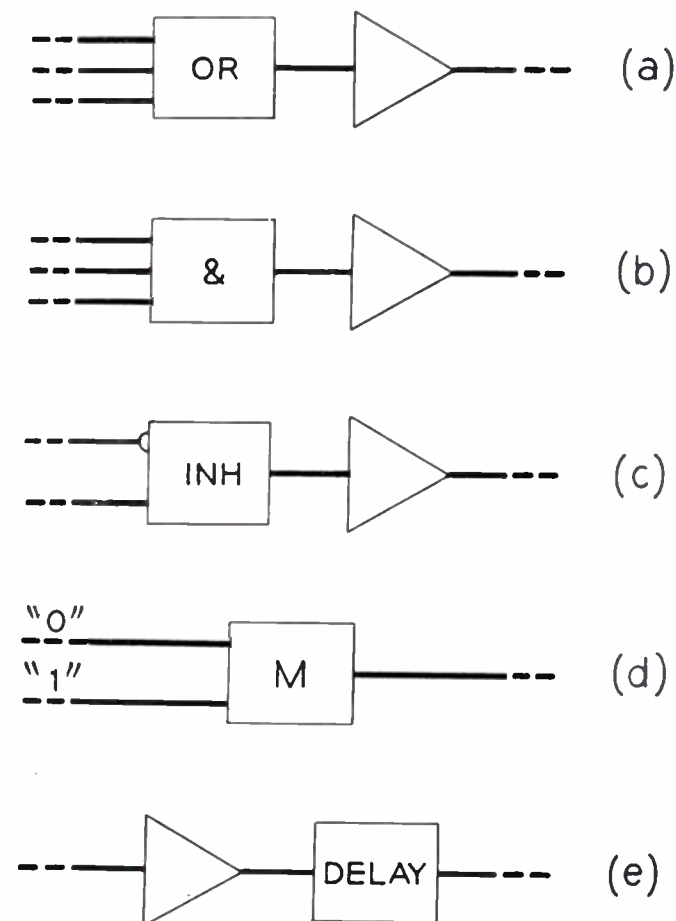
Fig. 10—Set of computer functions. (a) OR—circuit package. (b) AND—circuit package. (c) Inhibitor—circuit package. (d) Bit—storage-cell package. (e) Delay package.

B. Logic Packages

Fig. 11(a) shows a three-terminal OR-circuit combined with the basic amplifier. The only additional elements required besides those shown in Fig. 9 are the diodes $X6$, $X7$, and $X8$. If any of the inputs are carried above ground, the voltage across $R2$ will be carried above ground, diode $X1$ will cut off, and the transistor will fire. The diodes $X6$, $X7$, and $X8$ have the function of insuring that the current from any one input will flow into $R2$ rather than back into the other inputs.

A three-terminal AND-circuit can be mechanized, as in Fig. 11(b), where the additional elements required are $X6$, $X7$, $R6$, and $R7$. Since the transistor can only be

[6] Tung Chang Chen, "Diode coincidence and mixing circuits in digital computation," PROC. I.R.E., vol. 38, pp. 511–514; May, 1950.

[7] D. R. Hartree, "Calculating Instruments and Machines," University of Illinois Press, Urbana, Ill., pp. 98–104; 1949.

triggered to its high-current state by raising the emitter terminal above the peak point, the transistor will not fire in the circuit 11(b) unless the three diodes $X2$, $X6$, and $X7$ are all cut off. The absence of a signal at any of the three input terminals permits the corresponding diode to conduct continuously and clamp the emitter below the peak point. Diode enhancement properties react on this circuit in several ways. The current through diode $X1$, when the transistor is in its low-current state, is about 1,240 μa. Without the diodes $X6$ and $X7$ the

Fig. 11—Logic circuits added to amplifier. (a) Amplifier including 3-terminal OR circuit. (b) Amplifier with 3-terminal and circuit. (c) Amplifier with inhibition.

corresponding current is only 200 μa. The 1,240-μa current flows through diode $X1$ up to the time the diode is cut off, and enhances its back current. This increased back current may make triggering difficult. The enhanced back current of diodes $X6$, $X7$, and $X2$ may also cause trouble. If diode $X2$, for example, is the only one conducting in the forward direction, the enhanced back current of $X6$ and $X7$, if large enough, will develop a voltage across $R2$ that carries the emitter above the peak point. Neither of these enhancement problems have interfered with successful operation of the circuit of Fig. 11(b). If the number AND-terminals were increased, then these effects could be serious.

The inhibition circuit of Fig. 10(c) has been mechanized with the circuit of Fig. 11(c) which is fed by two simultaneous positive inputs. The inhibiting signal is inverted by the input transformer; when present, it carries diode $X5$ below ground, with the result that the emitter of the transistor is clamped below the peak point. Whenever the inhibiting signal is present, the transistor cannot fire even though the other signal may have cut-off diode $X2$. If the inhibiting signal is not present, then a positive signal across $R2$ will cut-off

diode $X2$ and trigger the circuit because diode $X5$ will not conduct until the emitter is above its peak point. A four-phase clock which generates four sine waves 90° out of phase is used in these applications. The relative phases of signals in Fig. 11(c) are important. The two input signals must be in phase; that is, they must come from amplifiers whose base signals were in step. If they were shut-off by phase-a of the clock, then the inhibiting amplifier they feed must operate on phase-b of the clock signal, where phase-b lags phase-a by one-quarter of a microsecond. To obtain successful inhibition, the inhibiting effect must be made to overlap both the rise and the fall of the signal to be inhibited. The fact that the transistor cannot fire until the phase-b clock has fallen to ground, by which time both signals on the input of the amplifier will have had a quarter of a period to establish themselves, makes it unnecessary for the inhibiting signal to arrive early. The phase-c clock signal fed to diode $X7$ holds the emitter below the peak point after the first quarter microsecond that the base has been at ground. Therefore, even though the signal to be inhibited may last longer than the signal which is to inhibit it, because of hole storage, no false operation of the amplifier results.

C. Memory Packages

The memory cell shown in Fig. 10(d) operates in the following manner: If a pulse is put on the "1" terminal, the circuit circulates the pulse and makes it available once every microsecond until it has been removed by the application of a pulse on the "0" terminal. The realization of the cell is shown in Fig. 12. If a half microsecond pulse is put on the "1" lead, it goes through the OR-circuit, the INHIBITION-circuit, the amplifier, and back through the three-quarter microsecond delay
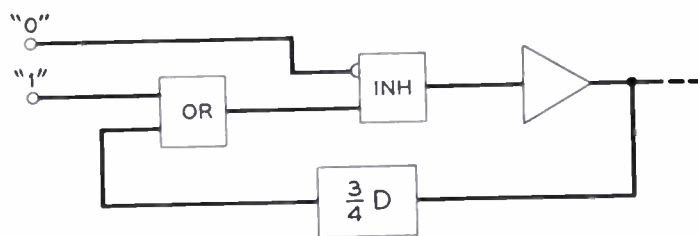
Fig. 12—Memory-cell block diagram.

line to the second terminal of the OR-circuit. The clock signal at the base of the transistor amplifier keeps the stored pulse in synchronism with the rest of the computer. If a signal is fed to the "0" terminal, it will inhibit the circulating pulse and the stored pulse is erased forever. The actual amplifier circuit is the same as in Fig. 11(c), except that two diodes are connected to resistor $R2$ to make up the OR-circuit. This is the form of memory that has been used as a substitute for leaving the bistable flip-flop locked up indefinitely to store a one.

The delay package, Fig. 10(e), consists of the standard amplifier followed by electric delay line capable of

storing several digits. It has been customary to use an amplifier after every 5 μsec of delay.

D. System Examples

The combination of these building blocks in a computing system is illustrated in Fig. 13, which is the block diagram of a serial binary adder that requires 11 standard amplifiers. The combinations 1-2-3 and 4-5-6 each
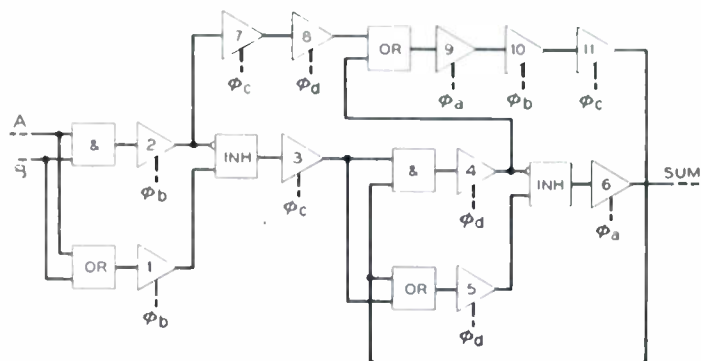
Fig. 13—Binary-adder block diagram.

consist of a half-adder.[7] The two amplifiers 7 and 8 are inserted to introduce a half microsecond of delay into the output of amplifier 2 so that it will be in phase with the output of amplifier 4 when it enters the OR-circuit feeding amplifier 9. Note that amplifiers 10 and 11 also serve only as delay units. It would have been possible to use passive delay lines to replace the amplifiers, 7, 8, 10, and 11, but the use of the standard amplifier results in a much more compact assembly.

Fig. 14—Multiplier and control unit.

The adder shown operates from a total dc power supply of less than one-half watt and is a part of the arithmetic unit shown in Fig. 14. The arithmetic unit has 38 amplifier packages, and can multiply two 16-digit binary numbers together in 272 μsec. The control unit on the left of the photograph contains 35 amplifier packages and develops the numbers which are handled

by the multiplier. It also provides test facilities. Each transistor package is mounted on a separate plug and all packages are interchangeable. It may be of interest that the dc power for the entire apparatus is obtained from a power supply using semiconductor rectifiers.

Fig. 15—A switching operation.

Another application of the regenerative amplifier will be illustrated with respect to Fig. 15. The memory cell and AND-circuit act as a switch which remembers whether it is to stay open or closed. The first pulse on A closes the switch and lets the data on lead E through to D. Seventeen microseconds later the pulse on B empties the memory cell. However, the next digit period, a pulse on A closes the switch again. At a later time the signal on B turns off the memory cell and nothing gets through the AND-circuit for some time. The switch is in effect closed again by the next signal on the A terminal.

Those familiar with digital computers may recognize the combination of Fig. 15 as the part of a binary multiplier which connects the multiplicand into an accumulator under control of the successive digits of the multiplier.

Conclusions

Transistor performance equal to that of vacuum tubes can be obtained in computer applications without sacrificing any of the obvious transistor advantages, such as low-power operation, ruggedness, and extremely small size. The circuit discussed herein has proven to be a versatile one, and is being studied for a number of computer applications. It is expected that it will in time be superseded by more efficient circuits. However, it has played a part in demonstrating the practicality of the transistor for high-speed switching and computing applications.

Acknowledgments

A Transistor Shift Register and Serial Adder*

JAMES R. HARRIS†, MEMBER, IRE

Summary—A small set of basic functions, such as binary memory and elementary binary logic, can be remarkably versatile; such functions are important in switching and computing. This paper describes a piece of computing equipment which can store a pair of binary numbers and add them, producing the sum a digit at a time. The equipment is built from a basic set of functional blocks, all of which are designed around transistors. This set of building blocks consists of a binary cell, a pulse amplifier, a pulse amplifier with delay, and logic circuits. The binary cell is a flip-flop; amplifiers are monostable circuits, and logic is performed in diode gates. Some interesting special features arise from the use of transistors. These features are discussed and the designs are evaluated.

INTRODUCTION

THE WORK reported here was part of a study, sponsored by the Joint Services, to determine the feasibility of using transistors in miniature packaged switching-type circuits. We hoped to design several types of useful packages, each type performing a basic function, and each compatible with the rest. At the same time, we hoped to learn how to specify and use transistors to better advantage.

In order to have a concrete goal for development and evaluation of packages, the work was directed toward certain selected larger functions that are commonly found in computing equipment. No attempt was made to achieve high computing speeds.

One of the "larger functions" chosen as a goal is comprised of receiving and storing a pair of binary numbers, adding them, and producing the sum in serial form. This was realized by means of a pair of shift registers[1] and a serial adder. A shift register is a piece of equipment which can receive a set of digits in time sequence or in parallel, hold the digits indefinitely, and move the digits along in response to a signal. A serial adder acts to produce proper sum digits when numbers to be added are fed into it.

It was found that shift registers and an adder, in common with most computing equipment, could be built from basic building blocks having these four functions: information storage (one binary digit), level restoration (gain), delay, and logic. A transistor or diode package was developed for each of these functions.

The plan of this paper is to describe first the general functioning of this system of storing and adding numbers, then the way in which the shift register and serial adder were designed from packages, and finally, the transistor circuits of the functional packages. Some of the circuit processes are explained in terms of the "Emitter V-I plot." The emitter can show an open-circuit stable negative resistance which, with a load

line, tells a great deal about a circuit. The reader is referred to an article by Anderson[2] for a full discussion of this method of analysis.

SYSTEM OPERATION

There is a wide choice of methods in storing and adding binary numbers, and we evaluated several against our requirements. In the system we adopted (Fig. 1), one digit of the sum is produced in response to each initiation pulse; initiation pulses can be applied at any rate whatever up to a maximum. The flip-flop type of circuit was chosen as the primary means of storing digits. Pulse delay is provided by active circuits, and the generation of sum and carry digits is done in diode circuits.
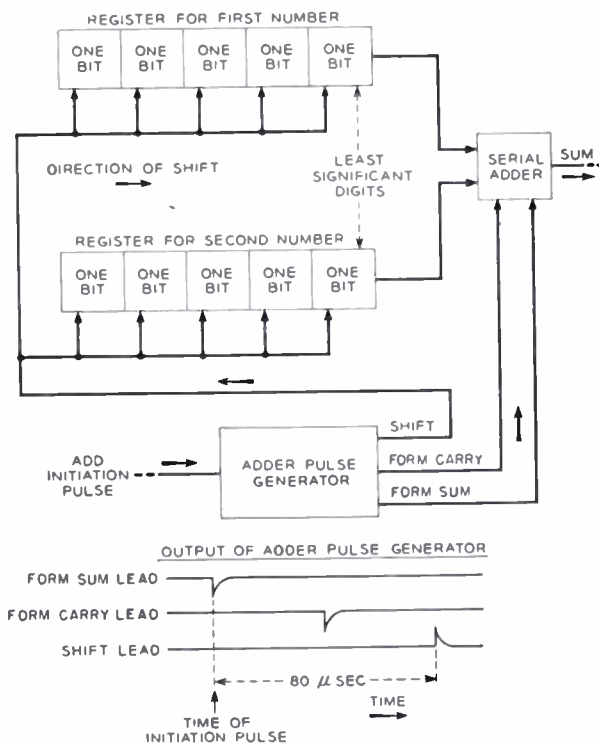
Fig. 1—This block diagram shows shift registers for a pair of numbers, a serial adder, and a pulse generator to control operations.

The arrangement in Fig. 1 acts very much like a person writing a pair of numbers on paper and then adding them. First the numbers are written into the registers. Then the following steps are repeated, starting with the least significant place:

1. Find the digit of the sum. (The least significant is found first.)
2. Find the carry digit and remember it. (In adding just two binary numbers, there is only one carry digit.)

[1] C. F. West and J. E. DeTurk, "A digital computer for scientific applications," PROC. I.R.E., vol. 36, p. 1452; December, 1948.
[2] A. E. Anderson, "Transistors in switching circuits," vol. 40, pp. 1541–1559; this issue.

3. Change the attention to the next more significant pair of digits. In the machine this is done by a rightward shift of the stored numbers. (Each digit moves one step to the right, and the rightmost pair, which is no longer needed, is lost.)

On paper or in the machine, the sum digit and carry digit are found by looking at three known digits (one of which is the old carry digit) and applying the rules of addition.

Shift Register

The shift register is shown in Fig. 2. Digits are stored in binary cells; our binary cell is a symmetrical flip-flop (Fig. 5) with a pair of inputs and a pair of outputs. The two inputs are used to "Set 0" and "Set 1."
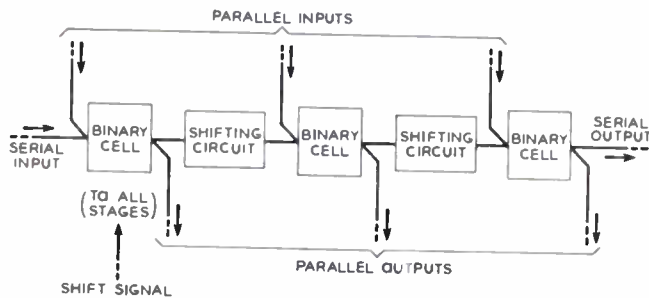
Fig. 2—The functional diagram of a shift register (a 3-bit register is shown).

The shifting method used is one of many, not necessarily the best. An amplifier-delay unit is used between cells as a shifting circuit. The function of this circuit is to provide a delayed pulse (rather than the memory of a state). The circuit can handle only one pulse at a time.

To shift, a pulse is applied to the "Set 0" inputs of all cells in the registers. Those cells which contained 1 switch to 0; the rest do not change. The transition from 1 to 0 produces on one of the output terminals of the cell a negativeward voltage step. This terminal is capacitor coupled to the amplifier-delay, which is triggered by a negative step. The output of the delay unit sets the next cell to 1.

Adder

As noted under system operation, the adder must examine three digits and then produce a sum digit and a new carry digit determined by these three digits. The sum and carry must be formed according to the rules of binary addition. The pertinent rules are as follows: $0+0+0=0$ and carry 0; $0+0+1=1$ and carry 0; $0+1+1=0$ and carry 1; $1+1+1=1$ and carry 1. Since each of the three "input" digits can be either 0 or 1, there are eight distinct possibilities; these are shown as the eight rows of the truth table in Fig. 3. To put the requirements briefly, the adder must produce a sum digit and a new carry digit in accordance with the truth table in Fig. 3. The new carry digit should appear in the storage cell that contained the old carry.

In order to facilitate understanding the design of the adder, several paragraphs will be used to present some of the concepts of two-state theory together with some of the conventions used here.

It is convenient to think of the sum digit and the new carry digit as dependent variables. They are functions of three independent variables. Dependent variables can be produced by means of logic circuits which have the independent variables as inputs; logic circuits are realized here by means of diode gates.

For circuit reasons, the diode gates of the adder are designed to put out a short negative pulse as a signal. We have found it convenient to think of a signal as 1, and the absence of a signal as 0. As a result, the more negative voltage level is called 1, both at the input and output of these logic circuits.

NEGATIONS OF INDEPENDENT VARIABLES			INDEPENDENT VARIABLES			DEPENDENT VARIABLES	
A' = (DIGIT OF FIRST NUMBER)'	B' = (DIGIT OF SECOND NUMBER)'	C' = (OLD CARRY DIGIT)'	A = DIGIT OF FIRST NUMBER	B = DIGIT OF SECOND NUMBER	C = OLD CARRY DIGIT	C_N = NEW CARRY DIGIT	S = SUM DIGIT
1	1	1	0	0	0	0	0
1	1	0	0	0	1	0	1
1	0	0	0	1	1	1	0
1	0	1	0	1	0	0	1
0	0	1	1	1	0	1	0
0	0	0	1	1	1	1	1
0	1	0	1	0	1	1	0
0	1	1	1	0	0	0	1

Fig. 3—The rules for forming sum and carry digits in serial addition of two binary numbers are included in the five right-hand columns of this truth table. The dependent variables can be generated by "logic circuits." The three left-hand columns are useful in circuit design.

Logic circuits are much simplified if the "opposite" of each independent variable is available in addition to the variable itself. The opposite (the Boolean negation) is represented in tables and equations by a prime mark on the symbol for the variable.

A binary cell with symmetrical outputs provides both the stored digit and the negation. One output terminal always has a voltage representing the stored digit; the other always has a voltage representing the negation of the stored digit. Specifically, the terminal that is negative when the cell is storing 1 is called here the "normal" digit output, and the other terminal is called the "prime" output.

With this background, the adder design can be stated briefly. The adder must produce S, the sum digit. Ac-

cording to the truth table, S is a specific function of A, B, and C. One way to state this function is as follows: $S = 1$ when A and B' and $C' = 1$, or when A' and B and $C' = 1$, or when A' and B' and $C = 1$, or when A and B and $C = 1$. This statement is realized with complete correspondence by the arrangement of Fig. 4. The statement can be written in a more compact and useful form using Boolean algebra; thus,

$$S = AB'C' + A'BC' + A'B'C + ABC.$$

The translation into Boolean form uses the convention that "AND" means Boolean product and "OR" means Boolean sum. This convention must be used when a signal is defined as 1 and "AND" and "OR" are defined as shown in the section on logic circuits.

The carry was realized by an arrangement that was suggested by Hussey and proved to be economical of parts. He suggested that the truth table be interpreted as follows: "The carry must change (C_n different from C) only when $A = B = 1$ or when $A = B = 0$. Furthermore, when $A = B = 1$, $C_n = 1$; and when $A = B = 0$, $C_n = 0$." This was realized as shown in Fig. 4. The carry cell receives a "Set 0" pulse when A, B, $C = 0$, 0, 0 or when A, B, $C = 0$, 0, 1. The first of these is harmless; the second is necessary. Similarly, there is a "Set 1" pulse when A, B, $C = 1$, 1, 1 or when A, B, $C = 1$, 1, 0. Again the first of these is harmless, and the second is necessary.

Fig. 4 has been simplified by omitting (1) the coupling circuits of the amplifiers and (2) a special circuit that applies some of the output signal of each amplifier to the input of the other amplifier. The latter circuit disables one amplifier whenever the other is triggered and is needed because this amplifier can be triggered by a signal on its output terminal. If no precautions were taken, the idle amplifier would be triggered falsely by the voltage step that appears on the idle input terminal of the flip-flop when it changes state.
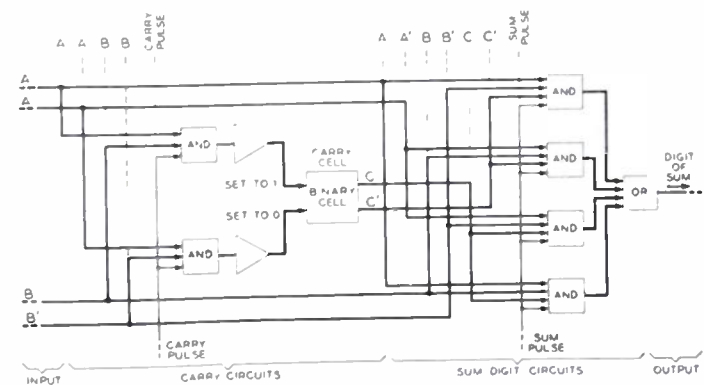
Fig. 4—A serial adder using 4 transistors.

BINARY CELL

The binary cell was realized in this system by means of a symmetrical flip-flop as shown in Fig. 5. Two point-contact switching-type transistors and seven resistors make the entire circuit.

One can question seriously the use of more than one transistor to make a flip-flop. These transistors have a large-signal current gain greater than unity, specifically, greater than 1.6. As a result, it is possible to make a bistable circuit with a single transistor. However, the symmetrical arrangement has some advantages. Logic circuits are simplified. Power-supply drain is independent of stored information, thus allowing very poor regulation of the power supply. Finally, there are some simplifications in circuit concept and design that were helpful in the exploratory work.
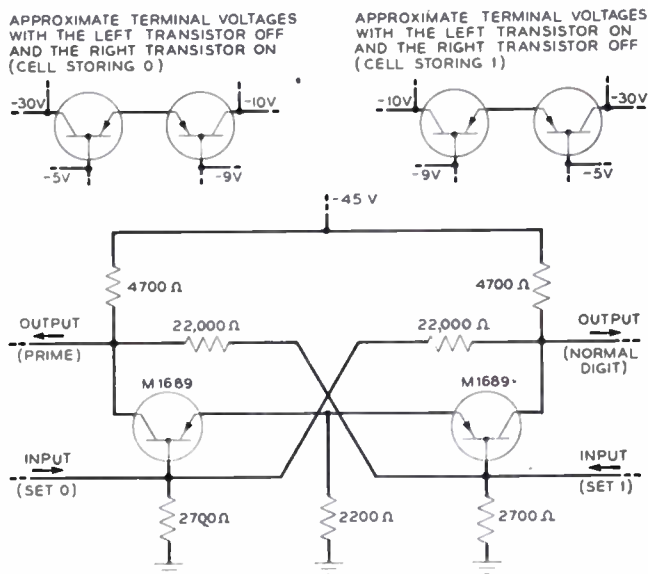
Fig. 5—This symmetrical flip-flop is used as a binary cell. To change state, apply a positive pulse to the proper input terminal. (A negative step applied to an input terminal is undesirable and can be prevented by using a diode in series with each input terminal.) The drawing labels in parentheses are based on these assumptions: (1) The state in which the left transistor is ON represents a stored 1. (2) At the output, the −30 volt level represents a 1.

Since transistors can be individually stable OFF (low current) or ON (high current), it is possible to have 4 stable states in the general two-transistor circuit. In this binary system, we want only the two states in which one is OFF and one is ON. The other two states are spurious and must be avoided. We avoid the OFF-OFF state by circuit design and the ON-ON state by a careful choice of triggering. A trigger signal always acts to turn OFF a transistor.

Emitter V-I plot analysis[2] is very useful in this circuit. The two transistors would have independent emitter V-I curves except for transmission through the 22,000 ohm resistors. This transmission can be taken into account by noting that its main effect is to produce a small voltagewise shift of an entire V-I curve. The load line of the plot is such that one transistor is always ON; in the absence of a trigger pulse, the emitter voltage of the ON unit is sufficiently negative to hold the other unit OFF.

These transistors have a remarkable property of slow switching from ON to OFF. In the grounded emitter connection, one expects an inversion of signal from input (base) to output. However, when a unit has been ON for a time, a positive step applied to the base appears as a positive step at the collector. It may require several microseconds to switch out of the saturation

region, the time depending on the transistor, the length of time saturated, and the switching current. This characteristic impairs the kind of operation usually found in vacuum-tube flip-flops and makes special measures necessary. In this case, the special measures are the use of emitter-to-emitter coupling together with a trigger arrangement wherein there is no dependence on the gain of the transistor that has been ON. The resulting circuit is reliable, but requires a large charge from the trigger source.

In terms of the present set of functions, a binary counter (scale-of-two counter) uses storage, logic, and delay. This flip-flop has been used with a pair of simple gates as a binary counter. In this case the delay need be only a brief memory of a state, and it is provided by the charge on a coupling capacitor.

Terminal characteristics of the flip-flop shown in Fig. 5 are approximately as follows:

DC power:

    8 to 10 ma, at −45 volts.

Input:

    Equivalent internal resistance and voltage:

       1,000 ohms and −9 volts or

       2,000 ohms and −5 volts.

    To change state:

       Apply a positive pulse to the −9 volt input.

    Pulse requirements:

       10 volts minimum; about 2×10^{-8} coulombs minimum.

Output:

    Equivalent internal resistance and voltage:

       1,500 ohms and −10 volts or

       4,000 ohms and −30 volts.

    Maximum permissible current in an output terminal:

       1 ma; more if the polarity of the current is favorable.

    Transition used as an output:

       The negativeward transition is preferable.

PULSE AMPLIFIER

In our data system, there is no need to preserve the original shape of a pulse when we amplify it. It is satisfactory to generate the pulse within the amplifier. This allows us to use a triggered monostable circuit as an amplifier. A simple and reliable single-transistor monostable circuit is available; it is shown in Fig. 6 with values chosen for the present application.

The circuit is normally OFF. When switched ON by a trigger signal, it can produce a peak power of a watt in a capacitor-coupled load. The amplitude of input signal at which the circuit triggers is under control of the circuit designer. With the values of Fig. 6, the circuit requires a step of 300 μa to trigger, a choice that gives good reliability. The required input power is estimated to be about a third of a milliwatt. This represents a "power gain" of 3,000, well beyond the gain of a similar transistor in a linear amplifier. In applications

where the resistor values can be adjusted, the circuit can be set to trigger on 10 μa or less.

The transistor switches OFF and the output pulse ends when the voltage on the emitter capacitor is no longer sufficiently positive to maintain the ON state. The relatively large capacitor used here allows a high-energy output pulse, but sets a limit on how often the circuit can be triggered. This limit is not important in the present system.
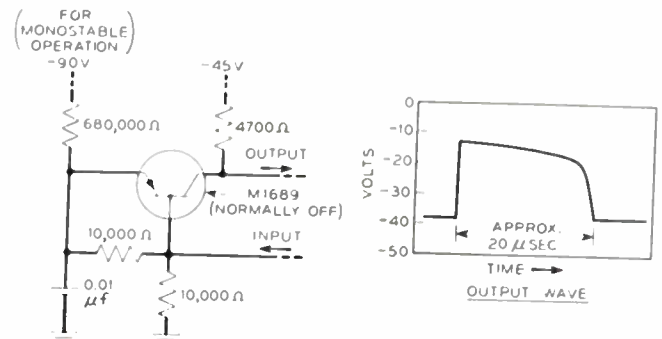
Fig. 6—This circuit can be made either monostable or free-running. The monostable circuit is used as a regenerative pulse amplifier. It is triggered by a negative step and is not sensitive to a positive step.

The output and input signal polarities of this amplifier are just what is needed by the binary cell that has been described. The output signal is a positive pulse, and the input signal required on the base is a negative step.

The switching times are remarkable. The total delay and rise time of the output signal is about 0.1 μsec. With a stronger than threshold trigger signal, many transistors approach 0.02 μsec. In contrast to this, the time required to switch OFF may be as much as 1 or 2 μsec unless the ON time is short.

In this application, a low internal output impedance in the ON state is desirable to develop a fast-rising, high-power output pulse. The output impedance when ON is essentially the emitter-to-collector path inside the transistor in series with the 0.01 microfarad emitter capacitor. When used in this way, the transistor is a good low-impedance switch, with a voltage drop of only 2 or 3 volts at currents in the range 10 to 100 ma. This is an important transistor feature that is not readily apparent from the usual static curves. The low impedance allows high power output with relatively low dissipation in the transistor. For a further discussion of efficiency, see "Transistors in Switching Circuits" in this issue.[2]

A shunt capacitance connected to the output increases positive feedback in switching ON and tends to speed the transition to high current. A remarkably high capacitance can be tolerated without affecting the output rise time. In one test, the voltage wave was little changed when 3,000 μμf was connected across the output. Peak current in this case must have been quite high, of the order of an ampere. It is not yet known whether present point-contact transistors are reliable with such high peak currents.

The terminal characteristics are more complex than those of the flip-flop, particularly in the ON state.

Output impedance in this state has been described. Input impedance in the ON state is low and variable and depends on the load. If this feature is troublesome, one can use a resistor in series with the input. The remaining characteristics of the circuit as shown in Fig. 6 are given approximately in the following table, which assumes that the 0.01-μf capacitor has negligible impedance at the frequencies of interest:

DC power:
    110 μa at -90 volts.
    1.5 ma at -45 volts; this increases
    with frequent triggering.
Input:
    Equivalent internal resistance and voltage in OFF
    state:
        4,000 ohms and -13 volts.
    Signal required to trigger:
        Negative current step of 300 μa or more, with
        rise time of 10 μsec or less.
    Maximum trigger rate for full sensitivity:
        600 per second.
Output:
    Equivalent internal resistance and voltage
    in OFF state:
        4,000 ohms and -38 volts.
    Signal to trigger on output terminal:
        A negative current step as little as 900 μa.
    Pulse voltage:
        23 volts.
    Pulse delay and rise time:
        0.02 to 0.1 μsec.
    Pulse duration:
        20 μsec; this is shortened by a load.
    Maximum available charge per pulse:
        0.1×10^{-6} coulombs.

Pulse-Amplifier Delay

The amplifier that has been described can be used as a delay unit by using the negative-going trailing edge of the output pulse as a signal. But this transition is rather slow, and the high internal impedance after the transition is a disadvantage. A fast and powerful delayed positive step can be generated with two pulse amplifiers in tandem, the second triggered from the trailing edge produced by the first. However, both of these arrangements have a weak point in this application because this pulse amplifier can be triggered by a negative step applied to its output.

A circuit which avoids these disadvantages is shown in Fig. 7. It is made by adding a simple normally ON monostable circuit following the pulse amplifier. Upon triggering, the base of the second transistor is driven positive. The collector current, which had flowed through the emitter, quickly switches to the base. The base-to-collector path in the transistor remains low resistance (a few hundred ohms) for 1 or 2 μsec. Then the collector resistance rises, the current falls to a low value, and the output terminal drifts negativeward as the capacitors charge. When the first transistor switches OFF, the second switches ON and the output terminal

snaps positiveward. The delayed positive pulse that is desired in this application can be obtained by "differentiating" the output wave (Fig. 7) with a series capacitor followed by a shunt-diode dc restorer.
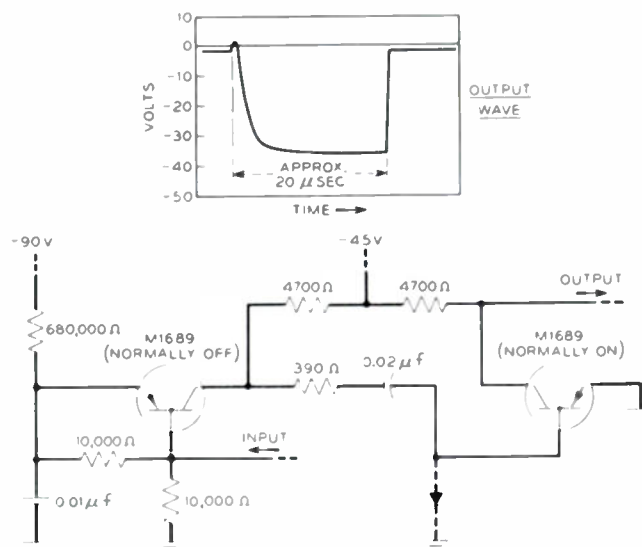
Fig. 7—This monostable circuit is used for delay and regeneration.

Since one transistor is normally ON, the current drain is relatively high. Slow switchoff causes the positive step applied to the base of the second transistor to appear on the load. This effect at the load can be greatly reduced, if desired, by a diode added as shown in Fig. 7. The diode also reduces delay time somewhat.

Input characteristics are given under pulse amplifier. The output has many of the characteristics of the pulse amplifier, especially the fast switch ON and the low internal impedance in the ON state.

Logic Circuits

Transistors and semiconductor diodes have many similarities, so that it seems natural to perform logic in this system with semiconductor diode gates. A complete set for two-state logic is "AND," "OR," and "NOT."[3] "NOT," which means opposite, is handled by the generation of the opposite in the binary cell. "AND" and "OR" are produced by diode gates as shown in Figs. 8
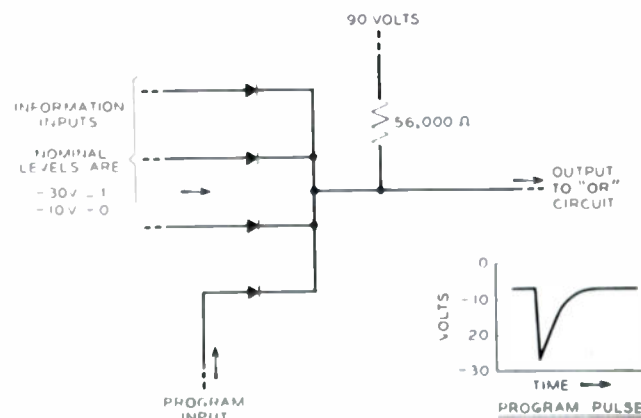
Fig. 8—A gate which acts as an AND circuit for negative signals.

[3] D. R. Hartree, "Calculating Instruments and Machines," University of Illinois Press, Urbana, Ill., p. 99; 1949.

and 9. The AND gate acts as follows: If (and only if) all inputs have signals (negative voltage), the output will have a signal (negative voltage). The output signal is produced by switching the bias current to the output. The OR gate follows this rule: If any input has a signal, the output will have a signal.
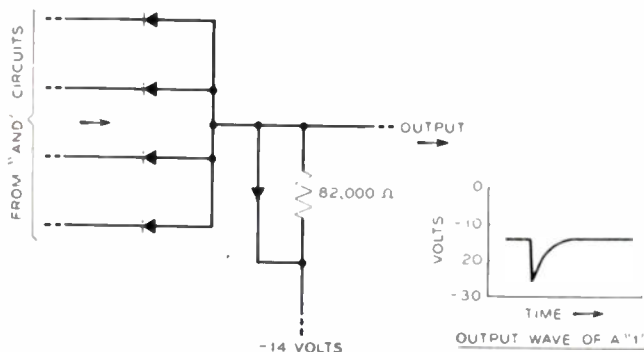
Fig. 9—A gate which acts as an OR circuit for negative signals.

Several advantages are realized by choosing the polarity of the gate circuits as shown. Current is drawn from an output terminal of a binary cell only when the internal impedance of the cell is low, giving good regulation. The current is of the polarity that is better tolerated by the cell. Finally, the gate output polarity will trigger a pulse amplifier without an inverting transformer.

One must be careful to assure that logic circuits do not put out spurious signals because of lack of synchronism in switching of input variables. In this case the gates are not enabled until the inputs have settled down.

Enabling can be done either by a separate tandem gate or by a program input directly on the logic circuits. The latter is more favorable in this case because it permits reduced drain on the binary cells. With proper dc level of the program pulse, the cells furnish no current in the interval between program pulses. There is also an opportunity for reduction of peak current in the binary cell output by inserting a low pass filter between binary cell and gate.

For proper operation, the reverse current in the information-input diodes of the "AND" gates must remain below some limit. To assure this, particularly at high temperatures, we used single-crystal PN junction germanium diodes (type M1727) in these positions. These diodes switch ON and switch OFF somewhat more slowly than the usual point contact type.[4]

In switching ON there is an initial high forward voltage drop, roughly 2 volts at 1 ma. This tends to cause a spurious output signal, which is prevented by the proper choice of gate bias voltages. In switching OFF, a diode may remain low resistance for several microseconds if it has carried a forward current for some time. This would cause trouble if the duty cycle were unfavorable. In this case, a PN diode carries

forward current for about a microsecond and has 1,666 μsec to recover, a cycle which is easily met.

Conclusions

The adder and shift register described showed concrete advantages over a similar vacuum-tube system. There was a power saving of about 10 to 1 and a saving in volume of about 5 to 1. Reduced volume stemmed partly from the smallness of the transistors and partly from the lower power dissipation. A specific conclusion was that power could and should be reduced much further by using lower supply voltages.

Some important characteristics that were studied are:
1. Probability of an error.
2. Tolerance of supply voltage changes.
3. Tolerance of temperature changes.

Errors were hunted by setting up the equipment so that it would perform all its functions over and over at full speed and would leave a permanent record if an error occurred in any operation whatsoever. It was found that the equipment would run at least for several days with no error. It may be necessary to shield the circuits from the outside world to prevent errors caused by false triggering of the pulse amplifiers.

Supply voltages to the entire equipment could be changed individually about ±25 or 30 per cent with no error. This was fairly satisfactory, and it should be improved by using transistors meeting current specifications.

At elevated temperatures, the OFF collector current and the switchoff time go up, so that this equipment was limited to about 50 degrees C. With higher ambient temperature, such equipment may be attractive because its low power makes cooling relatively easy.

Tests on the transistor were sufficient to insure about 98 per cent proper operation of the building-block circuits. The remainder showed slow triggering, for which there was no transistor test at the time. Tests on the building blocks insured operation of the equipment with proper safety margins.

It appears that transistors have unique advantages, such as the ability to deliver fast-rising, powerful pulses into a low-impedance load with very low drain from the power supply.

The packages that were developed are quite useful, even though this was an early project and there was no attempt to achieve high computing speeds. The feasibility of miniature switching circuits based on transistors was demonstrated. A number of basic qualities of transistors were brought out, and further transistor work was stimulated.

[4] L. A. Meacham and S. E. Michaels, "Observations of the rapid withdrawal of stored holes from germanium transistors and varistors," *Phys. Rev.*, vol. 78, no. 2, pp. 175–176; April 15, 1950.

Correspondence

Note on Gaseous Discharge Super-High-Frequency Noise Sources[*]

It is the object of this letter to present evidence that the fluctuation phenomena which prevent the use of gaseous discharges as a noise source at low currents and pressures are caused by the existence of moving striations. Although a discharge seems homogeneous to the eye, at low currents or pressures it can actually consist of bright and dark regions located periodically with distance along the discharge and traveling at thousands of centimeters per second. Pupp[1] has shown that the regions of high brightness are regions of high electron and ion concentrations and high electron temperature. He also demonstrated that the voltage gradient changes periodically.

The critical current below which a discharge will contain moving striations was given by Pupp[2] for He, Ne, A, and Kr. The straight lines in the upper right-hand portion of Fig. 1 represent his data. The curves plotted at the top indicate the currents and pressures below which the discharges of the given gases exhibit the objectionable fluctuations as noise sources. These are taken from the points plotted in Fig. 11 of the article[3] under discussion. The position of the curves on the graph and the similarity of the variation with atomic weight is taken as an indication that the objectionable fluctuations are caused by moving striations.

Johnson and Deremer state that the mixture of gases used in fluorescent-lamp discharges (mercury and 3.5-mm argon) result in "suppression" of the unwanted microwave fluctuations. This statement is consistent with the fact that the critical pressure for mercury vapor containing an inert gas above which the moving striations do not exist is considerably below the critical pressure for an inert gas at a given discharge current. This is shown by comparison with the lower curve in Fig. 1. The curve[4] for lamps containing Kr instead of A is a straight line through the same region and has about the same slope as the three upper curves from Pupp. The characteristic mercury color is maintained with or without striations in these mercury mixtures.

The difference between the power law for striations, demonstrated by the straight lines in Fig. 1 and the logarithmic relation, and for fluctuations, demonstrated[3] by Fig. 11, probably lies in the definition of critical definition of critical current. First, consider the critical current definition used in all the curves concerning striations. Moving striations travel from the anode toward the

* Received by the Institute, May 7, 1952.
[1] W. Pupp, "Oscillographic probe measurements on moving striations in the positive column of rare gases" (in German), *Phys. Zeit.*, vol. 36, p. 61; 1935.
[2] W. Pupp, "Moving striations in the positive column of the rare gases" (in German), *Phys. Zeit.*, vol. 33, p. 844; 1932. Reproduced by M. J. Druyvesteyn and F. M. Penning, "Mechanism of electrical discharges in gases of low pressure," *Rev. Mod. Phys.*, vol. 12, p. 87, fig. 80; 1940.
[3] H. Johnson and K. R. Deremer, "Gaseous discharge super-high-frequency noise sources," *Proc. I.R.E.*, vol. 39, pp. 908–914; August, 1951.
[4] H. Steele, "Characteristics of moving striations in a mercury-krypton mixture," *Phys. Rev.*, vol. 82, p. 571; 1951.

Fig. 1—Evidence that low current limitation to noise sources is caused by moving striation. Fluctuations or striation are observed at currents and pressures below the curves.

cathode. As current is increased toward the critical value, the moving striations first disappear from the cathode end of the positive column. As current is increased further, the homogeneous region near the cathode lengthens until at the critical current the striations have completely disappeared. The critical current at which the unwanted microwave effects disappear would be lower than that just defined because the anode end of the discharge is outside the guide.

The three sets of data, however, are inconsistent when the effect of diameter is considered. Pupp found no considerable difference in the critical current in discharges from $\frac{3}{4}$- to $2\frac{1}{2}$-inch diameter. The critical current increases with decreasing diameter for the curves of mercury with krypton. Johnson and Deremer find the microwave critical current increases with diameter.

The analysis of how the striations should influence the operation as a noise source is difficult because the number of striations within a given guide varies with all the parameters mentioned (gas pressure, current, tube diameter, and even tube length in precise measurements). Spacings between the bright regions of 1.25 and 3 times the $1\frac{1}{2}$-inch diameter have been observed in our mercury mixtures. Spacings of 2.5 and 5 times the diameter were observed by Pupp[5]

[5] W. Pupp, "Frequency and spacing of striations for moving striations in the positive column of rare gases" (in German), *Zeit. für tech. Phys.*, vol. 15, p. 257; 1934.

Correspondence

in 1-mm A discharges in $\frac{1}{2}$- and 2-inch diameters. In our mercury mixtures the shortest spacings are observed at currents just below the critical value. The unwanted microwave effects are thus observed with more than one bright region within the waveguide. If conditions were changed so that only one bright region or part thereof were within the guide at any instant, the amplitude of the microwave noise would oscillate at the frequency at which striations went through the guide. The study of this amplitude modulation should lead to a better understanding of moving striations.

HOWARD L. STEELE, JR.
Research Department
Westinghouse Electric Corp.

A Standard Waveguide Spark Gap*

Accurately rating the power-handling capacity of a microwave component is a frequent and difficult task. In studying this problem for the Bureau of Ships, we have found a simple waveguide spark gap to be a valuable tool and recommend its adoption by other microwave engineers.

The spark gap chosen consists of two hemispheres mounted three quarterwavelengths apart on the wide wall of the waveguide, as shown in Fig. 1. The radii of these hemispheres are about one quarter the guide height; each hemisphere is clamped to the wall by a single center screw. The spacing may be adjusted by mounting the hemispheres in separate waveguide sections and inserting spacers.
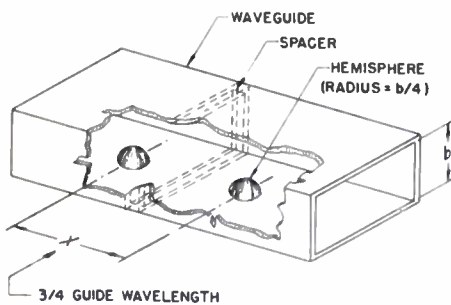
Fig. 1—Standard waveguide spark gap.

The hemisphere type of spark gap has many advantages over other types. It breaks down at about one-ninth the power capacity of the waveguide, as indicated by an approximate computation based on its simple geometric properties. It is well matched over a moderate bandwidth. Its structure insures that the arc will occur at a known location, which can readily be irradiated. In addition, it is small in size, convenient to inspect, easy to reproduce and inexpensive to fabricate for any waveguide size.

As a comparison standard of power capacity the spark gap is expected to prove preferable to straight waveguide. The latter is simple in concept, but extremely difficult

to measure in practice. Most magnetrons available today deliver power well below waveguide capacity; as a matter of practical experience, they are marginally capable of breaking down the hemisphere type of spark gap at atmospheric pressure. Furthermore, since the power capacity of a waveguide run is usually limited by some component having only a fraction of the power capacity of straight waveguide, it is desirable to use a standard comparable to such components in power capacity. Rating components relative to a spark gap would commonly simplify the entire design procedure by providing the engineer with realistic ratings.

If such a spark gap were universally used, it would serve as a common denominator in correlating measurements made on different test facilities, as well as checking the day-to-day operation of each. It is our recommendation that a waveguide spark gap of the hemisphere type be adopted for standard use with all microwave breakdown test facilities.

DAVID DETTINGER
Wheeler Laboratories
Great Neck, New York

Variable Slope Pulse Modulation*

A pulse-modulation scheme employing a train of equally spaced, equal-amplitude, triangular-shaped pulses has been developed, and appropriate modulator and detector circuits have been built and tested. The intelligence to be transmitted is carried in the slopes of the leading edges of the triangular pulses, the slope of a particular pulse being proportional to the amplitude of the intelligence signal at the time of the sampling. The rise time of a pulse is of the order of a few microseconds so that the intelligence (if an audio signal) will remain essentially constant during the time of one sampling, with the result that the slope of any one pulse is constant. In general, the amplitude of the modulating signal will change enough between samplings to produce a variation in the slopes of succeeding pulses.

The block diagram of the modulator and some modulator waveforms are shown in Fig. 1. The current generator supplies the current to charge the condenser C. This charging current is the cathode current of a pentode tube. The triangular-shaped pulse formed across C is coupled back to the screen grid of the pentode (via a cathode follower) in order to keep the screen grid to cathode potential constant. The cathode current, and so also the slope of the triangular-shaped voltage pulse formed across C, is then directly proportional to the intelligence signal, $a(t)$. The switch S is a thyratron which fires at some predetermined voltage across C, thus discharging the condenser and ending the pulse. A periodic positive pulse is applied across a resistor in the cathode circuit of the thyratron to stop conduction and so open the thyratron switch, resulting in the linear voltage rise across C as the current

from the current generator flows into the condenser.

The first step in the detection of these variable-slope pulses is to convert them to variable-amplitude pulses by means of a low time constant differentiating (RC) circuit.
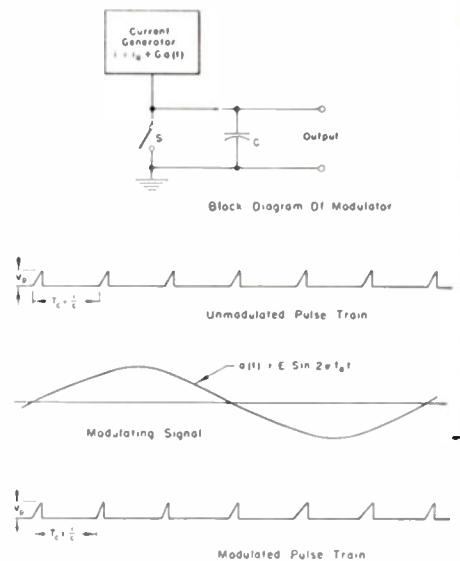
Fig. 1

The peak voltage of each pulse is directly proportional to the slope of the leading edge of the triangular-shaped pulse from which it was formed. From what has been said before, it follows that this peak voltage is directly proportional to the amplitude of the intelligence signal, $a(t)$, at the time of sampling at the modulator. The resulting variable amplitude pulses are then stretched in time and finally applied across the input of an appropriate low pass filter. The output of the filter is, ideally, a constant times $a(t)$.

Using a pulse repetition frequency of 8,000, an unmodulated pulse slope of 10,000,000 volts per second, and an intelligence signal of 500 cps, the distortion present in the output of the detector was determined for several degrees of modulation (the degree of modulation is defined as the ratio of maximum change in slope to unmodulated slope). It was found that for degrees of modulation less than 0.12 the distortion present was less than 0.5 per cent. As the degree of modulation was increased, the per cent distortion also increased until at a degree of modulation of 0.33, the distortion became 4.0 per cent. As the intelligence signal frequency was increased, the distortion percentages decreased, this because the filter passes fewer harmonics of the higher-frequency signals.

Plans are now underway to use slope-modulated trailing edges on the triangular-shaped pulses so that two channels of information may be carried by the same train of pulses.

OLAN E. KRUSE
Physics Dept.
Stephen F. Austin State College
Nacogdoches, Tex.

Contributors to Proceedings of the I.R.E.

Ernst F. W. Alexanderson (A'13–M'13–F'15) was born January 25, 1878, in Uppsala, Sweden. He is a graduate of Royal Technical University, Stockholm, Sweden, from which institution he received an honorary degree of Dr. of Technology. He has also received an honorary degree of Sc.D. from Union College, Schenectady, and honorary Ph.D. from Uppsala University, Sweden.

E. ALEXANDERSON

Mr. Alexanderson joined General Electric Co. in 1902 and served that company continuously until his retirement in January of 1952. When the Radio Corporation of America was formed in 1920, he served as chief engineer during its first four years, and as chief consulting engineer until 1930.

Dr. Alexanderson has done pioneer work in transoceanic communication (Alexanderson Alternator), radio reception, selective tuning, television, electronics for radio and power, and magnetic amplifiers.

Dr. Alexanderson is a registered professional engineer and recently rejoined RCA as a consultant.

❖

A. E. Anderson (S'38, A'41, M'45) was born on April 22, 1916, in Lima Ohio. He was graduated in 1939 with a B.S. degree in E.E. from Ohio State University, where that same year he received a M.Sc. degree.

A. E. ANDERSON

He then joined Bell Telephone Laboratories, where until 1942 he was engaged in the development of microwave electron tubes. During World War II he continued in this field with the U. S. Army at Evans Signal Laboratory, Belmar, N. J. He was closely associated with the Joint Army-Navy Electron Tube Committee, and was influential in standardizing electron tubes. From 1945 until 1948, Mr. Anderson worked on the development of electron-beam tubes, and since 1948 he has been active in the development of semiconductor devices including the transistor.

Mr. Anderson has served on a number of I.R.E. committees, and is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, Sigma Pi Sigma (physics), and the RDB semiconductor subpanel.

❖

Lorne D. Armstrong was born in Saskatchewan, Canada, on November 14, 1917. He received the B.A. degree in physics from the University of Saskatchewan in 1939. He received the degree of M.A. in physics from the University of Toronto in 1940 and the Ph.D. degree in 1949.

Dr. Armstrong was associated with the National Research Council of Canada as research physicist from 1941 to 1947. He joined RCA in 1949, working on ceramic seals and high-frequency power tubes at the Laboratories Division in Princeton, N. J. from 1949 until 1951. Since 1951 he has been engaged in research work on transistors.

L. D. ARMSTRONG

❖

Ralph Bray was born in New York City on September 11, 1921. He received his undergraduate training at Brooklyn College and joined the Purdue University semiconductor group in 1942. He has been connected with this group since then, and received the Ph.D. degree in physics in 1949 from Purdue.

Dr. Bray is now assistant professor of physics on leave at the Technical Institute, Delft, Holland where he is engaged in research on the electrical properties of metals. His interest is primarily in the investigation of rectifying barriers and their properties.

RALPH BRAY

Dr. Bray is a member of A.P.S., Sigma Xi, and Sigma Pi Sigma.

❖

George Bryan was born in Camden, N. J. on January 15, 1918. He first became active in the field of radio in 1934, and from this date until 1941 was engaged in Marine radio communications.

During World War II Mr. Bryan participated in work on receiver antijamming circuits for use in long-range radar systems. He was granted a patent on an antijamming radar receiving system.

GEORGE BRYAN

From the middle of 1941 to the present time Mr. Bryan has been employed at the Signal Corps Engineering Laboratories, Fort Monmouth, N. J. Since 1948 he has been engaged in work in the field of nucleonics.

Mr. Bryan participated in the recent A-bomb test at Eniwetok.

Abraham Coblenz was born in Poland on August 25, 1912. He received the B.S. degree from the College of the City of New York in 1935, attending the evening session. In 1944 he received the M.E.E. degree from the Polytechnic Institute of Brooklyn. Further post-graduate work at the latter institution was done in physics, while attending the evening session.

A. COBLENZ

From 1929 to 1940 Mr. Coblenz was engaged in various phases of radio work, joining the Department of Public Works of the City of New York as an electrical engineer in the latter year. In 1947 he became affiliated with the Signal Corps Engineering Laboratories at Fort Monmouth, N. J. as a physicist in radar and solid-state physics.

Mr. Coblenz served as a radar technician during World War II, and received training in radar principles at the Naval Research Laboratories, Washington, D. C.

❖

Esther M. Conwell (Mrs. Abraham A. Rothberg) was born in New York, N. Y. She received the B.A. degree, magna cum laude, from Brooklyn College in 1942. In 1945 the University of Rochester awarded her the M.S. degree and in 1948 the University of Chicago the Ph.D. degree.

E. M. CONWELL

Before joining Bell Laboratories in 1951, where she is now engaged in transistor research, Dr. Conwell taught physics at the University of Chicago, the University of Rochester, and Brooklyn College. In 1942–1943 she spent a few months working in the quality-control department of Western Electric's tube shop.

Dr. Conwell is a member of A.P.S. and A.A.A.S.

❖

Stanley F. Danko (M'46) was born on January 6, 1916 in New York, N. Y. He received the B.S. degree in electrical engineering from Cooper Union in 1937, with graduate studies in uhf techniques at the Moore School of Electrical Engineering, University of Pennsylvania. He has been employed in the U. S. Army Signal Corps Inspection Laboratory in Philadelphia for the pe-

STANLEY F. DANKO

riod, 1940–1946, joining the Signal Corps Engineering Laboratories at Fort Monmouth in 1946.

Mr. Danko has been associated with the electronic component development activities at the Signal Corps Engineering Laboratories, specializing in the development of miniaturization techniques, including printed circuits and packaging of electronic assemblies. He has been primarily responsible for the establishment of the autosembly (printed-circuit) system, and has authored a number of papers on the subject.

Mr. Danko is currently Assistant Chief of the Component Parts Section of the Signal Corps Engineering Laboratories, concerned with administration of research and development programs for electronic components.

❖

E. Dickten, Jr. was born in Elberfeld, Germany, on November 15, 1903. He received his education in the schools of Paterson, N. J., and the Brooklyn Polytechnic Institute Evening School.

In 1920 Mr. Dickten became a member of the Bell Telephone Laboratories, and after several years as an instrument maker, transferred to the Research Department, where he spent nine years on various phases of phonograph-disc recording. In 1935, as a member of the technical staff, he became engaged in relay and ringer research, maintained the Laboratories' frequency standard, and studied selective voice control. For the past four years he has conducted transistor research with special emphasis on transistor applications.

E. DICKTEN, JR.

❖

James M. Early (A'48), was born in Syracuse, N. Y. on July 25, 1922. He graduated from New York State College of Forestry with the B.S. degree, cum laude, in 1943, then served in the U. S. Army at Oak Ridge, Tenn. Discharged in 1945, Dr. Early attended Ohio State University, where he received M.S. and Ph.D. degrees in 1948 and 1951, respectively. Meanwhile he taught electrical engineering at Ohio State.

J. M. EARLY

In 1951 Dr. Early joined Bell Telephone Laboratories, where he is a member of the electronic apparatus development group. His field has been junction transistor development.

He is a member of Eta Kappa Nu and an associate of Sigma Xi.

❖

J. J. Ebers (S'46–A'48) was born in Grand Rapids, Mich. on November 25, 1921. He received the B.S. degree from Antioch

College in 1946, his education having been interrupted by three years' service in the U. S. Army. He obtained the M.S. degree in electrical engineering from Ohio State University in 1947, and the Ph.D. in 1950.

From 1947 to 1951 Dr. Ebers was an instructor and later an assistant professor in the electrical engineering department of Ohio State University, as well as research associate for the Ohio State University Research Foundation. In September of the latter year he joined the transistor development group of Bell Telephone Laboratories at Murray Hill, N. J.

J. J. EBERS

Dr. Ebers is a member of Eta Kappa Nu, Sigma Xi, and A.P.S.

❖

H. Y. Fan, a native of China, received the Sc.D. from the Massachusetts Institute of Technology in 1937. From 1937 to 1947 he was with the Radio Research Institute and Department of Physics of the National Tsing Hua University in Kunming and Peiping, China, in the capacity of associate professor (1937 to 1939) and professor (1939 to 1947). He worked on oxide-cathodes, photoelectric emission, and crystal rectifiers.

H. Y. FAN

Mr. Fan was a member of the staff of the Electronics Research Laboratory of M.I.T. from 1947 to 1948, working on thermionic and secondary emission of thoria cathodes. Since 1948 Mr. Fan has been a professor in the Department of Physics, Purdue University, working on semiconductors, mainly germanium and silicon.

Mr. Fan is a fellow of A.P.S. and a member of Sigma Xi.

❖

Belmont G. Farley (A'48) was born in Cape Girardeau, Mo. on December 29, 1920. He received the B.S. degree in electrical engineering and mathematics from the University of Maryland in 1941. After a short period of graduate work in mathematics at the Massachusetts Institute of Technology, he joined the staff of the Radiation Laboratory there, where he assisted in the development and overseas introduction of the SCR-615 radar, and served as project engineer in the development of the AN/TPS-10 radar.

B. G. FARLEY

In 1945 Dr. Farley resumed graduate study in physics at Yale University, receiving the Ph.D. degree in 1948 for experimental work in nucleon-velocity measurement. Since 1948 he has been associated with the Bell Telephone Laboratories, and is now working on high-speed switching transistors.

Dr. Farley is a member of the American Physical and Mathematical Societies and Sigma Xi.

❖

Jean H. Felker (A'46) was born March 14, 1919, in Centralia, Ill. In 1941 he received the B.S. degree in E.E. from Washington University in St. Louis, Missouri. After graduation he was associated with the Emerson Electric Company, and later he instructed at a Naval Training School in St. Louis.

He served in the U. S. Army from 1942 to 1945, first as a radar maintenance officer attached to the Royal Artillery in Great Britain, and later as an Army publications officer in this country. Since 1945, Mr. Felker has been a member of the technical staff of Bell Telephone Laboratories, where he is currently in charge of a group applying transistors to digital computers.

J. H. FELKER

❖

Harold Fleisher was born in Kharkov, Russia in 1921. He received the B.A. degree in 1942 and the M.S. degree in 1943 in physics from the University of Rochester. In 1951 he received the Ph.D. degree in physics from the Case Institute of Technology.

During 1942–1943 Dr. Fleisher was a research assistant in the Physics Department and the Institute of Optics of the University of Rochester. For the next three years he was a staff member of the Radiation Laboratory of the Massachusetts Institute of Technology, and contributed to volume 18 of the Radiation Laboratory Series. In 1946 he was a senior engineer with the Rauland Corporation of Chicago, and from that year to 1950 he was an Instructor at the Case Institute of Technology.

HAROLD FLEISHER

From 1950 to the present time Dr. Fleisher has been a physicist with the International Business Machines Corporation (Engineering Laboratories) at Poughkeepsie, New York, where he has been working on transistor research and development.

Dr. Fleisher is a member of A.P.S. and Sigma Xi.

❖

H. G. Follingstad (S'46, A'49) was born in Wanamingo, Minn., on January 6, 1922. He served in the Signal Corps from 1942–

46 as a radar technician concerned with several long-range, early-warning, search radars. In 1947 he was graduated, with highest distinction, from the University of Minnesota with the B.E.E. degree, and in 1948 he joined the technical staff of Bell Telephone Laboratories.

In 1950 Mr. Follingstad completed a Bell Telephone Laboratories Communications Development Training Program as well as a series of assignments in the specialized fields of switching-apparatus testing and the design of microwave-test equipment and other electronic-measuring apparatus. From 1950 to early 1951, he was engaged in transistor circuit development. Since that time he has been associated with an electronic-measuring apparatus group and is at present concerned with transistor-measurement problems and test equipments.

Mr. Follingstad is a member of Tau Beta Pi.

❖

Robert A. Gerhold (M'49) was born on May 14, 1916 in New York, N. Y. He received the B.S. degree in electrical engineering in 1936 from Cooper Union. From 1936 to 1939 he was Laboratory Engineer with the Signal Engineering and Manufacturing Company, New York. For the next three years he was employed by the Board of Transportation of the City of New York as a supervisor of the installation of electrical power, control, and supervisory equipment in the subway and substations. In 1941 he joined the staff of the Naval Material Laboratory in Brooklyn, N. Y., where he worked on the evaluation and development of audio-frequency communications systems and related equipments. He became Chief of the Communications Section in 1947 and consultant in electroacoustics in 1950.

In 1951 Mr. Gerhold was appointed Chief of the Miniaturization and Printed Circuits Group in the Components and Materials Branch of the Signal Corps Engineering Laboratories at Fort Monmouth, N. J. He is responsible for the Signal Corps' research and development program on systems and techniques for the miniaturization of military communications equipments.

Mr. Gerhold is a member of A.I.E.E., A.P.S., the Acoustical Society of America, and the Audio Engineering Society.

❖

L. J. Giacoletto (S'37–A'42–M'44–SM'48) was born in Clinton, Ind., November 14, 1916. He 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, while holding an appointment as research assistant there. From 1939 to 1941, while a teaching fellow in the department of electrical engineering at the University of Michigan, he engaged in frequency-modulation research. He received his Ph.D. degree in electrical engineering from Michigan in 1952.

From 1941 to 1945 Dr. Giacoletto served with the Signal Corps in laboratory assignments concerned with development activities in radio, navigational, and meteorological direction-finding equipments. Since 1946 he has been serving as a research engineer with the RCA Laboratories Division, Princeton, N. J.

Dr. Giacoletto is a member of the American Association for the Advancement of Science, Gamma Alpha, Iota Alpha, Phi Kappa Phi, Tau Beta Pi, and Sigma Xi. He has been chairman of the Monmouth Subsection and the Princeton Section of the I.R.E. and is now a member of the Board of Editors and the Administrative Committee of the Board of Editors.

❖

Norman J. Golden (A'51) was born in Newark, N. J. on September 6, 1924. He received the B.S. degree from Harvard College in 1944, concentrating in electronic physics. He did graduate work in theoretical physics in 1946 at the University of California and in 1949–1951 at Washington University in St. Louis, Mo.

During World War II Mr. Golden served as an electronics officer in the Navy. In 1946 and 1947 he did electronic instrumentation work at the E. I. DuPont Experimental Station. In the latter year he was also concerned with wide-band circuits at the Federal Telecommunications Laboratory. In 1948 he worked on gaseous discharge devices at the Raytheon Manufacturing Company, and in 1950 on radar-system design at Emerson Electric. At present he is in charge of the Circuits and Applications Section of the Semiconductor Engineering Department of Sylvania Electric Products, Inc., Electronics Division, Newton, Mass.

Mr. Golden is a member of Sigma Xi and A.P.S.

❖

Gordon N. Hall was born in Covington, Kentucky, on September 2, 1924. He attended the University of Kentucky where he received a B.S. degree in physics in 1951.

During the war years his time was devoted to airborne radar and communication equipment. From 1946 to 1952 he was actively engaged as a radio broadcast engineer. During the later period his time was devoted to design, installation, and maintenance of broadcast and high frequency short-hop communication equipment.

Mr. Hall joined the engineering staff of the General Electric Co. in February, 1952. He is currently occupied with the development of the germanium power rectifier.

❖

Robert N. Hall was born on December 25, 1919 in New Haven, Conn. He received the B.S. degree in physics in 1942 from the California Institute of Technology. Immediately after graduation, he joined the General Electric Company in Schenectady, where he worked on the development of cw magnetrons and other radar components.

Early in 1946, Dr. Hall returned to the California Institute of Technology as a graduate student in nuclear physics. His thesis work involved the development of a high-intensity proton source and the determination of the low-energy transmutation of carbon.

He received the Ph.D. degree in the summer of 1948 and returned to the General Electric Research Laboratory at The Knolls in Schenectady, N. Y. Since then he has worked on the purification of germanium and the preparation and study of p-n junction devices.

Dr. Hall is a member of A.P.S., Tau Beta Pi, and Sigma Xi.

❖

James Ridout Harris (A'43–M'49) was born in Lockhart, Tex. on April 14, 1920. After graduating from the University of Richmond (with a B.S. degree in physics in 1941), Mr. Harris spent a year and a half with the Chesapeake and Potomac Telephone Company of Virginia, first maintaining toll central office equipment, and later, writing specifications for its installation.

In 1942 he joined Bell Laboratories. During the next eight years he developed radio communication and navigation equipment for military air-

craft. Meanwhile, Polytechnic Institute of Brooklyn conferred the M.S.E.E. degree in 1948. Since then he has been concerned with the development of transistors and transistor circuits for digital computers.

Mr. Harris is a member of Phi Beta Kappa, Sigma Xi, and Sigma Pi Sigma.

❖

L. P. Hunter was born in Wooster, Ohio on February 11, 1916. He received the B.S. degree in physics from the Massachusetts

L. P. HUNTER

Institute of Technology, the B.A. degree from the College of Wooster, and the M.S. and D.Sc. degrees from Carnegie Institute of Technology, respectively.

In 1940 Dr. Hunter joined the Westinghouse Electric Corporation, where he did work on the physics of solids.

During World War II he was engaged in microwave magnetron development, and later spent a year at the University of California working on the calutron project. Dr. Hunter has since spent two years at Oak Ridge National Laboratory engaged in study of radiation effects in solids.

From 1948 to 1951 Dr. Hunter was manager of the solid-state electronics section of the Westinghouse Research Laboratories. In June, 1951 he joined the International Business Machines Corporation as a project engineer in charge of transistor research and development.

Dr. Hunter is a member of the A.P.S.

❖

Edward Keonjian (M'50–SM'52) was born in Tiflis, Caucasus, Russia on August 14, 1909. He received the M.S. degree in electri-

EDWARD KEONJIAN

cal engineering from the Leningrad Institute of Electrical Engineering in 1932.

After graduation Mr. Keonjian joined the Leningrad Central Radiolaboratory as a research engineer and in 1935 he was transferred to the Leningrad Research Institute of Electro-Physics as a senior

engineer. From 1939 to 1940 Mr. Keonjian was engaged in the project and development work for Naval Radio Communication.

After spending five years in Germany, Mr. Keonjian came to the United States and in 1947, joined the Westinghouse Electric Corporation in Jersey City, N. J. In 1949, he was appointed as a lecturer in electrical engineering at the City College of New York. After two years, in 1951, Mr. Keonjian joined the Engineering Staff of the General Electric Company, where he is engaged in the development work concerned with transistors and advanced electronics circuitry. He is a member of the General Electric Engineers Association.

Russell R. Law (A'34–M'40–SM'43) was born on January 11, 1907, in Hampton, Iowa. He received the B.S. and M.S. degrees

in electrical engineering from Iowa State College in 1929 and 1931. From 1929 to 1931 he taught electrical engineering there. In 1933 he received the D.S. degree from Harvard Engineering School. During 1933 and 1934 he was a research associate at Harvard University.

RUSSELL R. LAW

He joined the research and engineering department of RCA Manufacturing Company in 1934. In 1942 he transferred to their Laboratories Division at Princeton, N. J.

Dr. Law is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and the A.P.S.

❖

Kurt Lehovec (S'52) was born in Ladowitz, Czechoslovakia on June 12, 1918. He received the Ph.D. in physics from the Uni-

KURT LEHOVEC

versity of Prague in 1941. From 1942 to 1945 he was assistant professor in physics at the University of Prague and director of a research laboratory on selenium rectifiers, sponsored by the Süddeutsche Aparate Fabrik, Nürnberg. The following year he was a member of the staff of the Süddeutsche Aparate Fabrik.

From 1947 to 1952 Mr. Lehovec was consultant to the Signal Corps Engineering Laboratories, Fort Monmouth, N. J., for the field of semiconductors, and during 1951 to 1952 a consultant for the Radio Receptor Company, Brooklyn, N. Y. Since July 1, 1952, he has been head of the transistor group at the Sprague Electric Company, North Adams, Mass.

Mr. Lehovec is a member of A.P.S.

❖

Olof Lindberg was born in Woodlynne, N. J. in 1924. During World War II he served in the U. S. Army. In 1951 he received

the B.S. degree in engineering physics from the University of Illinois.

Following graduation, Mr. Lindberg went to work for the Westinghouse Electric Corporation. He is now at the Research Laboratories doing work on semiconductors and the preparation of ger-

OLOF LINDBERG

manium. He assisted in the design and construction of Hall measurement apparatus.

Arthur W. Lo (S'48–A'50) was born in Shanghai, China, on May 21, 1916. After receiving the B.S. degree in physics from

Yenching University in 1938, he taught in West China Union University and Yenching University until he came to the United States in 1945. He received his M.S. degree in physics from Oberlin College in 1946, and his Ph.D. in electrical engineering from the University of Illinois

ARTHUR W. LO

in 1949, while serving as a research associate. He spent the following year as assistant professor in electrical engineering at Michigan College of Mining and Technology, and the next as lecturer in electrical engineering at the College of the City of New York.

Dr. Lo joined RCA in 1951 as a member of the staff of the Advanced Development Engineering Section of the Engineering Products Department, assigned to advanced development work in transistor circuitry.

Dr. Lo is a member of Sigma Xi, Phi Kappa Phi, Pi Mu Epsilon, Eta Kappa Nu, and the American Association of University Professors.

❖

G. E. McDuffie, Jr. was born on April 4, 1925 in Washington, D. C. He received the B.E.E. degree in 1949 and the M.E.E. de-

gree in 1952, both from The Catholic University of America in Washington D. C.

Since receiving his Bachelor's degree, Mr. McDuffie has been a member of the faculty of the Department of Electrical Engineering at The Catholic University of America,

G. E. McDUFFIE, JR.

where he is presently engaged in research on the application of barium titinate to counter and computer elements. During the summers he has been employed by the Naval Research Laboratories, Washington, D. C.

Mr. McDuffie is an associate member of Sigma Xi.

❖

Harold C. Montgomery was born in Detroit, Mich., on April 6, 1907. He received an A.B. degree from the University of Southern California in 1929 and an M.A. degree from Columbia University in 1933.

In 1929 he joined Bell Laboratories and for 12 years studied problems in physiological acoustics and hearing acuity. After World War II he engaged in research in microphonic con-

H. C. MONTGOMERY

tacts. In 1936 Mr. Montgomery cooperated with the U. S. Public Health Service in a hearing survey, and in 1938 and 1939 made a similar study for the Bell System at the New York World's Fair. From 1941 to 1945 he worked on N.D.R.C. contracts for the navy, and now conducts research in semiconductors—particularly their noise aspects.

Mr. Montgomery is a member of Phi Beta Kappa and Phi Kappa Phi.

❖

J. A. Morton (A'37–SM'49) was born in St. Louis, Mo. on September 3, 1913. He received the B.S. degree in electrical engineering from Wayne University in 1935 and the M.S. degree from the University of Michigan in 1936.

Mr. Morton joined Bell Telephone Laboratories in 1936. He is currently in charge of the development of the transistor and other semiconductor devices. In the past he has been concerned with research on coaxial cables, microwave amplifier circuits, and radar receivers, and with vacuum-tube development. He designed a microwave tube used in the New York–San Francisco microwave relay system.

Mr. Morton is a member of Eta Kappa Nu, Alpha Delta Psi, Phi Kappa Phi, Sigma Xi, and the Mackenzie Honor Society. He has served on several technical committees of the I.R.E. in the past.

❖

Charles W. Mueller (S'35–A'36–VA'39–SM'45) was born in New Athens, Ill., in 1912. He received the B.S. degree in electrical engineering from the University of Notre Dame in 1934 and the S.M. degree in electrical engineering from the Massachusetts Institute of Technology in 1936.

From 1936 to 1938 Dr. Mueller was associated with the Raytheon Production Corporation, first in the engineering supervision of factory production of receiving tubes, and then in the development of gas-tube voltage regulators and cold-cathode thyratrons.

In 1942 he received the degree of Sc.D. in physics from M.I.T. While there Dr. Mueller worked on a government contract on the development of gas-filled special-purpose tubes for counting operations. Since 1942 he has been a member of the technical staff at RCA Laboratories in Princeton, N. J. where he has been engaged in research on high-frequency receiving tubes, secondary electron emission phenomena, and solid-state devices.

Dr. Mueller is a member of the American Physical Society and Sigma Xi. He was the 1946–1947 chairman of the Princeton Section of the I.R.E.

❖

David H. Navon was born in New York City on October 28, 1924. He received the B.S. degree in electrical engineering from the City College of New York in 1947. Following graduation he served two years in the U. S. Army.

After teaching college physics at Mohawk College, Utica and Queens College, Flushing, he obtained the M.S. degree from New York University in 1950. Since that date he has been with the solid-state group at Purdue University, where he is doing work leading to a doctorate in physics.

Mr. Navon is a member of Sigma Xi and Sigma Pi Sigma.

❖

Reinald S. Nielsen was born on February 26, 1927 at Bridgeport, Conn. He attended the Junior College of Connecticut from 1946 to 1948, and received the B.E.E. degree from Syracuse University in 1951.

Mr. Nielsen has worked for the General Electric Company at Bridgeport, Conn. and Syracuse, N. Y., and for Syracuse University from 1946 to 1951. Upon graduation he joined Sylvania Electric Products Company, Electronics Division, Boston, Mass. In March of 1952 he transferred from the Equipment Development Group to the Circuits and Applications Group of the Sylvania Electronics Semiconductor Engineering activity.

In the U. S. Navy service from 1944 to 1946, Mr. Nielsen attended the Electronic Technicians school, where he became an instructor. At the Junior College of Connecticut he became a member of Phi Theta Kappa. He is also a member of Tau Beta Pi and the Syracuse University E.E. Honor Society.

❖

Harry L. Owens (M'46) was born on August 11, 1920 in Indianola, Miss. He received the B.S. degree in electrical engineering from Mississippi State College in 1942. During World War II he was on active duty with the Signal Corps and studied electronics and radar at Harvard University and the Massachusetts Institute of Technology, later serving with the Signal Corps Engineering Laboratories in the VT fuse tube program. In 1946 he was discharged with the rank of Captain.

Since 1946 Mr. Owens has been employed by the Signal Corps Engineering Laboratories and assigned to the Thermionics Branch, Evans Signal Laboratory, where he is in charge of work on semiconductor devices, including transistors.

❖

Jacques I. Pankove (formerly Pantchechnikoff) was born in Russia on November 23, 1922. He grew up in France where he completed his secondary education. From 1942 to 1944 he attended the University of California in Berkeley, Calif., where he received his B.S. degree.

The following two years he served with the United States Army. He then returned to the University of California where he received his M.S. degree in 1948. Since then, he has been engaged in semi-conductor research at the RCA Laboratories in Princeton, N. J.

Mr. Pantchechnikoff is an associate member of Sigma Xi and holds membership in the American IEE.

❖

Gerald L. Pearson was born in Salem, Ore., on March 31, 1905. He was graduated from Willamette University in 1926 with an A.B. degree, and was awarded an M.A. degree by Stanford University in 1929. He then became a member of the physical research department of Bell Telephone Laboratories.

For several years Mr. Pearson studied noise in resistors, vacuum tubes, and carbon microphones, then engaged in semiconductor research related to the development of thermistors. Since 1945 he has been investigating the properties of semiconducting materials for transistor applications.

During World War II Mr. Pearson worked on N.D.R.C. contracts for the army and the navy. In 1948 the A.I.E.E. awarded him a prize for the best paper in research, "Properties and Uses of Thermistors."

Mr. Pearson is a fellow of the American Physical Society and a member of Sigma Xi.

Richard L. Petritz was born on October 24, 1922 in Rockford, Ill. He received the B.S. degree in 1944, the B.S.E.E. degree in 1946, and the M.S.E.E. degree in 1947 from Northwestern University. In the interim, 1944–1946, he served as radar officer aboard a destroyer.

R. L. PETRITZ

Dr. Petritz worked part time on an analogue computer project at Northwestern University from 1946 to 1947. He received a two-year Murphy Foundation Fellowship, during which time he did graduate work in physics at Northwestern. He obtained the Ph.D. degree in theoretical physics in 1950.

From 1949 to 1950 Dr. Petritz was an instructor in the Department of Electrical Engineering at Northwestern. During the summer of 1949 he was associated with the Theoretical Physics Division at Los Alamos, N. M. Since 1950 he has been a member of the faculty of the Department of Physics at the Catholic University of America and a member of the Semiconductor Research Branch at the Naval Ordnance Laboratory, White Oak, Md. He has been concerned with research in semiconductors, particularly fluctuation problems and photoconductivity.

Dr. Petritz is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, A.P.S., and the Washington Philosophical Society.

❖

Robert L. Pritchard (S'45–A'51) was born in Irvington, N. J. on September 8, 1924. From 1942 to 1943 he worked as a technical assistant in the Circuit Research Department of Bell Telephone Labs.

R. L. PRITCHARD

Dr. Pritchard spent nine months on active duty (recruit training and underwater sound project at Solomons, Md.) with the U. S. Navy. He entered the V-12 program and attended M.I.T. He was graduated from Brown University in 1946 with the B.S. degree. The following year he obtained the M.S. degree in engineering sciences and applied physics at Harvard University. He then worked as a research assistant, full time for one year and half time for two years at the Acoustics Research Laboratory, Harvard University. He received the Ph.D. degree in acoustics there in 1950.

Since that date Dr. Pritchard has been affiliated with the General Electric Research Laboratory in Schenectady, N. Y. He has worked on a variety of projects in communications, including some work in acoustics. For the past year he has been doing research on the transistor from an electric-circuit point of view.

Dr. Pritchard is a member of the Acoustical Society of America and Sigma Xi.

❖

Louise Roth was born in 1910 in Evansville, Ill., and attended Evansville College. She came to Purdue University in 1945. She worked first as an analyst in the Chemistry Department and, in 1948, joined the Physics Department, where she was engaged in the growing of anthracene crystals and the luminescence of anthracene under irradiation. Since 1950 she has worked on semiconductors, particularly in connection with the growing and purification of germanium. At present she is in charge of chemical technology in the Physics Department.

LOUISE ROTH

Miss Roth is a member of Sigma Xi, the American Chemical Society, and the American Association for the Advancement of Science.

❖

Charles L. Rouault was born on April 17, 1916 in Jersey City, N. J. He received the degree of M.E. from Stevens Institute of Technology in 1939 and joined the General Electric Company, completing the Test and Advanced Course Program.

C. L. ROUAULT

During World War II, Mr. Rouault was a mechanical and project engineer on shipborne and ground radar sets and was awarded the Navy Department Certificate of Commendation. From 1945 to 1947 he was engaged in general consulting practice. He then rejoined the General Electric Company, working on wide-band amplifiers and general feedback problems. He received the Coffin Award of General Electric Company in 1952 for the development of a selective signaling system.

At present, Mr. Rouault is project engineer in charge of applications and advanced development in the semi-conductor engineering section.

❖

John S. Saby (SM'52) was born on March 21, 1921 in Ithaca, N. Y. He received the B.A. degree in physics at Gettysburg College in 1942 and the M.S. and Ph.D. degrees from the Pennsylvania State College in 1944 and 1947, respectively. He taught physics at the latter College, and did research in atmospheric acoustics, X-Rays, and wave mechanics. From 1947 through 1950, he was an instructor on the faculty of the Physics Department at Cornell University.

JOHN S. SABY

In 1951, Dr. Saby joined the Electronics laboratory of the General Electric Company in the capacity of assistant section head. He is engaged in research and development of semiconductor devices, particularly junction transistors.

Dr. Saby is a member of A.P.S., Phi Beta Kappa, and Sigma Xi.

❖

Baldwin Sawyer (A'52) was born in Narragansett, R. I. on July 21, 1922. He was graduated from Yale University, in the class of 1944, with the B.E. degree in metallurgy.

For three years Mr. Sawyer did metallurgical research on the Manhattan Project at the University of Chicago. After World War II he decided to go more deeply into the science of metals and other solids, and began graduate studies in physics at the University of Pennsylvania. These he continued at Carnegie Institute of Technology, receiving the D.Sc. degree in physics in 1952. In the fall of 1951 he joined the technical staff of the Bell Telephone Laboratories to work in the development of semiconductor devices.

BALDWIN SAWYER

Mr. Sawyer is a member of Tau Beta Pi and Sigma Xi.

❖

J. S. Schaffner (S'48–A'51) was born on December 21, 1921 in Schwanden, Switzerland. He received a degree in electrical engineering from the Swiss Institute of Technology in 1946 and his M.S. and Ph.D. degrees from the University of Illinois in 1948 and 1950, both in electrical engineering. From 1950 to 1951, he was research assistant professor at the University of Illinois, and also worked on the theory of nonlinear oscillators in the Electron Tube Laboratory.

J. S. SCHAFFNER

Dr. Schaffner joined the General Electric Company in 1951. Since then, his work has been mainly with transistor circuitry.

Dr. Schaffner is a member of Eta Kappa Nu, Sigma Xi, and A.P.S.

L. G. Schimpf (SM'46) was born in New Washington, Ohio, on January 29, 1914. He was graduated in 1937 from Ohio State University with a B.E.E. degree. That year he became a member of the switching research department at Bell Laboratories, occupying himself with the application of electronic devices to switching, with special attention given to cold cathode tubes.

L. G. SCHIMPF

After World War II Mr. Schimpf joined the transmission research group to study transmission circuits and acoustics as applied to local subscriber stations. He has also made analyses of dialing methods.

Mr. Schimpf is a member of the Acoustical Society of America, Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

❖

Richard F. Shea (A'29–M'32–SM'43) was born on September 13, 1903 in Boston, Mass. He received the B.S. degree from the Massachusetts Institute of Technology in 1924.

From 1925 to 1937 Mr. Shea was a member of the engineering departments of American Bosch, Amrad, Kolster, Atwater Kent, Pilot, Freed-Eiseman and Fada, and was chief engineer of the last three.

RICHARD F. SHEA

Mr. Shea joined the General Electric Company in 1937 as development engineer and group leader of the Receiver Department. He was transferred to the Specialty Division in 1945 as Consulting Engineer and to the Electronics Laboratory in 1947. He is currently Section Engineer in charge of Advanced Circuits Section.

Mr. Shea is a fellow of the Radio Club, and has served on several I.R.E. committees in the past.

❖

Jacob Shekel (A'52) was born in Bialystok, Poland on January 6, 1926. He received his engineering education at the Hebrew Institute of Technology in Haifa, Israel, and graduated with the degree Ingénieur (E.E.) in 1951.

Mr. Shekel is employed by the Scientific Department of the Ministry of Defense of Israel, where he does research and development work in the fields of network

JACOB SHEKEL

analysis and synthesis and uhf techniques.

During the summer of 1949 Mr. Shekel was a guest student on the Foreign Students Summer Project at the Massachusetts Institute of Technology, where he took a course on microwaves.

❖

John N. Shive, since joining Bell Telephone Laboratories in 1939, has devoted most of his time to the development of semiconductor devices, and is responsible for the invention of the phototransistor, announced to the public in 1950. His contributions to transistor development were preceded by work on selenium rectifiers.

JOHN N. SHIVE

A native of Baltimore, where he was born February 22, 1913, Dr. Shive was graduated in 1934 from Rutgers University with a B.S. degree in physics and chemistry. He received the Ph.D. degree in 1939 from Johns Hopkins University.

Dr. Shive is a fellow of the A.P.S. and a member of the A.A.A.S., Phi Beta Kappa, and Sigma Xi.

❖

William Shockley is in charge of the Transistor Physics group at Bell Telephone Laboratories. He directed the group which developed the point-contact transistor, and on the basis of theory, he predicted the junction transistor more than three years before working units were announced publicly.

WILLIAM SHOCKLEY

Dr. Shockley was born in London, England on February 13, 1910, and came to the United States three years later. He was graduated from California Institute of Technology in 1932 with the B.S. degree. Continuing his studies at the Massachusetts Institute of Technology on a teaching fellowship, he received the Ph.D. degree in physics in 1936 and in September of that year joined the Physical Research Department of Bell Telephone Laboratories. His work has included vacuum-tube and electron-multiplier design, studies of various physical phenomena in alloys, radar development, solid-state physics, magnetism, and semiconductors.

In 1942, Dr. Shockley was assigned to Columbia University Division of War Research on an O.S.R.D. contract as Director of Research for the Antisubmarine Warfare Operations Research Group. In 1944 he transferred to the War Department in the capacity of expert consultant in the Office of Secretary of War, and the following year, he returned to Bell Telephone Laboratories.

Dr. Shockley was awarded the Medal for Merit in 1946, and in 1951 was elected to the National Academy of Sciences. He won the I.R.E.'s Morris Liebmann Memorial Prize in 1951 "in recognition of his contribution to the creation and development of the transistor." He is a member of A.P.S., Tau Beta Pi, and Sigma Xi.

❖

Bernard N. Slade (M'50) was born on December 21, 1923 in Sioux City, Iowa. He received the B.S. degree in electrical engineering from the University of Wisconsin in 1948.

He joined the RCA Victor Division immediately after graduation and became associated with the Advanced Development Group at Harrison, N. J., where he is working on the development of crystal triodes. At the same time he is doing graduate work in electrical engineering at Stevens Institute of Technology.

BERNARD N. SLADE

Mr. Slade is an associate member of the A.I.E.E.

❖

Earl L. Steele (SM'48, A'52) was born in Denver, Colo. in 1923. After receiving the B.S. in physics from the University of Utah, Salt Lake City, in 1945, he continued work in this field at Cornell University, Ithaca, N. Y., receiving his Ph.D. degree in 1952.

While at Cornell, he taught elementary physics courses for engineers and pursued research activities on dielectric properties of luminescent materials and solid-state theory.

EARL L. STEELE

Taking a leave of absence from Cornell in 1951, he joined the General Electric Electronics Laboratory where he is presently engaged in research on semiconductor physics and devices.

Dr. Steele is a member of the American Physical Society, American Association of Physics Teachers, Sigma Xi, and Gamma Alpha Graduate Scientific Fraternity.

❖

Otmar M. Stuetzer (A'50) was born in Nuremberg, Germany in 1912. He received the lower degrees in mathematics and physics at the University and the Technische Hochschule in Munich, working part time as an instructor.

From 1936 to 1945 Dr. Stuetzer was employed by the Drahtlos-telegraphische Versuchsstation Graefelfing, later taken over by the Flugfunk-Forschungsinstitut Oberpfaffenhofen. From 1940 he served there as

director of the general science and the microwave departments. He was awarded the D.Sc. degree in 1938 and the Dr. habil. degree in 1943, both from the Technische Hochschule in Munich. In parallel with his research work he lectured, and for a year held a responsible position with the German Research Council. In 1941 he received a Lilienthal Society award for microwave work.

O. M. STUETZER

Since 1946 Dr. Stuetzer has been working for the U. S. Air Force at Wright-Patterson Air Force Base, Ohio.

❖

William E. Taylor was born in Worland, Wyo. on November 15, 1921. He received the B.S. degree in metallurgical engineering in 1942 and the Ph.D. degree in 1950, both from Purdue University.

From 1943 to 1946 Dr. Taylor served in the U. S. Army Field Artillery. He was then engaged as a metallurgist at the Oak Ridge National Laboratory from 1950 to 1952, participating in an investigation of the effects of neutron irradiation on solids. He is now employed at the Phoenix Research Laboratory of Motorola, Inc.

W. E. TAYLOR

Dr. Taylor is a member of the American Institute of Mining and Metallurgical Engineers, A.P.S., and Sigma Xi.

❖

Donald E. Thomas (A'47–SM'47) is a member of the Transmission Development Department of Bell Telephone Laboratories, where he is currently engaged in the development of transistor devices for special applications.

Mr. Thomas was born on May 12, 1907, in Hanover Township, Pa. He joined Bell Telephone Laboratories in 1929 after graduating from Pennsylvania State College with a B.S. degree in electrical engineering. In 1932 he was awarded the M.A. degree by Columbia University.

D. E. THOMAS

His first assignment was in submarine telephone-cable development. Just prior to World War II he became engaged in the development of sea and airborne radar, and continued in this work until he left for military duty in 1942. Upon rejoining the Labo-

ratories in 1946, Mr. Thomas was active in the development and installation of the first deep-sea repeatered submarine telephone cable—between Key West and Havana—which went into service in 1950.

During World War II Mr. Thomas was made a member of the Joint and Combined Chiefs of Staff Committees on Radio Countermeasures, and is now a civilian member of the Research and Development Board Panel on Electronic Countermeasures.

Mr. Thomas is a member of Tau Beta Pi and Phi Kappa Phi.

❖

Robert L. Trent (M'45) was born in New York City on March 21, 1919. He received the B.S. degree in E.E. and the M.S. degree in E.E. from Columbia University in 1941 and 1946, respectively.

In 1941 Mr. Trent joined Bell Laboratories, where until 1947 he was concerned almost exclusively with war developments, taking part in development of sonar, airborne radar, and pulse-

R. L. TRENT

code-modulation communication equipment. During the period from 1947 to 1949 he worked on various aspects of short and long haul radio-relay television links. He is now engaged in the development of transistor circuits and devices.

Mr. Trent is a member of the A.I.E.E.

❖

L. B. Valdes (A'46) was born in Marianao, Cuba, on June 12, 1928. He moved to New Orleans for his high school and college education, and received the B.E. degree in electrical engineering from Tulane University in 1946. That fall he began graduate studies in electrical engineering at Northwestern University, where he remained for three years. He received the M.S. degree there in 1947.

L. B. VALDES

Mr. Valdes next joined the transistor-development group at Bell Telephone Laboratories, where he has been a worker in the field since 1949.

Mr. Valdes is a member of Eta Kappa Nu and Sigma Xi.

❖

Robert L. Wallace, Jr., was born in Callina, Tex. on February 21, 1916. He attended Westminster Junior College and then the University of Texas, from which he received the B.A. degree in physics and mathematics in 1936, followed by the M.A.

in physics in 1939. He then attended Harvard University, where he combined teaching and studying. From 1941 to 1945 he was a special research associate at Harvard, in the field of military communications.

In 1946 Mr. Wallace joined Bell Telephone Laboratories, Inc., where he has been concerned with some problems in magnetic recordings and with transistors. He is a member of

R. L. WALLACE, JR.

the Acoustical Society of America, Phi Beta Kappa, and Sigma Xi.

❖

Maynard C. Waltz is engaged in electronic apparatus development at Bell Telephone Laboratories, especially that work related to germanium and silicon varistors.

Mr. Waltz was born in Damariscotta, Maine on August 19, 1916. He was graduated from Colby College in 1938 with a B.A. degree in physics and two years later received the M.A. degree in physics from Wesleyan University,

M. C. WALTZ

Middletown, Conn.

From 1940 to 1942 he taught physics at Wesleyan, and from 1942 to 1946 worked in the Radiation Laboratory at the Massachusetts Institute of Technology. In 1946 he joined Bell Telephone Laboratories.

Mr. Waltz is one of the authors of "Radar Receivers," a volume in the Radiation Laboratory Series. He is a member of Phi Beta Kappa, Sigma Xi, and Sigma Pi Sigma.

❖

Robert E. Yaeger was born on March 1, 1921, in Chicago, Ill. After receiving the B.S. degree from Worcester Polytechnic Institute in 1942, he joined Bell Laboratories, where his first assignment was with the trial installation department. He engaged in the design of military equipment in 1943 and 1944, when he was called to active duty by the U. S. Naval Reserve. As a radio-radar officer, he served in the

ROBERT E. YAEGER

Asiatic Pacific Theatre until 1946.

Returning to Bell Laboratories, Mr. Yaeger designed voice-frequency and carrier-telephone equipment and, later, worked on the development of program-transmission circuits. He became a member of the transistor development group in 1950, first studying applications of transistors to military use, and currently to carrier telephony.

Institute News and Radio Notes___

LANDMARKS IN ELECTRONICS
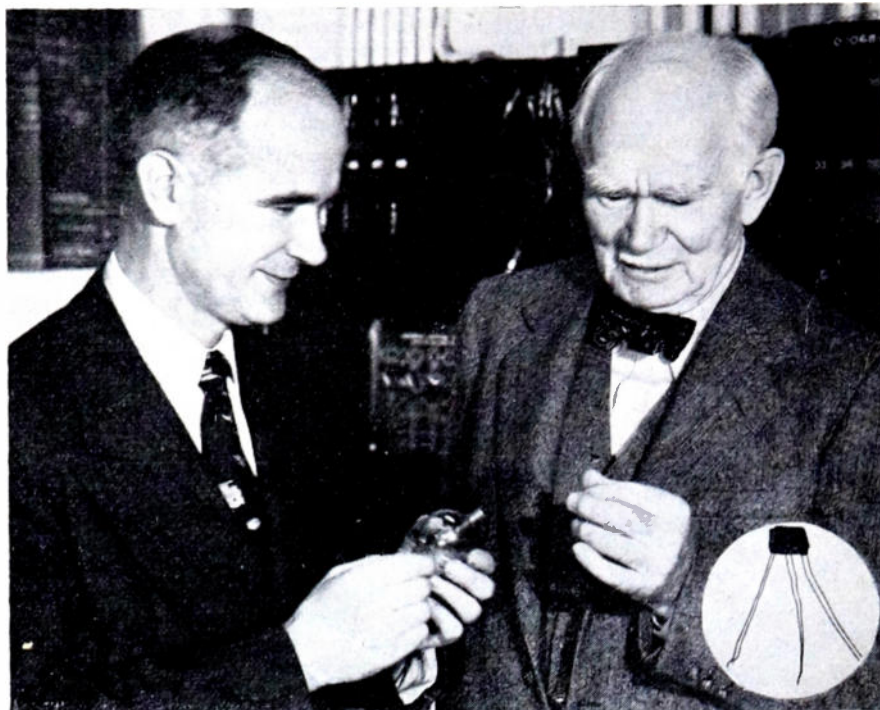
Two major landmarks in the history of electronics are represented by the devices in the hands of the men pictured above. On the left is Dr. William Shockley of Bell Telephone Laboratories, who directed the research program leading to the invention of the transistor. On the right is Dr. Lee De Forest, who in 1907 invented the audion, forerunner of modern vacuum tubes, and the cornerstone of modern electronics. In this photograph Dr. Shockley is holding Dr. De Forest's audion and Dr. De Forest holds the transistor. (A closeup of the transistor is shown in the inset at right.)

COMPUTER CONFERENCE TO STRESS INPUT-OUTPUT

A Joint IRE-AIEE-ACM Computers Conference will be held December 10–12, at the Hotel Park Sheraton, New York, N. Y. The Professional Group on Electronic Computers will be the IRE sponsor.

The following 19 papers will be presented successively to assure that no paper will be missed because of simultaneous presentation.

Wednesday, December 10

"Recording Techniques for Digital Coded Data," A. Tyler, Eastman Kodak Co.

"Converters for Teletype Tape to IBM Cards," IBM

"Cards to Magnetic Tape Converters for UNIVAC," E. Blumenthal and F. Lopey, Rem. Rand Inc.

"Devices for Transporting the Recording Media," R. L. Snyder, R. L. Snyder Co.

"Buffering Between the Input-Output and the Computer," A. L. Leiner, NBS

Thursday, December 11

"SEAC Input-Output System," Sidney Grunwald, J. L. Pike, P. A. Mantek, Ernest Ainsworth, NBS

"Univac Input-Output System," H. F. Welsh, H. Kukoff, L. D. Wilson, E. Roggenstein, E. Masterson, Rem. Rand Inc.

"RAYDAC Input-Output System," Walter Gray, Kenneth Rehler, Louis Fern,

Frank Dean, Raytheon Mfg. Co.

"Type 701 Input-Output Organization," L. D. Stevens, IBM

"Type 726 Magnetic Tape Reader and Recorder," W. S. Buslik, IBM

"Magnetic Tape Technique and Performance," H. W. Nordyke, IBM

Friday, December 12

"Survey of Analog-to-Digital Data Converters," H. E. Burke, Consol. Eng. Corp.

"Survey of Mechanical Type Printers," J. Hosken, A. D. Little, Inc.

"Survey of Nonmechanical Type Printers," R. J. Rossheim, Geo. Wash. Univ.

"Analex Printer," Leo Rosen, Anderson-Nichols Co.

"Eastman Printer," Russel Thompson, Eastman Kodak Co.

"Ferranti Input-Output Equipment," B. W. Pollard, Ferranti Ltd., London

"Garment Tag Equipment," O. G. Hessler, Sears Roebuck and Co.

"Numerically Controlled Machine Tool," W. Pease, M.I.T.

"Summary and Forecast," S. N. Alexander, NBS

At-the-door registration fee for the conference is $3.50 and advance registration is $3.00. Registration cards and hotel reservation forms may be obtained from A. C. Holt, Chairman of the Registration Committee, International Business Machines Corp., 590 Madison Ave., New York, N. Y.

All checks should be made payable to Ralph R. Batcher.

Technical Committee Notes

The **Standards** Committee met on September 11, under the Chairmanship of A. G. Jensen. Mr. Jensen stated that the Executive Committee had deferred action on the Proposed Standards on Receivers: definitions of terms pending the final report on the noise figure (noise factor) definitions. Mr. Jensen reported that he had written to the chairman of the AIEE standards committee in regard to the "Feedback Control Systems Report," published in the October, 1951 issue of *Electrical Engineering*, and had received assurance that the situation will be discussed at their next standards meeting. A. G. Clavier desired clarification on the procedure to be followed by his Symbols Committee in connection with the Proposed American Standard—Device Function Letter Designations for Electric Control Equipment. The proposed standard covers three questions: I—Designations for Identification of Electric and Electronic Parts (Reference-Designations); II—Designations for Indicating Functions of Electric Control Devices; III—Designations for Identification of Telephone, Telegraph, Computer, and Related Equipment. Since the standard has not reached ballot stage, Mr. Jensen suggested that Parts I and II be discussed at the next meeting of the Symbols Committee and that the Standards Committee be informed of any decision. The Committee feels that the last section of the proposed standard might be circulated for the general information of interested parties. The next item on the agenda was the discussion of the definition Noise Factor (Noise Figure), also Noise Factor (Noise Figure), Average, and Noise Factor (Noise Figure), Spot. With the approval of these definitions, the Receivers Standard will be resubmitted to the Executive Committee. The Committee then discussed the supplement to Standard 51 IRE 17. S1 "Practical Consideration in Measuring VHF Receiver Oscillation Radiation." R. F. Shea reviewed the background work on this proposed standard which is to be considered as a supplement to 51 IRE 17. S1 and made available in mimeographed form. There is no necessity for printing this in the Proceedings although a notice of its availability may appear. Some changes were made by the Standards Committee members concerning this proposed standard and it will be submitted to the Executive Committee at an early date. Following was a discussion of the Proposed Modulation Definitions (51 IRE 11. PS1). A portion of these definitions were considered and the remaining terms were deferred until the next Committee meeting. Reports were submitted concerning ASA subcommittees C42.1 on General Terms, C42.13 on Communications, and C42.14 on Electron Tubes by J. G. Brainerd, J. C. Schelleng, and L. S. Nergaard.

The **Antennas and Waveguides** Committee, under the Chairmanship of D. C. Ports, met on September 10. The Chairman asked for qualified men to edit the material for the antenna section of the Annual Review. Suggestions were made by the Committee. There was some discussion as to membership and the future work of the West Coast subcommittee. The remainder of the meeting was devoted to a review of the Definitions of Wave Guide Terms.

Under the Chairmanship of M. R. Briggs, the **Radio Transmitters** Committee met on September 5. A report was submitted by J. B. Heffelfinger, Chairman of Subcommittee on Double Sideband AM Transmitters. He stated that an informal meeting of his subcommittee had been held at the NARTB Convention and that some progress had been made on the proposed Standard on Double Sideband Transmitters: Methods of Testing. Specific tasks had been assigned to members of the subcommittee, and these reports, when correlated, will be submitted as a draft for consideration at the next Committee meeting. Harold Goldberg, Chairman of the Subcommittee on Pulse Modulated Transmitters, advised the Committee that work on part I of the pulse definitions has been completed and that a final report soon will be ready for submission to the main Committee. A. E. Kerwien, Chairman of the Subcommittee on Single Sideband Radio Communication Transmitters, reported that a list of definitions has been prepared by his subcommittee and will be reviewed for submission to the main Committee. The Committee then turned its attention to a discussion on the Radio Transmitters' contribution to the Annual Review.

The **Audio Techniques** Committee met on September 10, under the Chairmanship of C. A. Cady. A good part of the meeting was spent in a detailed discussion by the Committee on a list of Audio Techniques Definitions dated September 24, 1951. This list was approved by the Committee and is to be sent on the "Grand Tour."

IRE Award Presentation

The speakers table at the Western IRE Convention Luncheon during presentation of the new Pacific Region IRE Electronics Achievement Award. From left to right, Commissioner Hyde of the Federal Communications Commission; J. M. Pettit, Stanford University and recipient of the Award; R. L. Sink, IRE Director of Region 7; A. M. Zarem, Stanford Research Institute; and, H. R. Lubke, Consulting Engineer, Hollywood, Calif.

Electronics Achievement Award Established in Pacific Region

The first "Electronic Achievement Award of Pacific Region IRE" was presented to Joseph L. Pettit, by the Director of IRE Region 7 (and instigator of the award), R. L. Sink, during a luncheon held at the Western IRE Electronic Show and Convention, Long Beach, Calif., August 27–29.

The award, which was established in 1951 by the Pacific Regional Committee, is presented annually during an IRE Convention or Conference. The recipient is selected from among the 5,000 IRE members on the West Coast and is to have made important contributions to the electronic arts and industry in one or more of the general categories of education, research and invention, system engineering, contributions to the literature, or contributions to IRE activities.

The qualifications of the recipient are judged by a committee consisting of two representatives from each IRE Section within the Pacific Region, thus enabling the entire West to participate in the selection. One of the features of the award is to give early recognition to accomplishments which had not previously received national attention.

As the first to receive the award, Dr. Pettit, a Senior Member of the Institute and associate professor of electrical engineering at Stanford University, was cited for his, "Multiple contributions to the electronic arts in the fields of education and high-frequency communications and through his leadership in Institute affairs."

Dr. Pettit also received the Presidential Certificate of Merit in 1949 for his work during World War II with the government's Radio Research Laboratory and his contributions in the field of electronics.

A native of Minnesota, Dr. Pettit received the B.S. degree from the University of California and the E.E. and Ph.D. degrees from Stanford University. He has worked closely with the government in directing the development of high-frequency receivers used in combating enemy radar, and in 1944, he served in the China-Burma-India Theater instructing in their use.

Dr. Pettit has been a member of the Executive Committee of the San Francisco IRE Section, Chairman of the Palo Alto Subsection, and General Chairman of the West Coast IRE 1949 Convention.

Professional Group News

ANTENNAS AND PROPAGATION

The San Diego Section has been holding successful Antennas and Propagation Group Meetings, J. L. Heritage as Acting Chairman. The Group plans to hold a technical meeting lecture and demonstration of radar backscatter sounding for determination of maximum usable frequency.

AUDIO

The Audio Group of the San Diego Section, now sponsored by F. X. Byrnes and Matt Brady of the Navy Electronics Laboratory, has held two successful meetings recently. The Group met with officials at the Western IRE Convention to discuss the affiliation of the San Diego Chapter with the National Group.

During the Chicago IRE Section's meeting, the Audio Group session was represented with a talk by B. B. Bauer, on "Electroacoustic Analogies for the Radio Engineer."

BROADCAST AND TELEVISION RECEIVERS

The "Turret Tuner for Television Receivers," by M. G. Beier, J. F. Bell, Albert Cotsworth, and J. F. White, was given at the Broadcast and Television Receivers Group session during the recent meeting of the Chicago IRE Section.

CIRCUIT THEORY

At the first meeting held by the Los Angeles Chapter of the Circuit Theory Group at the Institute for Numerical Analysis, W. R. Evans spoke on, "The Root Locus Method."

The speaker for the November meeting will be Louis Weinberg who will give an expository talk on network synthesis.

The Circuit Theory Group session at the Chicago Section meeting sponsored a talk on "Unusual Extensions of Electromechanical Equivalence," by Robert Adler.

COMMUNICATIONS

The title of the IRE Professional Group on Communications has been officially changed to the IRE Professional Group on Communications Systems.

ELECTRONIC COMPUTERS

Recent meetings of the Los Angeles Chapter Professional Group on Electronic

Computers have been held at the Institute for Numerical Analysis, National Bureau of Standards, Westwood, Calif., with H. T. Larson presiding officer. Among the papers presented at these meetings have been, "Recent Developments in the Digital Computer Activity of the University of California," by P. L. Morton; "The Selection," by W. F. Gunning; "Navy Computers Acceptance Tests," by D. E. Dufford; and, "Test Programs for the SWAC," by Roselyn Lipkis. A demonstration of test routines on the SWAC also was given.

M. M. Astrahan Chairman of the Electronics Computers Group has announced that Walter Hamstead has been appointed Sectional Activities Chairman. Mr. Hamstead will promote the organization of Group Chapters in New York City, Washington, D. C., and Boston.

INSTRUMENTATION

The Professional Group on Instrumentation, in conjunction with the Instrument Society of America, sponsored a session at the National Instrument Conference and Exhibit, September 8–12, 1952, Public Auditorium, Cleveland, Ohio.

A successful program by the IRE Group was attended by over 150 persons to hear the papers and discussions presented. Among the papers presented were: "A Peak-reading, Single-pulse Voltmeter," by E. F. Carome; "Surface Finish Measurement with Electronics," by D. R. Christian; "A Prescheduled Automatic Actuator Control System," by W. T. Harrigill, Jr.; "Geophysical Instrumentation," by F. S. Segesman; "Sound and Vibration Instrumentation," by Frank Mass; "Computers and Servomechanisms," by H. R. Hegbar; and, "High-Speed Recording," by S. J. Begun.

The San Diego IRE Section has announced the meetings of an Instrumentation Group under the Chairmanship of William Burnett. One meeting was sponsored by John Day, Junior Past Chairman of the San Diego Section, and D. C. Kalbfell, Student Advisor. A symposium on subminiaturization also was held. The Group plans to hold four meetings per year.

INDUSTRIAL ELECTRONICS

A symposium, "Industrial Electronic Transducers," by James Everett, Dave

Elam, and Gordon Kenett, was featured by the Group during the Chicago IRE Section's Meeting held in September.

MICROWAVE CIRCUITRY

The San Diego Section has shown an interest in the formation of a chapter of the Microwave Circuitry Group. The Group may combine with the Antennas and Propagation Group in its formation.

VEHICULAR COMMUNICATIONS

The Executive Committee has approved the formation of the Washington Chapter of the Professional Group on Vehicular Communications.

With this Group in attendance the Washington IRE Section held a meeting recently at which a paper was presented by Jacob Rabinow of NBS, entitled, "Information Storage Devices."

Vehicular Communications Meeting

The third Annual Meeting on Vehicular Communications will be held December 3–5, 1952, at the Statler Hotel, Washington, D. C. The Washington IRE Section will be host to the meeting, and the newly formed Washington Professional Group Chapter on Vehicular Communications and the National Group will act as sponsors.

The theme will be "Spectrum Conservation" Papers will be presented on vehicular and kindred subjects, supplemented with a great variety of exhibits by manufacturers.

A luncheon and an evening banquet featuring prominent speakers in the vehicular field will highlight the social program. An interesting program for the ladies will include tours, teas, and style shows.

W. A. Shipman, P.O. Box 215, Falls Church, Va., is Chairman for the conference. He will be ably assisted by Colonel E. White and Joseph Kelly, Arrangements Chairmen; Roland Davies and Robert Tall, Publicity Chairmen; Granville Klink, Exhibits Chairman; Mark Swanson, Chairman of the Washington IRE Section; Tom Jacocks, Chairman of the Ways and Means Committee.

IEE PLANS SPRING SYMPOSIUM

The Institute of Electrical Engineers will hold a symposium dealing with the selection and use of dielectrics, March 16–18, 1953, Institute of Electrical Engineers Building, London, England.

With the presentation of approximately 50 papers anticipated, the provisional plan for the subsequent sessions is as follows: permittivity and dielectric losses in solids; permittivity and dielectric losses in liquids; electric strength and breakdown mechanisms; mechanical, physical, chemical properties, and their stability; current problems in insulation design for high frequencies and lower frequencies; classification, specification, and testing.

TRANSACTIONS OF IRE PROFESSIONAL GROUPS

The following issues of TRANSACTIONS have recently been published by IRE professional Groups and additional copies are available from the Institute of Radio Engineers, Inc., 1 East 79 Street, New York 21, N. Y., at the prices listed below.

Sponsoring Group	Publication	Group Members	IRE Members	Non-members*
Airborne Electronics	PGAE-4; "The Selectivity and Intermodulation Problem in UHF and Communication Equipment" (12 pages)	$0.45	$0.65	$1.35
Audio	PGA-9; September–October Issue (28 pages)	0.60	0.90	1.80

* Public libraries and colleges can purchase copies at IRE Member rates.

IRE People

William C. Ballard (SM'46), professor of electrical engineering at Cornell University, died recently.

Professor Ballard was born on September 1, 1888, in Baltimore, Md. After studying at Baltimore City College, he attended Cornell University where he received the M.E. degree in 1910, in electrical engineering. He then was placed in charge of communications and electronics courses at Cornell, and was responsible for the beginning and development of that department.

In addition to his appointment at Cornell, Professor Ballard was engaged as an engineering consultant for various companies, among them Atwater Kent, Westinghouse, Fox Film, Western Electric, Sylvania Electric Products, Inc., and the Automatic Signaling Company.

Professor Ballard was the Vice-Chairman of the IRE Professional Group on Medical Electronics.

Ralph I. Cole (A'20–SM'46) has joined the engineering staff of Melpar Inc.

Mr. Cole who was born in St. Louis, Mo., on August 7, 1905, received the B.S. degree from Washington University in 1927. He then became associated with the United States Signal Corps as a radio engineer engaged in research and development at the Fort Monmouth Laboratory. Under his direction that the first modern tank radio was developed.

RALPH I. COLE

In 1936, Mr. Cole received the M.S. degree in physics at Rutgers University. He then was put in charge of research and development of radio direction finding and fighter control equipment for the Signal Corps and Air Force, and at the beginning of World War II, he was commissioned a major in the army assigned to the Signal Corps Laboratory. He was promoted to lieutenant colonel in 1944 in charge of all engineering activity concerning development of ground electronics equipment at the Watson Laboratories.

Mr. Cole received the Army commendation Ribbon and Legion of Merit for his efforts in the research and development field, and still holds the rank of colonel in the Air Force Reserve.

Mr. Cole is Chairman of the IRE Professional Group on Engineering Management and a member of the Research and Development Boards Radar Panel. He is the holder of several patents on improvement in radio direction finding systems and special radio components.

Benedict K. V. French (A'24–M'30–SM'43) has become a member of the field engineering staff of the General Instrument Corporation.

Mr. French began his radio career in 1923 as a development engineer with the Federal Telegraph and Telephone Company. He held subsequent positions with the American Bosch Company, P. R. Mallory Company, Howard W. Sams Company, RCA's License Division Laboratory, and the Allen B. DuMont Laboratories, Inc. He was responsible for the introduction of push-button station selection and wave-band switching.

During World War I, Mr. French served on the joint Army-Navy Standardization Board, and in World War II, he was the supervisor for Mallory research in the development of a mercury-type dry battery, extensively used in the armed forces radio equipment.

Mr. French has served as Chairman of the IRE Connecticut Valley and Indianapolis Sections and is a member of the IRE Professional Group on Airborne Electronics.

❖

Alfred N. Goldsmith (M'12–F'15), a Founder of The Institute of Radio Engineers, has been awarded the Radio Pioneers 1952 Special Citation for Outstanding Services and Contributions to Radio.

The award was cited to Dr. Goldsmith, "who, through continuing and essential contributions over a period of forty years to the growth and development of radio, broadcasting and television, has made the growth of these industries a monument not only to science, but to himself," and was signed by H. F. Kaltenborn, Founder, and Carl Haverlin, President.

Dr. Goldsmith, Editor of the PROCEEDINGS of the I.R.E., is well known for his work as a consulting engineer and for his inventions in the radio, phonograph, TV, color TV, and motion picture fields.

❖

Ernest A. Marx (A'44) has been named vice president of DuMont Television and Electronics, Ltd., in Canada. He is the director of the company's international division.

Mr. Marx, who is a native of New York, received the B.S. degree in electrical engineering in 1917, from Columbia University. From 1917–1920, he was a radio instructor for the United States Naval Radio School. From 1920–1935, he was a statistician and analyst for companies engaged in engineering activities, and in 1935–1941, he was a technical advisor for American banking houses in Europe and the representative to foreign engineering companies.

Before joining the DuMont Laboratories in New Jersey, Mr. Marx was a radar engineering specialist for the United States Navy Radar Installation and Radar School.

Gordon N. Thayer (SM'47–F'51) has been named a vice president of the Bell Telephone Laboratories, Inc. In his new post, he will be responsible for the laboratories' military development program and its relations with the Army, Navy, and Air Force.

Mr. Thayer was born October 6, 1908, in Delta, Colo., and received the M.E. degree from Stevens Institute of Technology in 1930. He then joined Bell Telephone engaged in the development of mobile radio communications equipment and systems.

G. N. THAYER

In 1940, Mr. Thayer became affiliated with a group developing radar systems, and later turned his attention to microwave radio-relay systems. Since 1949, he has been concerned with the development of communications systems, including the cross-country radio-relay system, the Key West-Havana submarine cable, and overseas radio projects. Mr. Thayer was appointed assistant director of transmission systems development at Bell Telephone in 1949, and director of transmission development in 1951.

❖

Ben Kievit (M'34–SM'43) has been named manager of sales engineers for the radio and television picture tube divisions of Sylvania Electric Products, Inc.

A native of Toledo, Ohio, Dr. Kievit received the B.S. and M.S. degrees from the University of Kentucky in 1924 and 1926, and the Ph.D. degree from the University of Michigan in 1930. While working for his graduate degrees, Dr. Kievit was an instructor in physics.

BEN KIEVIT

Dr. Kievit joined Sylvania in 1930 as a research physicist doing developmental work on photoflash lamps. In 1932 he was made supervisor of the radio tube application department and also edited the *Sylvania Technical Manual* and the *Sylvania Engineering News Letter*.

In 1938, Dr. Kievit was promoted to assistant director of the tube division's commercial engineering department, and in 1940, he was named supervisor of customer services. In 1943, he became field engineer, eastern district, of Sylvania's tube sales department.

Dr. Kievit is an active author of technical articles, technical manuals, and engineering news letters. He is a fellow of the American Association for the Advancement of Science, and a member of the American Physical Society and Sigma Xi.

Sections*

Chairman		Secretary
I. L. Knopp 628 Ecton Rd. Akron, Ohio	AKRON (4)	Buford Smith, Jr. 1831 Ohio Ave. Cuyahoga Falls, Ohio
W. L. Fattig Box 788 Emory Univ., Ga.	ATLANTA (6)	H. W. Ragsdale 654 Cooledge Ave., N.E. Atlanta, Ga.
C. E. McClellan 1306 Tarrant Rd. Glen Burnie, Md.	BALTIMORE (3)	C. D. Pierson, Jr. 1574 Waverly Rd. Baltimore, Md.
L. W. McDaniel 3385 Timberwood Lane Beaumont, Tex.	BEAUMONT- PORT ARTHUR (6)	C. B. Trevey 2555 Pierce St. Beaumont, Tex.
J. H. Merchant 2 Cedar St. Binghamton, N. Y.	BINGHAMTON (4)	R. F. New 654 Chenango St. Binghamton, N. Y.
F. D. Lewis 275 Massachusetts Ave. Cambridge, Mass.	BOSTON (1)	A. J. Pote 71 W. Squantum St. North Quincy, Mass.
I. C. Grant San Martin 379 Buenos Aires, Arg.	BUENOS AIRES	C. A. Cambre Olazabal 5255 Buenos Aires, Arg.
R. T. Bozak 90 Montrose Ave. Buffalo, N. Y.	BUFFALO- NIAGARA (4)	R. R. Thalner 254 Rano St. Buffalo, N. Y.
J. L. Hollis 2900 "E" Ave., N.E. Cedar Rapids, Iowa	CEDAR RAPIDS (5)	J. W. Smith 1136 27, N.E. Cedar Rapids, Iowa
R. M. Krueger 5143 N. Neenah Ave. Chicago, Ill.	CHICAGO (5)	J. J. Gershon 2533 N. Ashland Ave. Chicago Ill.
J. P. Quitter 509 Missouri Ave. Cincinnati, Ohio	CINCINNATI (5)	D. W. Martin Box 319-A, RR 1 Newtown, Ohio
I. L. Hunter 3901 E. Antisdale Rd. S. Euclid, Ohio	CLEVELAND (4)	R. K. Olsen 8042 Brecksville Rd. Brecksville, Ohio
J. H. Jaeger 361 Oakland Park Ave. Columbus, Ohio	COLUMBUS (4)	R. W. Masters 1633 Essex Rd. Columbus, Ohio
John Merrill 16 Granada Terr. New London, Conn.	CONNECTICUT VALLEY (1)	H. E. Rohloff The So. New Eng. Tel. Co. New Haven, Conn.
R. A. Arnett 4073 Rochelle Dr. Dallas, Tex.	DALLAS-FORT WORTH (6)	J. A. Green 6815 Oriole Dr. Dallas, Tex.
J. L. Dennis 3005 Shroyer Rd. Dayton, Ohio	DAYTON (5)	A. B. Henderson 801 Hathaway Rd. Dayton, Ohio
W. R. Bliss 1426 Market St. Denver, Colo.	DENVER (5)	R. E. Swanson 1777 Kipling St. Denver, Colo.
W. L. Cassell Iowa State College Ames, Iowa	DES MOINES- AMES (5)	R. E. Price 1107 Lyon St. Des Moines, Iowa
P. L. Grundy 55 W. Canfield Ave. Detroit, Mich.	DETROIT (4)	E. J. Love 9264 Boleyn Detroit, Mich.
L. R. Maguire 4 E. 6 St. Emporium, Pa.	EMPORIUM (4)	J. B. Grund 104 W. 6 St. Emporium, Pa.
H. L. Thorson General Electric Co. Owensboro, Ky.	EVANSVILLE- OWENSBORO (5)	A. P. Haase 2230 St. James Ct. Owensboro, Ky.
R. B. Jones 4322 Arlington Ave. Fort Wayne, Ind.	FORT WAYNE (5)	J. J. Iffland 1008 Madison St. Fort Wayne, Ind.
Arthur Ainlay RR 6, Mt. Hamilton Hamilton, Ont., Canada	HAMILTON (8)	John Lucyk 77 Park Row S. Hamilton, Ont., Canada
F. L. Mason Electronics Office Naval Shipyard Pearl Harbor, Oahu, T.H.	TERRITORY OF HAWAII (7)	J. W. Anderson 4035 Black Pt. Rd. Honolulu, T. H.
H. T. Wheeler 802 N. Avenue "A" Bellaire, Tex.	HOUSTON (6)	J. K. Hallenburg 1359 DuBarry Lane Houston, Tex.
F. D. Meadows 1449 N. Emerson Indianapolis, Ind.	INDIANAPOLIS (5)	J. T. Watson 2146 Admiral Dr. Indianapolis, Ind.
D. L. Ewing 108-A Byrnes China Lake, Calif.	INYOKERN (7)	F. S. Howell 313-B Tyler St. China Lake, Calif.
D. G. Wilson Univ. of Kansas Lawrence, Kan.	KANSAS CITY (5)	Mrs. G. L. Curtis Radio Industries, Inc. Kansas City, Kan.
G. H. Scott American Radio & TV Inc. No. Little Rock, Ark.	LITTLE ROCK (5)	V. L. Dillaplain 203 S. Pine St. Little Rock, Ark.
R. B. Lumsden 332 Hale St. London, Ont., Canada	LONDON, ONTARIO (8)	J. D. B. Moore 27 McClarey Ave. London, Ont., Canada
W. G. Hodson 10806 Smallwood Ave. Downey, Calif.	LOS ANGELES (7)	B. S. Angwin 238 N. Frederic St. Burbank, Calif.
M. C. Probst Rt. 7, Box 415 Louisville, Ky.	LOUISVILLE (5)	G. W. Yunk 2236 Kaelin Ave. Louisville, Ky.
H. W. Mehrling 365 La Villa Dr. Miami Springs, Fla.	MIAMI (6)	M. C. Scott, Jr. Station WIOD Miami, Fla.
D. E. Mereen 3260 N. 88 St. Milwaukee, Wis.	MILWAUKEE (5)	H. J. Zwarra 722 N. Broadway Milwaukee, Wis.
N. R. Olding Canadian Broadcasting Corp. Montreal, P.Q., Canada	MONTREAL, QUEBEC (8)	R. W. Cooke 17 De Castelneau St. Montreal, P. Q., Canada
C. W. Carnahan 3169-41 Pl., Sandia Base Albuquerque, N. M.	NEW MEXICO (7)	L. E. French 107 S. Washington Albuquerque, N. M.
H. T. Budenbom 82 Wellington Ave. W. Short Hills, N. J.	NEW YORK (2)	A. B. Giordano 85-99 Livingston St. Brooklyn, N. Y.
V. S. Carson N. C. State College Raleigh, N. C.	NORTH CAROLINA- VIRGINIA (3)	J. G. Gardiner 3502 Kirby Dr. Greensboro, N. C.
C. E. Harp 524 E. Macy St. Norman, Okla.	OKLAHOMA CITY (6)	E. G. Crippen 3829 N.W. 23 St. Oklahoma City, Okla.
C. W. Rook Univ. of Nebraska Lincoln, Neb.	OMAHA-LINCOLN (5)	V. H. Wight 1411 Nemaha St. Lincoln, Neb.
E. L. R. Webb 31 Dunvegan Rd. Ottawa, Ont., Canada	OTTAWA, ONTARIO (8)	D. V. Carroll Box 527 Ottawa, Ont., Canada
C. M. Sinnett 103 Virginia Ave. Westmont, N. J.	PHILADELPHIA (3)	S. C. Spielman Walton Rd. Huntingdon Valley, Pa.
R. E. Samuelson 1401 E. San Juan Ave. Phoenix, Ariz.	PHOENIX (7)	Z. F. McFaul 4242 N. 2nd Dr. Phoenix, Ariz.
J. G. O'Shea 30 E. Orchard Ave. Pittsburgh, Pa.	PITTSBURGH (4)	J. H. Greenwood 166 N. Sprague Ave. Pittsburgh, Pa.
L. C. White 3236 N. E. 63 Ave. Portland, Ore.	PORTLAND (7)	G. C. Ellison 11310 S.E. Market St. Portland, Ore.
W. H. Bliss 300 Western Way Princeton, N. J.	PRINCETON (3)	J. S. Donal, Jr. RCA Labs. Princeton, N. J.
Garrard Mountjoy 100 Carlson Rd. Rochester, N. Y.	ROCHESTER (4)	R. N. Ferry 196 Lafayette Pkwy. Rochester, N. Y.
W. F. Koch 1340-33 St. Sacramento, Calif.	SACRAMENTO (7)	H. C. Slater 1945 Bidwell Way Sacramento, Calif.
H. G. Wise 1705 N. 48 St. East St. Louis, Ill.	ST. LOUIS (5)	E. F. O'Hare 8325 Delcrest Dr. University City, Mo.

* Numerals in parentheses following Section designate Region number

Sections ——— | Subsections ———

Chairman		Secretary
Stanley Benson Box 1707 Salt Lake City, Utah	SALT LAKE CITY (7)	M. E. Van Valkenburg Univ. of Utah Salt Lake City, Utah
A. H. LaGrone Box F, Univ. Station Austin, Tex.	SAN ANTONIO (6)	Paul Tarrodaychik 215 Christine Dr. San Antonio, Tex.
	SAN DIEGO (7)	R. A. Kirkman 6306 Celia Dr. San Diego, Calif.
W. E. Noller 1229 Josephine St. Berkeley, Calif.	SAN FRANCISCO (7)	O. J. M. Smith Univ. of Calif. Berkeley, Calif.
E. S. Sampson 1243 Chrisler Ave. Schenectady, N. Y.	SCHENECTADY (2)	D. E. Norgaard 1908 Townsend Rd. Schenectady, N. Y.
J. E. Hogg 4107 Sunnyside Ave. Seattle, Wash.	SEATTLE (7)	H. M. Swarm Univ. of Washington Seattle, Wash.
Samuel Seely Syracuse University Syracuse, N. Y.	SYRACUSE (4)	W. H. Hall Gen. Elec. Co. Syracuse, N. Y.
W. M. Stringfellow 136 Huron St. Toledo, Ohio	TOLEDO (4)	G. H. Eash 845 W. Woodruff Ave. Toledo, Ohio
G. E. McCurdy 74 York St. Toronto, Ont., Canada	TORONTO, ONTARIO (8)	Clive Eastwood 658 Pharmacy Ave. Dawes Rd. P.O. Toronto, Ont., Canada
W. D. Fuller 73 E. 51 Pl. Tulsa, Okla.	TULSA (6)	W. J. Weldon 2530 E. 25 St. Tulsa, Okla.
O. A. Schott 4224 Elmer Ave. Minneapolis, Minn.	TWIN CITIES (5)	F. S. Hird 224 S. 5th St. Minneapolis, Minn.
A. H. Gregory 150 Robson St. Vancouver, B. C. Canada	VANCOUVER (8)	Miles Green 2226 W. 10th Ave. Vancouver, B. C. Canada
M. W. Swanson 1420 Mt. Vernon Memorial H'way Alexandria, Va.	WASHINGTON (3)	T. B. Jacocks 777 14 St., N.W. Washington, D. C.
J. H. Canning 1701 Chestnut St. Williamsport, Pa.	WILLIAMSPORT (4)	R. C. Lepley R.D. 2 Williamsport, Pa.

Chairman		Secretary
R. F. Lee 2704-31 St. Lubbock, Tex.	AMARILLO-LUBBOCK (6) (Dallas-Ft. Worth Subsection)	C. M. McKinney Texas Tech. College Lubbock, Tex.
Carl Volz 160 W. Hamilton Ave. State College, Pa.	CENTRE COUNTY (4) (Emporium Subsection)	R. L. Riddle Penn. State College State College, Pa.
F. G. McCoy Rt. 4, Box 452-J Charleston, S. C.	CHARLESTON (6) (Atlanta Subsection)	C. B. Lax Sergeant Jasper Apts. Charleston, S. C.
D. E. Reynolds 4116 Memphis St. El Paso, Tex.	EL PASO (7) (New Mexico Subsection)	J. E. Hoefling Box 72, Ft. Bliss El Paso, Tex.
S. L. Johnston 207 Edgewood Dr. Huntsville, Ala.	HUNTSVILLE (6) (Atlanta Subsection)	R. C. Haraway 603 College Hill Apts. Huntsville, Ala.
L. B. Headrick RCA Victor Div. Lancaster, Pa.	LANCASTER (3) (Philadelphia Subsection)	C. G. Landis Safe Harbor, Box 6 Conestoga, Pa.
C. J. Hirsch 49-23 Forest Dr. Douglaston, N. Y.	LONG ISLAND (2) (New York Subsection)	B. F. Tyson 4916 Douglaston Pkwy. Douglaston, N. Y.
R. T. Blakely Merry Hill Titusville R'., Poughkeepsie, N. Y.	MID-HUDSON (2) (New York Subsection)	
S. D. Robertson Box 107 Red Bank, N. J.	MONMOUTH (2) (New York Subsection)	G. F. Senn 81 Garden Rd. Little Silver, N. J.
A. G. Richardson 180 Vreeland Ave. Boonton, N. J.	NORTHERN N. J. (2) (New York Subsection)	P. S. Christaldi Box 111 Clifton, N. J.
O. G. Villard, Jr. 2050 Dartmouth St. Palo Alto, Calif.	PALO ALTO (7) (San Francisco Subsection)	J. V. Granger 772 Paul Ave. Palo Alto, Calif.
A. A. Kunze Lee Center New York, N. Y.	ROME (4) (Syracuse Subsection)	J. M. Thompson Box 1245 Haselton Br. P.O. Rome, N. Y.
George Weiler 1429 E. Monroe South Bend, Ind.	SOUTH BEND (5) (Chicago Subsection)	A. R. O'Neil 1525 N. Adams St. South Bend, Ind.
H. G. Swift Rte. 2 Derby, Kan.	WICHITA (Kansas City Subsection)	P. A. Dunbar 1328 N. Lorraine Wichita, Kan.
R. F. Tinkler 166 Portage Ave., E. Winnipeg, Canada	WINNIPEG (8) (Toronto Subsection)	H. R. Gissing 65 Rorie St. Winnipeg, Canada

Professional Groups ———

Chairman	
AIRBORNE ELECTRONICS	George Rappaport Wright Field, Dayton, Ohio
ANTENNAS AND PROPAGATION	A. H. Waynick Pennsylvania State College State College, Pa.
AUDIO	J. J. Baruch Massachusetts Institute o fTechnology Cambridge, Mass.
BROADCAST AND TELEVISION RECEIVERS	D. D. Israel 111 8 Ave. New York, N. Y.
BROADCAST TRANSMISSION SYSTEMS	Lewis Winner 52 Vanderbilt Ave. New York, N. Y.
CIRCUIT THEORY	R. L. Dietzold 34 W. 11 St. New York, N. Y.
COMMUNICATIONS SYSTEMS	G. T. Royden 67 Broad St. New York, N. Y.
ELECTRON DEVICES	George D. O'Neill Sylvania Electric Products, Inc. Bayside, L. I., N. Y.
ELECTRONIC COMPUTERS	M. M. Astrahan I.B.M. Plant. no. 2 Poughkeepsie, N. Y.
ENGINEERING MANAGEMENT	Ralph I. Cole Griffiss A.F.B., Rome, N. Y.

Chairman	
INDUSTRIAL ELECTRONICS	Eugene Mittlemann 549 W. Washington Blvd., Chicago, Ill.
INFORMATION THEORY	Nathan Marchand Sylvania Electric Rroducts Inc. Bayside, L. I., N. Y.
INSTRUMENTATION	I. G. Easton General Radio Co. Cambridge, Mass.
MEDICAL ELECTRONICS	L. H. Montgomery, Jr. Vanderbilt University Nashville, Tenn.
MICROWAVE THEORY AND TECHNIQUES	Ben Warriner Gen. Elec. Co., Advanced Electronics Center, Cornell Univ., Ithaca, N. Y.
NUCLEAR SCIENCE	L. R. Hafstad Atomic Energy Comm. Rm. 132, 1901 Constitution Ave. Washington, D. C.
QUALITY CONTROL	Leon Bass General Elec. Co. Schenectady, N. Y.
RADIO TELEMETRY AND REMOTE CONTROL	M. V. Kiebert, Jr. Bendix Aviation Corp. Teterboro, N. J.
VEHICULAR COMMUNICATIONS	F. T. Budelman Budelman Radio Corp. Stamford, Conn.

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
Wireless Engineer, London, England

NOTE: The Institute of Radio Engineers does not have available copes 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 I.R.E.

The number in heavy type at the upper 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.

ACOUSTICS AND AUDIO FREQUENCIES

016:534 **2675**
References to Contemporary Papers on Acoustics—R. T. Beyer. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 330–334; May, 1952.) Continuation of 2406 of October.

534.2 **2676**
Connection between the Problem of Acoustic Diffraction at a Sphere and the Reciprocity Theorem—S. N. Rzhevkin. (*Zh. Tekh. Fiz.*, vol. 21, pp. 1224–1227; October, 1951.) It is shown by solving the diffraction problem that the sound pressure at a remote point *B* caused by a radiator at a point *P* on the surface of a fixed rigid sphere is equal to that which would be caused at the point *P* by a radiator of the same output at the point *B*.

534.231 **2677**
A More Rapidly Convergent Expansion for the Velocity Potential of a Piston Source—H. Elrod, Jr. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 325–326; May, 1952.) A variation of the expansion given by Carter and Williams (1815 of 1951) is presented which is particularly useful for calculations in the paraxial regions.

534.231:534.321.9 **2678**
The Acoustical Field Near a Circular Transducer—E. W. Guptill and A. D. MacDonald. (*Canad. Jour. Phys.*, vol. 30, pp. 119–122; March, 1952.) An approximate formula is derived for the case of a circular diaphragm clamped at its periphery. Particle velocity in the near field is greater than that for a plane wave; in the case of a crystal of 1-cm radius operating at about 1 mc in water, the difference is roughly 0.1 per cent.

534.232+534.321.9 **2679**
➤➤ **New Ultrasonic and Sonic Generators**—R. Göbel. (*Nachrichtentechnik*, vol. 2, pp. 7–11; January, 1952.) Illustrated review of many different types of generator available commercially.

534.232 **2680**
Radiation Loading of Cylindrical and Spherical Surfaces—M. C. Junger. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 288–289; May, 1952.) Series of curves are presented for the acoustic reactance and resistance ratios of partial waves emitted by cylindrical and spherical sources whose dynamic configuration can be expressed by infinite series.

534.232:534.141.4 **2681**
Acoustical Characteristics of Jet-Edge and Jet-Edge-Resonator Systems—W. L. Nyborg, M. D. Burkhard and H. K. Schilling. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 293–304; May, 1952.)

534.24 **2682**
The Reflection of a Pulse by a Spherical Surface—D. V. Anderson, T. D. Northwood and C. Barnes. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 276–283; May, 1952.) The Kirchhoff integral is applied to obtain an approximate solution for the reflection of a pulse back to a receiver near the source. For a minimum-time path the reflected pulse develops a tail, and for a maximum-time path a head. Experiments confirmed this.

534.24 **2683**
Reflection of a Pulse by a Concave Paraboloid—D. V. Anderson. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 324–325; May, 1952.) The existence of a secondary pressure maximum about $3f/4$ from the axis ($f=$focal length of reflector), as predicted by Friedlander (98 of 1943), is confirmed experimentally.

534.24:534.321.9 **2684**
On the Failure of Plane-Wave Theory to predict the Reflection of a Narrow Ultrasonic Beam—M. S. Weinstein. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 284–287; May, 1952.) Reflection measurements at 3.35 mc in water for air-backed Al plates of thicknesses from 0.25 to 0.025 inch and for an Al plate 2 inches thick indicate that plane-wave theory is not strictly applicable to the reflection of a narrow beam. The excess pressure of the reflected wave is considerably lower than that predicted by the plane-wave theory when the angle of incidence is such that the change of phase on reflection varies greatly with a small change of the angle of incidence.

534.26+535.43 **2685**
Multiple Scattering of Radiation by an Arbitrary Planer Configuration of Parallel Cylinders and by Two Parallel Cylinders—V. Twersky. (*Jour. Appl. Phys.*, vol. 23, pp. 407–414; April, 1952.) The solution previously given (1803 of August) is applied to the particular case where all the axes lie in the same plane. The scattered wave is expressed as an infinite sum of orders of scattering, the first order being the usual single-scattering approximation. Both em and acoustic waves are considered.

534.26 **2686**
On Acoustic Diffraction Cross-Sections for Oblique Incidence—J. W. Miles. (*Jour. Acous. Soc. Amer.*, vol. 24, p. 324; May, 1952.) The variational method of Schwinger furnishes a convenient means for relating the plane-wave transmission cross-section of an aperture for oblique incidence to that for normal incidence. Bouwkamp's results (2160 of 1951) are then used to obtain the oblique-incidence transmission cross-section for moderate values of ka, where k is the wave number and a the radius of the aperture.

534.373:534.414 **2687**
The Damping of Acoustic Resonators—K. Voelz. (*Z. angew. Phys.*, vol. 4, pp. 18–19; January, 1952). Addendum to 2087 of 1951.

534.61:621.317.78.029.4 **2688**
Measuring the Mean Power of Varying-Amplitude Complex Audio Waves—H. W. Curtis. (PROC. I. R. E., vol. 40, pp. 775–779; July, 1952.) An instrument for measuring af signals in which the power varies rapidly (e.g., speech sounds) is essentially a computer with circuits performing the operations of squaring, integrating, putting limits on the integral, and taking a logarithm. The result is presented as a continuous strip recording of oscilloscopic deflections, giving readings accurate to within 1 db over a 25-db range.

534.612.4 **2689**
Absolute Calibration of Acoustic Transducers by the Method of Reciprocity in a Tube with Standing Waves—M. V. Kazantseva. (*Zh. Tekh. Fiz.*, vol. 21, pp. 1213–1223; October 1951.) The method described previously (1060 of 1950) was applicable at frequencies up to about 1,500 cps. This range is extended to 10 kc by using a hydrogen-filled tube of diameter 4 cm and length 86 cm. The results are plotted and the theory of the method is discussed.

534.612.4:621.395.61 **2690**
Absolute Method of Calibrating Microphones at Audible and Ultrasonic Frequencies—V. Gavreau, M. Dratz and A. Calaora. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1603–1605; April 16, 1952.) For frequencies up to 40 kc, a source consisting of a solid duralumin cylinder vibrating longitudinally is used [see 2455 of 1941 (St. Clair)]. An optical method is used for measuring the amplitude, the end of the cylinder being polished to form a mirror and the displacement of the image of a razor edge

observed with a stroboscopic arrangement. Tables are given of measurements made in an anechoic room on an electrodynamic and on a piezoelectric microphone.

534.614:533.5　　　　2691
The Velocity of Sound in Gases at Low Pressures—R. L. Abbey.—(*Aust. Jour. Sci. Res., Ser. A*, vol. 5, pp. 223–225; March, 1952.) Report of experiments made to clear up the discrepancy between Maulard's results (3009 of 1949) and those of Abbey and Barlow (*Aust. Jour. Sci. Res., Ser. A*, vol. 1, p. 175; 1948). A tube of diameter 3.5 cm was used. The new results indicate that the discrepancy was largely due to inadequate tube correction for diameter and frequency variations.

534.75:534.79　　　　2692
The Anomalous Behaviour of the Threshold of Hearing in Relation to the Equal-Loudness Contours—R. Guelke and H. Helm. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 317–322; May, 1952.) Tests on 48 subjects show that the intensity jump from threshold to 10 phons is greater at 100 cps than at 1,000 cps. This is contrary to experience at higher intensities. A possible explanation is suggested.

534.78:534.791　　　　2693
On the Measurement of the Loudness of Speech—I. Pollack. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 323–324; May, 1952.) A loudness scale for speech is presented which is based on available experimental data. Apparatus and methods used were the same as for measurements on white noise (1501 of July).

534.84:[621.395.61+621.395.623.7　　2694
Quality Assessment of Rooms for Electroacoustic Transmissions—H. Etzold. (*Funk u. Ton*, vol. 6, pp. 191–197; April, 1952.) Discussion of the relation between the acoustical properties of rooms and the quality of the reproduction in a microphone-amplifier-loudspeaker system, the microphone and loudspeaker being in different rooms, account being taken of physiological and psychological factors.

534.845　　　　2695
Note on Beranek's Theory of the Acoustic Impedance of Porous Materials—J. W. McGrath. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 305–309; May, 1952.) The theory previously given by Beranek (1410 of 1942) is modified so as to avoid two approximations. The modified theory appears to be quite adequate for very porous materials with low flow resistance.

534.845　　　　2696
An Experimental Investigation of Resonant Sound Absorbers—K. A. Velizhanina. (*Zh. Tekh. Fiz.*, vol. 21, pp. 1087–1099; September, 1951.) A detailed report is presented on experiments with sound-absorbing systems consisting of perforated sheets mounted at a distance from a wall. The theory of such systems is discussed and design formulas are derived.

534.845　　　　2697
The Use of Perforated Facings in Designing Low-Frequency Resonant Absorbers—D. B. Callaway and L. G. Ramer. (*Jour. Acous. Soc. Amer.*, vol. 24, pp. 309–312; May, 1952.)

534.851/.852:534.321　　　　2698
Measurement of Pitch Fluctuations—H. Vollmer. (*Funk u. Ton*, vol. 6, pp. 169–175; April, 1952.) Various possible methods of measurement of fluctuations in sound-reproduction equipment are considered briefly. The method adopted was a phase-measurement bridge, the indicating instrument in which gives the frequency deviation direct, with maximum scale readings on the two ranges of ±0.5 per cent and ±1.5 per cent. A standard test frequency of 5 kc is used.

621.395.61/.62+534.85/.86　　2699
I.R.E. Show Review—1952—H. K. Richard-

son. (*Audio Eng.*, vol. 36, pp. 26–28, 53; April, 1952.) Brief notices of some of the af equipment exhibited.

621.395.62　　　　2700
Equipment for Sound Reproduction—*Wireless World*, vol. 48, pp. 255–258; July, 1952.) Illustrated descriptions of disk and magnetic recorders, amplifiers, loudspeakers, pickups and microphones shown at recent exhibitions.

621.395.623.73　　　　2701
Electrical Loudspeakers—L. Iglück. (*Nachrichtentechnik*, vol. 2, pp. 11–15; January, 1952.) A short review of the basic principles of various types of loudspeaker, including high-power types, and details of the construction of a giant 1-kw loudspeaker and of the means adopted for increasing its efficiency.

621.395.625.6:534.862.3　　　2702
New Sound-Recording System—J. P. Shields. (*Radio-Electronics*, vol. 23, pp. 26–28; April, 1952.) Description of a system using spiral scanning of an intensity-modulated electron beam to record 30-minute programs on disks of ordinary photographic film. For reproduction, the developed film is interposed between the screen of the scanning cr tube and a lens focusing the light from the screen on a photocell, whose output voltage is fed to a loudspeaker via an amplifier.

621.395.813:621.317　　　2703
Objective Measurements of Intelligibility on Subscribers' Telephones—G. Fontanellaz. (*Tech. Mitt. schweiz. Telegr.-Teleph Verw.*, vol. 29, pp. 445–466; December 1, 1951. In German and French.) Report of tests made by the Swiss Post Office in conjunction with the C.C.I.F., complementary to the subjective measurements made previously by the latter body (1191 of June). The methods of measurement are described; results are in satisfactory agreement with those obtained from the subjective tests.

ANTENNAS AND TRANSMISSION LINES

621.392.09:621.392.26　　　2704
On the Excitation of Surface Waves—G. Goubau. (PROC. I.R.E., vol. 40, pp. 865–868; July, 1952.) A theoretical proof is given of the existence of surface waves. The complete solution of the field equations for cylindrical surface waveguides includes terms corresponding to the surface wave and to supplementary fields; these are separable by means of certain orthogonality relations. The amplitude of the surface wave excited by a dipole or other source can be determined without solving the entire excitation problem.

621.392.121　　　　2705
Etched Sheets serve as Microwave Components—R. M. Barrett. (*Electronics*, vol. 25, pp. 114–118; June, 1952.) The flat-strip transmission line has the advantages of a coaxial line and a form factor suited to the printed-circuit technique. The characteristic impedance of the line is a function of the width of the central strip. A plot of the field distribution in such lines shows that the field is largely concentrated near the inner strip, and no energy is propagated laterally. A practical method of feeding the line from a coaxial system is to insert a normal N-type connector between the edges of the dielectric. All conventional microwave components such as hybrid junctions, directional couplers, power-division networks and filters are readily manufactured by the technique. When losses due to a continuous dielectric sheet are too high, or when high power is used, a compensated stub-supported transmission line is useful.

621.392.22　　　　2706
Determination of Conditions for Integrating the Equations of Propagation of Electricity along a Heterogeneous Line—M. Parodi. (*Compt. Rend. Acad. Sci.*, (Paris), vol. 234,

pp. 1674–1676; April 21, 1952.) Solutions are obtained by an iterative process for the differential equations of steady-state propagation.

621.392.26　　　　2707
The Impedance of Unsymmetrical Strips in Rectangular Waveguides—L. Lewin. (*Proc. IEE*, Part IV, vol. 99, pp. 168–176; July, 1952.) Full paper. See 2434 of October.

621.392.26　　　　2708
Single- and Multi-Iris Resonant Structures—I. Reingold, J. L. Carter and K. Garoff. (PROC. I.R.E., vol. 40, pp. 861–865; July, 1952.) An experimental investigation was made of the effect of varying the position of a rectangular window in a thin transverse diaphragm in a waveguide. The Q of a window of given dimensions is increased as the distance of the window from the broad wall of the waveguide is decreased. Low-Q multiwindow structures can be obtained in which the Q of the individual windows is comparatively high.

621.392.26:517.932　　　2709
Generalized Telegraphist's Equations for Waveguides—S. A. Schelkunoff. (*Bell Sys. Tech. Jour.*, vol. 31, pp. 784–801; July, 1952.) "Maxwell's partial differential equations and the boundary conditions for waveguides filled with a heterogeneous and nonisotropic medium are converted into an infinite system of ordinary differential equations. This system represents a generalization of 'telegraphist's equations' for a single-mode transmission to the case of multiple-mode transmission. A similar set of equations is obtained for spherical waves. Although such generalized telegraphist's equations are very complicated, it is very likely that useful results can be obtained by an appropriate modal analysis."

621.392.26:538.614　　　2710
Faraday Rotation of Guided Waves—H. Suhl and L. R. Walker. (*Phys. Rev.*, vol. 86, pp. 122–123; April 1, 1952.) An extension of the theory of Faraday rotation to include the case of transmission in a bounded waveguide system, as opposed to that of propagation in an unbounded medium, for which the theory already exists. A modified Verdet's constant, appropriate to waveguide transmission, is derived. See also 2911 of 1951 (Goldstein et al.), 1233 of June (Hogan) and 1587 of July (Wicher).

621.392.26:621.314.25.088　　2711
Errors in a Microwave Rotary Phase Shifter—A. J. Simmons. PROC. I.R.E., vol. 40, p. 869; July, 1952.) A discussion of the effect of nonlinearity of the phase-shift/rotation characteristic on the performance of the adjustable waveguide phase changer described by Fox (1255 of 1948.)

621.392.26:621.392.52　　　2712
The Calculation of the Dispersion in a Waveguide Filter with Slot Couplings—G. N. Rapaport. (*Zh. Tekh. Fiz.*, vol. 21, pp. 1076–1086; September, 1951.) Discussion of propagation of waves in a cylindrical waveguide divided into sections by partitions with annular slots. The field inside each section is expanded in series with respect to the characteristic functions of the closed cylindrical resonator. Using the continuity of the magnetic field and the flux of the Umov-Poynting vector through the slot, equations are derived determining dispersion and the electric field of the slot. The spectrum of the frequencies passed by the filter is considered and also the relation between the bandwidth and the width of the slots.

621.392.43:621.385.029.6:621.396.65　2713
Impedance Matching of a Travelling-Wave Valve [Type M8] to a Waveguide.—Chavance and L. Moutte. (See 2940.)

621.392.43:621.385.029.6:621.396.65　2714
Impedance Matching of Travelling-Wave Valves [Types M8 and M11] to Coaxial Lines—Clostre and Wallauschek. (See 2939.)

621.396.67 **2715**

Cylindrical Aerials—B. Storm. (*Wireless Eng.*, vol. 29, pp. 174 176; July, 1952.) A new method of solving Hallen's integral equation for the current in an antenna is presented which is simpler and more direct than the Hallén-King iteration method. The current is determined as the sum of a dominant sinusoidal current and a series of higher harmonics corresponding to a trigonometric series, which are integrated term by term, giving a series of sums of Si and Ci functions. Theoretically it is possible to calculate the antenna current and impedance with great accuracy. The amount of labor increases with the number of points considered along the antenna, but no new mathematical difficulties are encountered. Five points give sufficient accuracy for most practical cases. Calculated results for a λ/2 dipole and a λ dipole are compared with 1st-order and 2nd-order calculations by King.

621.396.67 **2716**

The Absorption Gain and Back-Scattering Cross-Section of the Cylindrical Antenna— S. H. Dike and D. D. King. (*Proc. I.R.E.*, vol 40, pp. 853–860; July, 1952.) The method of Hallén (2763 of 1939) as modified by Middleton and King (1771 of 1946) is used to derive first-order formulas for the gain and back-scattering cross-section of a receiving dipole arranged broadside on. Calculated values are compared with values obtained from measurements on unloaded and on matched-loaded dipoles; explanations are advanced for the discrepancies found.

621.396.67 **2717**

A Slot Aerial with Directing Elements— M. L. Levin. (*Zh. Tekh. Fiz.*, vol. 21, pp. 795–801; July, 1951.) In order to increase the directivity of a slot antenna a passive dipole is placed in front of the slot. The theory of such a system is discussed.

621.396.67 **2718**

Derivation of the Fundamental Equation of the Theory of Slot Aerials—M. L. Levin. (*Zh. Tekh. Fiz.*, vol. 21, pp. 787–794; July, 1951.) The voltage distribution along a narrow slot cut in a thin conducting surface is derived from the theory of metal antennas.

621.396.67 : 621.392 **2719**

Multiple Reflections in Long Feeders— J. Lewin. (*Wireless Eng.*, vol. 29, pp. 189–193; July, 1952.) The signal distortion produced by reflections at many points along a mismatched waveguide feeder is analyzed in terms of harmonic content for a single-tone signal, and in terms of interchannel interference in the case of a multichannel source. Such distortion is negligible for short feeders, but dominant for long ones. Measurements on a 120-ft feeder, made up of 10 foot sections, confirm the incoherent nature of the reflected waves. Considerable improvement in transmission should result from reduction of reflections at waveguide joints. The use of long sections on very long runs is desirable.

621.396.67 : 621.392.26 **2720**

On the Theory of Slot Aerials in a Circular Waveguide—M. L. Levin. (*Zh. Tekh. Fiz.*, vol. 21, pp. 772–786; July, 1951.) Results obtained by Pistol'kors (1266 and 1267 of 1948) for the radiation from transverse slots are shown to be erroneous; correct formulas are derived.

621.396.67 : 621.392.26 : 538.221 **2721**

Microwave-Antenna Ferrite Applications— N. G. Sakiotis, A. J. Simmons and H. N. Chait. (*Electronics*, vol 25, pp. 156–166; June, 1952.) For general radar scanning purposes an axial dc magnetic field is applied to a ferrite-cored transmission line, the variation of permeability effecting a phase shift of the circularly polarized input. The nonreciprocity of the simple system for transmission and reception is overcome with a combination of ferrite cylinders; by eliminating a 45° phase-shift element at one end the system can be used as a TR switch. Phase shift per unit length of a ferrite rod is a function of the length, increasing as the length increases, due to the change in the demagnetization factor.

621.396.671 **2722**

Input Impedance of Horizontal Dipole Aerials at Low Heights above the Ground— R. F. Proctor. (*Proc. I.R.E. (Australia)*, vol. 13, pp. 58–61; February, 1952.) Reprint. See 2134 of 1950.

621.396.671 : 537.311.5 : 538.569 **2723**

On the Current Induced in a Conducting Ribbon by the Incidence of a Plane Electromagnetic Wave—E. B. Moullin and F. M. Phillips. (*Proc. IEE* (London), Part IV, vol. 99, pp. 137–150; July, 1952.) Full paper. See 2441 of October.

621.396.676 **2724**

The Slot Aerial and its Application to Aircraft—R. H. J. Cary. (*Proc. IEE* (London), Part III, vol. 99, pp. 187–196; July, 1952. Discussion, pp. 210–213.) Measurements on full-scale and model aircraft show that the types of polar diagram frequently required for aircraft can be obtained with slot antennas, for which there is some agreement between measured polar diagrams and those calculated from ray and diffraction theories. Exact diagrams must be determined experimentally, since aircraft do not have simple shapes which permit accurate calculation. The installation of slots for metre or shorter wavelengths presents little difficulty, but their application for longer waves depends on what can be achieved in matching devices and on the limit of possible modification of aircraft structure.

621.396.676 **2725**

A Survey of External and Suppressed Aircraft Aerials for Use in the High-Frequency Band—R. H. J. Cary. (*Proc. IFE*, Part III, vol. 99, pp. 197–210; July, 1952.) Discussion, pp. 210–213.) The characteristics of external wire antennas are compared with those of "suppressed" antennas within the aircraft skin. Some of the latter type have performances which compare favorably with those of external wire types and in some respects have better electrical characteristics. In particular, folded or near-end-fed antennas, or those which involve an insulated stabilizer unit, have a large bandwidth and are reasonably efficient. The impedance characteristics of some of the suppressed antennas render them suitable for frequencies lower than the hf band. Methods of excitation for suppressed antennas are discussed. Measurements of the characteristics of aircraft antennas are most economically carried out on scale models.

621.396.677 **2726**

Radiation Field of a Square Helical Beam Antenna—H. L. Knudsen. (*Jour. Appl. Phys.*, vol. 23, pp. 483–491; April, 1952.) Rigorous formulas are derived for the field of an antenna with a uniformly progressing current wave of constant amplitude; these are compared with the Kraus approximation (643 and 1860 of 1949) for a circular-helix antenna. The two methods of calculation give results in fair agreement, the rigorous formulas being simpler and of direct application to the square helices often used in the metre-wave band. See also 1844 of 1951.

621396.677 **2727**

Waves in a Pyramidal Horn—E. G. Zelkin. (*Zh. Tekh. Fiz.*, vol. 21, pp. 1228–1239; October, 1951.) A theoretical investigation in which simplification is introduced by considering an approximation to the pyramidal horn consisting of a biconical sector with two additional walls (Fig. 1). The walls of such a horn coincide with the co-ordinate surfaces of a spherical system. The propagation of waves and the configuration of the field inside this horn is discussed. The polar diagram is also considered.

621.396.677.3.012† **2728**

Design Data for Horizontal Rhombic Antennas—E. A. Laport. (*RCA Rev.*, vol. 13, pp. 71–94; March, 1952.) A method is described which greatly simplifies design calculations. Using tables which give the spherical co-ordinates for all radiation-pattern lobes up to those of the sixth order, the performance of a horizontal rhombic antenna over a range of frequencies can easily be investigated. The complete radiation pattern is then determined from stereographic charts of the type suggested by Foster (119 of 1938). A set of 13 transparent charts for leg lengths of 2–7λ covers the usual practical range.

621.396.679 **2729**

Design of Optimum Buried-Conductor R.F. Ground System—F. R. Abbott. (*Proc.* I.R.E. vol. 40, pp. 846–852; July, 1952.) Design formulas and charts are presented for a radial-conductor earthing system particularly suitable for operation at frequencies below 1 mc and leading to antenna radiation efficiencies of 50 per cent or more.

CIRCUITS AND CIRCUIT ELEMENTS

621.3.015.7 : 621.387.4 **2730**

Millimicrosecond-Pulse Techniques—N. F. Moody, G. J. R. Maclusky and M. O. Deighton. (*Electronic Eng.*, vol. 24, pp. 214–219, 287–294 and 330–333; May–July, 1952.) A report of experimental work in the development of pulse circuits, particularly for laboratory measurements. Particular attention is given to circuits using secondary-emission pentodes. Descriptions are included of oscillographic equipment for viewing rapidly repeated waveforms, and of devices for expanding pulses on an analogue principle. Application of the techniques to the study of short nuclear half-lives is discussed.

621.314.3† **2731**

Compensating for Quiescent Current in Multistage Magnetic Amplifiers—A. S. Fitzgerald. (*Elec. Eng.*, vol. 71, pp. 206–211; March, 1952.) Essential text of 1952 A.I.E.E. Winter General Meeting paper. Three different compensation circuits are discussed and a fourth circuit, designed to combine the best features of the others, is described. This includes, in addition to the single saturable reactor with the usual ac and dc windings, a transformer and two resistors.

621.314.3† **2732**

Improvements extend Magnetic-Amplifier Applications—F. Benjamin. (*Electronics*, vol. 25, pp. 119–123; June, 1952.) The characteristics of modern core materials and rectifiers are noted and their bearing on magnetic-amplifier performance is discussed.

621.316.726 : 621.396.615.141.2 **2733**

Frequency Control of Modulated Magnetrons by Resonant Injection System—L. L. Koros. (*RCA Rev.*, vol. 13, pp. 47–57; March, 1952; *Proc. NEC* (Chicago), vol. 7, pp. 39–45; 1951.) Two systems are described by which the output frequency f_m of a modulated magnetron can be crystal-controlled. "Pushing" and "moding" are thereby suppressed and the carrier can be modulated 100 per cent with high fidelity. In one system the crystal drive excites the input cavity of a grounded-grid frequency doubler at a frequency $f_m/2$; the anode cavity, tuned to f_m, is coupled to the magnetron output which is adjusted for correct loading. In the second system the injection frequency for the magnetron is a subharmonic of f_m; in a particular case the anode cavity of the multiplier was to $f_m/2$, the crystal-drive frequency being $f_m/6$.

621.316.726.078.3:621.396.615   2734
The Attainment of Very High Frequency Stability—W. Herzog. (*Arch. elekt. Übertragung*, vol. 6, pp. 159–162; April, 1952.) Methods are indicated for improving the Q of the feedback circuits of various well-known oscillators by the inclusion of a compensating resistance.

621.316.729:621.396.615.072.9   2735
Representation of Oscillator Synchronization by Nonlinear Equation—A. Blaquière. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1741–1743; April 28, 1952.) Theory previously given (335 of March) is extended to the study of forced oscillations, and the limits of synchronization are determined by consideration of a diagram in the complex plane. A method for investigating synchronization stability is outlined. Application of the graphical treatment to the van der Pol equation has given all the results established by other methods by van der Pol.

621.316.84:539.23   2736
Metal-Film Resistors—R. J. Heritage. (*Electronic Eng.*, vol. 24, pp. 324–327; July, 1952.) The production of such resistors by chemical reduction, evaporation, sputtering, or firing-on methods is discussed briefly. The firing-on process appears to be the most suitable for mass-production methods; the technique is described and examples are shown on various practical types.

621.316.86:546.281.26   2737
Industrial Applications of Semiconductors: Part 2—Silicon Carbide Resistors—R. W. Sillars. (*Research* (London), vol. 5, pp. 169–175; April, 1952.) The manufacture and properties of SiC resistors of different types are described, and their applications for overvoltage protection, voltage discrimination in communication circuits, voltage regulation etc., are considered. Conduction in n- and p-type SiC, and barrier-layer effects, are discussed.

621.318.57   2738
Switching Action in Circuits including Valves—W. Taeger. (*Funk u. Ton*, vol. 6, pp. 198–201; April, 1952.) Analysis of effects in circuits such as that of a pentode with a choke and resistor in the anode lead. The time required to reach the tube saturation current is the shorter the lower the value of the saturation current and the less the resistance in the anode circuit.

621.318.572:681.142   2739
Universal High-Speed Digital Computers: Serial Computing Circuits—Williams, Robinson and Kilburn. (See 2839.)

621.319.4   2740
R.F. Characteristics of Capacitors—R. Davidson. (*Wireless World*, vol. 58, pp. 301–304; August, 1952.) Examination of the inductance and other properties of different types of capacitor used for decoupling and interference suppression.

621.319.43   2741
Profile Calculation for Rotary [plate] Capacitors—O. Schmid. (*Frequens.*, vol. 6, pp. 105–107; April, 1952.) A general formula relating the plate shape and insertion angle is given and applied to the types with linear variation of (a) capacitance, (b) wavelength, (c) frequency, and with logarithmic variation of capacitance with insertion angle.

621.392.5:512.31   2742
Network Synthesis using Tchebycheff Polynomial Series—S. Darlington. (*Bell. Sys. Tech. Jour.*, vol. 31, pp. 613–665; July, 1952.) "A general method is developed for finding functions of frequency which approximate assigned gain or phase characteristics, within the special class of functions which can be realized exactly as the gain or phase of finite networks of linear lumped elements. The method is based upon manipulations of two Tchebycheff polynomial series, one of which represents the assigned characteristic, and the other the approximating network function. The wide range of applicability is illustrated with a number of examples."

621.392.5:537.228.1   2743
Piezoelectric Transducers for Ultrasonic Delay Lines—H. N. Beveridge and W. W. Keith. (PROC. I.R.E., vol. 40, pp. 828–835; July, 1952.) "The over-all frequency response and loss of the transducers of a delay line are derived from Roth's equivalent circuit [3021 of 1949]. The transient response of a delay line is computed from an equivalent transmission-line circuit. The bandwidths and transient responses measured on practical delay lines are given for comparison with theoretical values."

621.392.5:621.385.3:512.83   2744
Matrix Theory applied to Thermionic Valve Circuits—S. R. Deards. (*Electronic Eng.*, vol. 24, pp. 264–267; June, 1952.) Extension of linear-quadripole theory to include tubes as network elements. The method is illustrated by analysis of a cathode-follower circuit.

621.392.5.029.3:621.3.012.3   2745
The Prediction of Audio-Frequency Response: No. 3—Single Complex Impedance in Resistive Network—N. H. Crowhurst. (*Electronic Eng.*, vol. 24, pp. 241–243; May, 1952.) Formulas and charts are given for computing the response of networks with series or shunt complex impedances. No. 2: 1540 of July.

621.392.5.029.3:621.3.012.3   2746
The Prediction of Audio-Frequency Response: No. 4—Step Circuits—N. H. Crowhurst. (*Electronic Eng.*, vol. 24, pp. 337–339; July, 1952.) Circuits using a single reactance to produce a step in the amplitude response are discussed and basic and practical types are shown. Charts are provided from which the amplitude and phase response can be found. Numerical examples illustrate their use. No. 3: 2745 above.

621.392.52   2747
Elementary Introduction to Filter Theory—C. Wisspeintner. (*Funk u. Ton*, vol. 6, pp. 29–40, 89–97, 141–153 and 202–213; January–April, 1952.) A practical treatment based on vector diagrams. The various types of filter are considered in turn, starting with low-pass filters, and phase relations and the effect of circuit losses are discussed.

621.392.52   2748
Construction of Band-Pass Filters from Basic Elements—W. Nonnenmacher. (*Frequenz*, vol. 6, pp. 107–113; April, 1952.) Detailed treatment of the transmission properties of various basic circuits and their use in the construction of band-pass filters with symmetrical and with asymmetrical transmission curves. Rumpelt's template method of filter design (729 of 1943) is explained.

621.392.52   2749
Some Considerations in the Design of a Miniature Filter-Transformer—T. T. Brown. (*Marconi Rev.*, vol. 15, pp. 90–94; 2nd Quarter 1952.) Means of attenuating upper and lower frequencies in audio circuits are discussed, with reference to miniature airborne equipment. The design of a filter unit is described which includes the input transformer as an integral part and has a voltage step-up of 12:1 within 1db from 300 cps to 5 kc, and an attenuation slope of over 24 dg per octave outside these limits. It is mounted in a can 2×1.6.×1.2 inches.

621.392.52:517   2750
Some Problems in the Analysis of Curves by Calculus Methods—M. Lévy. (*Canad. Jour. Phys.*, vol. 30, pp. 147–158; March, 1952.) A discussion of transform methods in relation to filter theory. The process considered consists of performing combinations of arithmetical operations on the successive ordinates of the curve representing a signal function; it is inherently frequency selective. One problem is to find combinations of only a few operations yielding pass-band selectivity, the second problem is, whether the results of the analysis are unique. Consideration is given first to continuous transforms, yielding integrals, and then to discontinuous transforms, yielding sums and differences.

621.392.52:518.4   2751
New Graphical Methods for Analysis and Design—G. R. Schneider. (*Wireless Eng.*, vol. 29, pp. 194–195; July, 1952.) Comment on paper noted in 1546 of July (Saraga and Fosgate), calling attention to similar methods of determining filter insertion loss described in 1939 by Piloty (46 and 959 of 1940).

621.392.52:534.11–18   2752
A Band-Pass Mechanical Filter for 100 kc/s—L. L. Burns, Jr. (*RCA Rev.*, vol. 13, pp. 34–46; March, 1952.) The filter is a development of the neck-type design previously described [59 of 1950 (Roberts and Burns)]. It comprises eight λ/2 rods of Ni-Span C coupled by thin steel λ/4 necks silver-soldered to the rods. The terminations are lossy lines consisting of 5-foot lengths of rubber-coated Cu wire coiled on ⅞-inch formers. Two separate compartments are provided in the chassis to isolate the magnetostriction input and output circuits. The filter response compares favorably with that of a 3-section quartz-crystal filter. The insertion loss of the filter alone is about 8 db. Design procedure is described in an appendix.

621.392.52:621.396.611.3:621.396.621.54   2753
The Calculation of Coupled H.F. Band-Pass Filters comprising Any Number n of Circuits with Equal Q Factor—W. Pfost. (*Arch. elekt. Übertragung*, vol. 6, pp. 135–142; April, 1952.) A previous paper by Behling (380 of 1947) gives the calculation for a four-circuit filter. In the method now described a multi-circuit filter is reduced by successive steps to more elementary forms whose input admittances are simply evaluated. Formulas are established also for the voltage transfer, efficiency and selectivity.

621.396.6   2754
New Techniques for Electronic Miniaturization—R. L. Henry, R. K. F. Scal and G. Shapiro. (*Proc. I.R.E.* (*Australia*), vol. 13, pp. 75–81; March, 1952.) Reprint. See 307 of 1951.

621.396.6:621.317.755   2755
Slow-Speed Circular Timebase—A. M. Hardie and P. A. V. Thomas. (*Wireless Eng.*, vol. 29, pp. 177–183; July, 1952.) Two methods of cro display in polar form, in correct phase relation with reference to a rotating member, are described. In the first method, currents proportional respectively to sin θ and cos θ are derived from an af oscillator in conjunction with an Admiralty magslip transmitter and applied to the deflection coils of cro with a central deflecting electrode sealed through the tube face, to produce a constant rotating magnetic field. A method of correcting the effect of spurious phase shifts in this system is described. The second method makes use of a specially designed sine-cosine potentiometer, full details of which are given, in conjunction with a normal es cr tube.

621.396.611.1:517.51   2756
Resonant Circuit with Periodically Varying Parameters—D. G. Tucker: P. Bura and D. M. Tombs. (*Wireless Eng.*, vol. 29, pp. 222–224; August, 1952.) Discussion of paper abstracted in 2471 of October.

21.396.611.21   **2757**

Note on Safe Resonator Current of Piezo-lectric Elements—I J. Post. (PROC. I.R.E., ol. 40, p. 835; July, 1952.) Values of the opti-lum current for commonly used quartz cuts re tabulated.

21.396.615.17   **2758**

Some New Multivibrators—Chang Sing n 1 Chu Yao-I. (*Electronic Eng.*, vol. 24, pp. 70 271; June, 1952.) Circuits and waveforms re shown for modified multivibrators in which racitors are inserted between cathodes and rth, the dc paths being completed by resist-connecting cathode to grid.

21.396.615.17:621.317.755   **2759**

Feedback in Time-Base Circuits—A. E. rguson. (*Electronic Eng.*, vol. 24, pp. 280–1; June, 1952.) A comparison of the Miller 1 constant-current-pentode timebase cir-its.

1.396.645   **2760**

Distributed Amplification—R. W. A. Scarr-Cormack. (*Electronic Eng.*, vol. 24, p. 295; 1e, 1952.) Comment on 2167 of September l author's reply.

1.396.645   **2761**

Heater-Voltage Compensation for Alternat-g-Current Amplifiers—N. W. Broten. (PROC. R E., vol. 40, pp. 843–845; July, 1952.) ations of gain due to heater-voltage varia-is are reduced by means of a dc feedback vork. The design procedure is indicated, l a comparison is made of the results of surements on (*a*) a compensated and (*b*) uncompensated wide-band IF amplifier for ration at 30 mc.

1 396.645.015.3   **2762**

Calculating Transient Response—T. Rod-n. (*Wireless World*, vol. 58, pp. 292–295; gust, 1952.) (Description of a simplified hod of compounding frequency and phase aracteristics and applying a superposition nciple based on trapezoidal response curves.

21.396.645.211   **2763**

Theory of the Resistance-Coupled Ampli-r—R. Rücklin. (*Arch. elekt. Übertragung*, l. 6, pp. 163–170 and 198–205; April and lay, 1952.) By plotting against normalized quency and introducing the concept of andwidth factor, amplification and phase ngle of all possible resistance-coupled ampli-rs are represented by a single family of urves. The noise voltage at the output is valuated as an integral which has a closed olution. On replacing *t* by the normalized time *t* in the equations representing the output oltage, functions of general applicability are btained; these are used to find the form of the utput voltage for any input, in particular for riangular pulses. The optimum frequency re-ponse is determined for separating triangular ulses of given duration from background noise. Build-up transients are treated by splitting he step-function response into a step-voltage mplification and a decay function, the latter ffording a criterion for the amplifier distor-ion. The decay function is calculated for an mplifier with feedback via the supply battery.

21.396.645.371   **2764**

Negative-Feedback Amplifiers. Overload-ng under Pulse Conditions—J. E. Flood. *Wireless Eng.*, vol. 29, pp. 203–212; August, 952.) "The overloading of simple resistance-oupled negative-feedback amplifiers is investi-ated for an input signal which rises from zero o its final value at a uniform rate. The ampli-iers are assumed to be linear unless the ap-lied signal exceeds the permitted value. For he single-stage amplifier, the permissible in-ut voltage decreases as the rise time of the ignal is reduced. Two-stage and three-stage ritically damped amplifiers can be made to andle quickly changing signals which are as

large as the maximum permissible slowly chang-ing signal, provided that the time-constant of the first stage of the amplifier is sufficiently large compared with that of the second stage. Overloading of the amplifier by step voltages and by sinusoidal signals is studied in appen-dices."

621.385:[621.396.621+621.396.645   **2765**

Electronic Valves: Book V—Application of the Electronic Valve in Radio Receivers and Amplifiers; Vol. 2—A.F. Amplification, the Output Stage, Power Supply. [Book Review]—B. G. Dammers, J. Haantjes, J. Otte and H. van Suchtelen. Publishers: Cleaver-Hume Press, London, 431 pp., 45s. (*Wireless Eng.*, vol. 29, pp. 197–198; July, 1952.) "The authors have chosen to make it almost a text-book of broadcast receiver practice and have succeeded admirably in doing so." See 991 and 2888 of 1950 for previous books in this series.

GENERAL PHYSICS

530.12:530.145.6   **2766**

The Relativistic Relations between Fre-quency, Wavelength, Phase Velocity and Group Velocity—J. L. Synge. (*Compt. Rend. Acad. Sci.* (Paris) vol. 234, pp. 1669–1670; April 21, 1952.) Formulas established by de Broglie (76 of 1948) are here derived from con-siderations of relativistic invariance.

530.12:538.3   **2767**

The Physics of the Electromagnetic Uni-verse—B. Jouvet. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234. pp. 1532–1534; April,7, 1952.) Discussion of the significance of changes of co-ordinates of the electromagnetic universe, leading to a fundamental principle of electro-magnetic relativity; the experimental conse-quences are indicated.

534.26+535.43   **2768**

Multiple Scattering of Radiation by an Arbitrary Planar Configuration of Parallel Cylinders and by Two Parallel Cylinders—Twersky. (See 2685).

535.13:538.3   **2769**

Is there an Aether?—L. Infeld: P. A. M. Dirac. (*Nature* (London), vol. 169, p. 702; April 26, 1952.) Comment on paper noted in 1573 of July and author's reply.

535.312   **2770**

On the Theory of the Displacement of Rays in Total Internal Reflection of a Spherical Wave—L. M. Brekhovskikh. (*Zh. Tekh. Fiz.*, vol. 21, pp. 874–880; August, 1951.) A for-mula (1) is quoted determining the displace-ment of a beam along the boundary between the two media where the beam undergoes total reflection. This formula also applied to the case of a spherical wave.

535.42   **2771**

On the Diffraction of Cylindrical Electro-magnetic Waves—N. V. Zernov. (*Zh. Tekh. Fiz.*, vol. 21, pp. 1066–1075; September, 1951.) Diffraction at a circular aperture in an ideally conducting plane is considered. The boundary problem is reduced to an infinite system of lin-ear algebraic equations which are solved by the reduction method. For an approximate calcu-lation of the field it is sufficient to use only a few of the equations. A relation is established between the proposed method and the varia-tion method of solving the same boundary problem. The problem can also be reduced to an integral equation of the second kind.

537.226.3   **2772**

The Dielectric Properties of Systems Con-taining Straight Polar Chains—R. A. Sack. (*Aust. Jour. Sci. Res., Ser. A*, vol. 5, pp. 135–145; March, 1952.) Mathematical theory is developed which provides an explanation of the high losses at low frequencies observed in solids containing hydroxyl groups, and also of

the low frequency absorption in ionic crystals containing lattice imperfections.

537.315:539.163.001.8   **2773**

Study of the Distribution of Surface Poten-tial by Means of Radioactive Deposits—T. Westermark and L. G. Erwall. (*Nature*, (London), vol. 169, pp. 703–704; April 26, 1952.)

537.523.4   **2774**

Observations on the Electrical Breakdown of Gases at 2800 Mc/s: Part 2—Relative Breakdown Stresses, Statistical Lags and Formative Lags—W. A. Prowse and W. Ja-sinski. (*Proc. IEE*, Part IV, vol. 99, pp. 194–203; July, 1952. Summary, *ibid.*, Part III, vol. 99, pp. 215–217; July, 1952.) The forma-tive time is found to be zero, within the limits of experimental error, for spark breakdown in air, N, O and H in all conditions examined, but is appreciable in Ne, Ar and He. Statistical time-lags are much greater in air and N than in O and H. The results are discussed with ref-erence to the mechanism of breakdown. Part 1: 650 of April.

537.713   **2775**

A Simple System of Dimensions in Elec-trical Engineering—M. Eskenazi. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1673–1674; April 21, 1952.) It is proposed to base di-mensional equations on the system R, E, L, T (resistance, emf, length, time); this has the advantage over Tarboureich's system (696 of 1947) that the fundamental magnitudes are all represented by commonly used material stand-ards.

538.3   **2776**

Electromagnetic Energy Density and Flux—C. O. Hines. (*Canad. Jour. Phys.*, vol. 30, pp. 123–129; March, 1952.) A comparison is made of the validity and usefulness of Poynting's theorem and of Macdonald's theorem; reasons are given for preferring the latter. Only non-magnetic media are considered, though this restriction may not be necessary.

538.56:537.525   **2777**

Plasma Oscillations and Striations—G. V. Gordeev. (*Compt. Rend. Acad. Sci.*, (U.R.S.S.), vol. 79, pp. 771–774; August 11, 1951. In Russian.) Previous investigators have estab-lished a connection between the striations and one of the plasma waves. The method is ex-tended, and striations are considered as a group of waves. Equations (9) and (15) are derived in which the group velocity and wavelength of the almost monochromatic group of waves are related to the parameters of the discharge.

538.56:537.525   **2778**

Oscillations in a Discharge as a Source of Travelling Striations—A. A. Zaitsev. (*Compt. Rend. Acad. Sci.* (U.R.S.S.), vol. 79, pp. 779–781; August 11, 1951. In Russian.) Oscillatory phenomena were studied in pure neon, in neon and argon with molecular impurities, and in air. A close connection between the traveling striations and the oscillations of potential in the discharge was observed.

621.317.35:518.4   **2779**

Graphical Method applicable to Harmonic Analysis and to Symbolic Calculus.—G. La-ville. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1728–1730; April 28, 1952.)

GEOPHYSICAL AND EXTRATERRESTRIAL PHENOMENA

523.72+523.82]:621.396.822   **2780**

Electromagnetic Radiation produced by Electron Collisions in a Very Strongly Ionized Medium—B. Kwal. (*Jour. Phys. Radium*, vol. 13, pp. 35–38; January, 1952.) The prob-lem of em emission on long wavelengths due to electron collisions, when the electrons have nonrelativistic velocities, cannot be satis-factorily treated from the quantum theory

point of view, and classical theory must be applied. In the case of ionic plasma, for radiation frequencies below a critical value 100 mc for interstellar ionized regions and 10 kc for the solar corona, the emissions should have a constant intensity, though propagation at these frequencies cannot take place unless the adjacent region is one of lower density. For frequencies sufficiently low with respect to the critical frequency, classical theory no longer holds.

523.72 : 621.396.822   **2781**
The Position and Movement on the Solar Disk of Sources of Radiation at a Frequency of 97 mc/s: Part 3—Outbursts—R. Payne-Scott and A. G. Little. (*Aust. Jour. Sci. Res., Ser. A*, vol. 5, pp. 32–46; March, 1952.) Observations indicate that initially the position of an outburst almost coincides with that of the accompanying flare, but the outburst moves rapidly toward, and sometimes off, the solar limb. Its displacement is consistent with the assumption that the outburst is initiated by an agency, such as corpuscular streams, originating at the flare and moving outward with a velocity between 500 and 3,000 km. The polarization of the outbursts is discussed. Part 2: 1902 of August.

523.746 : 523.72   **2782**
Reversal of Polarisation of Microwaves from Sunspots—U. C. Guha. (*Indian Jour. Phys.*, vol. 25, pp. 8–16; January, 1951.) The reflection coefficients of the ordinary and the extraordinary waves are calculated for a barrier with a parabolic distribution of ions. The results are applied to explain the observed reversal of the polarization of microwaves escaping from sunspots.

523.746 "1952.01/.03"   **2783**
Provisional Sunspot-Numbers for January to March, 1952—M. Waldmeier. (*Jour. Geophys. Res.*, vol 57, p. 310; June, 1952; *Z. Met.*, vol. 6, p. 158; May, 1952.)

523.78   **2784**
The Total Solar Eclipse of February 25, 1952—R. O. Redman. (*Nature* (London), vol. 169, pp. 686–688; April 26, 1952.) A preliminary report of the results of observations made in the Sudan by various groups of astronomers and technicians, nearly all from Europe and the United States, during the eclipse.

523.78   **2785**
Combined Observations of the Total Eclipse of the Sun at Khartoum (Sudan) and of the Partial Eclipse observed by Radio-Telescope at Meudon, 25th February 1952—M. Laffineur, R. Michard, J. C. Pecker, M. d'Azambuja, A. Dollfus and I. Atanasijević. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1528–1530; April 7, 1952.) A brief account of the radio and optical equipment used and the measurements made by the French expedition to Khartoum. In the middle of totality the intensity of the radiation received on 550 mc was 19.5 per cent and on 255 mc was 30.5 per cent of that from the unobscured sun. At Meudon, the intensity of radiation received at the maximum of the partial eclipse was 83 per cent of that from the unobscured sun.

523.78 : 523.72   **2786**
The Ellipsoidal Form of the Sun observed at Metre Wavelengths—É. J. Blum, J. F. Denisse and J. L. Steinberg. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1597–1599; April 16, 1952.) Curves showing the variation of intensity of radiation received from the sun on 169 mc during the eclipses of 1st September 1951 and 25th February 1952 are interpreted as indicating that the coronal radiation from the equatorial regions is greater than that from the polar regions, the radiation at this frequency corresponding to that from a greatly flattened ellipsoid with nearly uniform brightness. See also 1282 of June (Bosson et al.).

523.78 : 523.72   **2787**
Khartoum Expeditions for Total Solar Eclipse of February 25th, 1952—M. K. Aly. (*Observatory*, vol. 72, pp. 63–72; April, 1952.) Report of meetings held prior to the eclipse, with statements of the measurements planned by the various research groups. Skeleton descriptions are given of equipment for measurement of solar rf radiation at wavelengths 8.5 mm, 9.4 cm, 0.55 m and 1.17 m.

523.85 : 621.396.822   **2788**
Radio-Frequency Radiation from the Constellation of Cyngus—J. H. Piddington and H. C. Minnett. (*Aust. Jour. Sci. Res., Ser. A*, vol. 5, pp. 17–31; March, 1952.) Observations were made at frequencies of 1,210 and 3,000 mc, using apparatus and methods described previously (1906 of August). At the lower frequency two sources were observed, one being the known radio star Cyngus A and the other a diffuse source or "radio nebula," probably due to thermal emission from clouds of ionized interstellar gas. Neither of these sources was observed at the higher frequencies.

550.38 "1952.01/.03"   **2789**
Indices of Geomagnetic Activity of the Observatories Abinger, Eskdalemuir and Lerwick, January to March 1952—(*Jour. Atmos. Terr. Phys.*, vol. 2, no. 4, pp. 256–258; 1952.) Three-hour K-indices are tabulated.

550.38 "1952.01/.03"   **2790**
Cheltenham Three-Hour-Range Indices K for January to March, 1952—R. R. Bodle. (*Jour. Geophys. Res.*, vol. 57, p. 310; June, 1952.)

550.385 "1951.10/1952.03"   **2791**
Principal Magnetic Storms [Oct. 1951–March 1952]—(*Jour. Geophys. Res.*, vol. 57, pp. 311–313; June, 1942.

550.386 "1951.10/.12"   **2792**
International Data on Magnetic Disturbances, Fourth Quarter, 1951—J. Bartels and J. Veldkamp. (*Jour. Geophys. Res.*, vol. 57, pp. 305–309; June, 1952.)

551.311.234 : 621.317.335.3   **2793**
The Effect of Moisture on the Electrical Properties of Soil—A. Cownie and L. S. Palmer. (*Proc. Phys. Soc.* (London), vol. 65, pp. 295–301; April 1, 1952.) Measurements of the permittivity of samples of soil containing 4.1–47.7 per cent moisture were made at a frequency of 430 mc, using a coaxial transmission line terminated by the sample under test. The permittivity increases with moisture content, the observed values ranging from 4.0 to 31.4. Comparison is made with the results of other investigators.

551.510.41 : 546.214   **2794**
Direct Measurements of the Vertical Distribution of Atmospheric Ozone to 70 Kilometers Altitude—F. S. Johnson, J. D. Purcell, R. Tousey and K. Watanabe. (*Jour. Geophys. Res.*, vol. 57, pp. 157–176; June, 1952.) Measurements in rockets at White Sands, New Mexico, showed that the ozone concentration above 35 km height decreased approximately exponentially.

551.510.535   **2795**
The Dependence of Ionospheric Disturbances on Local Time—J. M. Ardillon. (*Compt. Rend. Acad. Sci.* (Paris) vol. 234, pp. 1568–1571; April 7, 1952.) Observations made at Poitiers and at Washington during 1949–1950 are reported. The onset of a decrease of F_2-layer critical frequency is never observed during the daytime. Comparison is made of the diurnal variation of this critical frequency for quiet and disturbed days by plotting the differences between mean values for the five quietest and five most disturbed days of the month. These curve exhibits minima at 0600 and 1800 hours, with a maximum between those times. These

diurnal variations are more marked at the equinoxes than at other seasons.

551.510.535   **2796**
On Negative Ions of Molecular Oxygen in the D Layer—D. R. Bates and H. S. W. Massey. (*Jour. Atmos. Terr. Phys.*, vol. 2, no. 4, pp. 253–254; 1952.) Note supplementary to paper abstracted in 983 of April. Original estimates of the daytime ratio of negative ions to electrons may be too high.

551.510.535   **2797**
Theory of Formation of an Ionospheric Layer below E Layer Based on Eclipse and Solar Flare Effects at 16 Kc/s—R. N. Bracewell. (*Jour. Atmos. Terr. Phys.*, vol. 2, no. 4, pp. 226–235; 1952.) An anomaly in the height of reflection of steeply incident 16-kc waves during the partial solar eclipse of 1949 and observed effects of solar flares are incompatible with the theory that reflection takes place from the bottom of a Chapman layer. An alternative mechanism is proposed which does not require that electron density be negligible compared to the density of ionizable particles.

551.510.535   **2798**
The Effect of Collisions between Electrons on the Absorption of Radio Waves in the F-Layer and in the Solar Corona—V. L. Ginzburg. (*Zh. Tekh. Fiz.*, vol. 21, pp. 943–947; August, 1951.) A mathematical discussion showing that collisions between electrons do not contribute appreciably to the absorption of radio waves.

551.510.535 : 523.745 : 525.624   **2799**
Effect of Vertical Transport of Ions Caused by Solar Tides in F_2 Region—D. C. Choudhury. (*Indian Jour. Phys.*, vol. 25, pp. 1–7; January, 1951.) According to Martyn's theory (1053 of 1949 and back references) there is a marked vertical drift of ions. If the drift velocity is represented by the relation $v = v_0 e^{-\gamma z} \cos(\omega t + \delta)$, the night-time Chapman distribution will be markedly affected even if electron reduction by recombination is negligible. Calculated results are in general agreement with observations if the value of γ is about unity.

551.510.535 : 523.78   **2800**
Ionospheric Effects of Solar Eclipse at Sunrise, September 1, 1951—H. W. Wells. (*Jour. Geophys. Res.*, vol. 57, pp. 291–304; June, 1952.) Results obtained during the annular eclipse, which started before sunrise, show no evidence of any considerable ion production at any of the three observation stations until more than a third of the sun's disk was exposed during the recovery phase. Ion density increased rapidly about 15 minutes after maximum phase, and the high rate of increase, about double the normal rate, continued for half an hour after the end of the eclipse.

551.510.535 : 525.624   **2801**
The Calculation of the Probable Error of Determinations of Lunar Daily Harmonic Component Variations in Geophysical Data: a Correction—S. Chapman. (*Aust. Jour. Sci. Res., Ser. A*, vol. 5, pp. 218–221; March, 1952.) Comment on 636 of 1950 (Tschu), indicating certain inaccuracies.

551.510.535 : 550.385   **2802**
The Morphology of Storms in the F_2 Layer of the Ionosphere: Part 1—Some Statistical Relationships—E. V. Appleton and W. R. Piggot. (*Jour. Atmos. Terr. Phys.*, vol. 2, pp. 236–252; 1952.) A statistical study is made of ionospheric storm phenomena at a number of stations. The phenomena show marked diurnal variations and for correlation purposes must be referred to local time. Sequences of storm events at auroral, temperate and equatorial latitudes are contrasted; in temperate latitudes at least, such sequences are subject to seasonal variation.

51.510.535:621.396.11   **2803**
Lunar Variations in F-Region Critical Frequency at Singapore—Osborne. (See 2869.)

51.510.535:621.396.11.029.45   **2804**
An Explanation of Radio Propagation at 16 c s in Terms of Two Layers below E Layer— N. Bracewell and W. C. Bain. (*Jour. Atmos. Terr. Phys.*, vol. 2, pp. 216–225; 1952.) An ionospheric model is proposed to explain observed features of propagation over distances of 90 km and 535 km. At a range of 90 km, a wave is received by reflection from the upper of the postulated layers. At a range of 535 km, waves reflected from both layers contribute to the total field, especially the wave reflected once from the lower layer and those reflected more than once from the upper layer. Anomalous effects at distances of about 300 km from the transmitter are explained by the proposed model.

51.510.535(71)   **2805**
Ionospheric Disturbances in Canada—J. Meek. (*Jour. Geophys. Res.*, vol. 57, pp. 181–190; June, 1952.) Analysis of variations of F-region ionization and of abnormally high absorption of radio waves in northern latitudes indicates that disturbances appear first in one part of the auroral zone and then move round the earth with the sun for several days. The effect of a disturbance is enhanced and extends further south near certain geographic latitudes. Diurnal and seasonal characteristics of the disturbances are described. It is suggested that the disturbances are connected with similar magnetic disturbances and are due to a narrow stream of solar particles moving into the earth's path.

51.594.13   **2806**
Electric Conductivity and Small-Ion Concentration of the Atmosphere at One Metre above Ground, and Conductivity at Ground Level—G. A. O'Donnell. (*Jour. Atmos. Terr. Phys.*, vol. 2, pp. 201–215; 1952.) Results of measurements at five locations within 120 miles of New York show that conductivity values at ground level and at 1m are not equal, nor is their ratio constant. This applies to both polar and total conductivity.

51.594.21   **2807**
Distribution of Electrical Conduction Currents in the Vicinity of Thunderstorms—R. E. Holzer and D. S. Saxon. (*Jour. Geophys. Res.*, vol. 57, pp. 207–216; June, 1952.)

51.594.22   **2808**
Thunderbolts: The Electric Phenomena of Thunderstorms—E. Gold. (*Nature* (London), vol. 169, pp. 561–563; April 5, 1952.) Report of discussion at the Royal Astronomical Society, January 1952.

51.594.6:621.317.35   **2809**
The Waveforms of Atmospherics—P. G. F. Caton and E. T. Pierce. (*Phil. Mag.*, vol. 43, pp. 393–409, April, 1952.) Convenient equipment and technique for recording waveforms of atmospherics are described; a time resolution of $3\mu s$ is achieved on photographs with an exposure time of 20 ms. The results of an extensive series of observations made at Cambridge, England, during the period 1947–1951 are presented and analyzed. Both day and night conditions, all seasons of the year, and a wide range of distance of source of atmospheric (100–4,000 km) were covered. A new and more complete classification of atmospheric waveforms is derived and the characteristics of each type are studied in relation to distance of origin. At night, the type of waveform observed depends on the geographical location of the source; for storms to the southwest, a transition in waveform type occurs at a distance of about 1,600 km, but no similar transition is observed for sources to the southeast. A critical discussion is presented of the applicability of the theory of multiple ionospheric reflections

to the various types of waveform, and of the accuracy with which the effective height of reflection and the distance of the source can be determined. Only a small proportion of atmospherics recorded, rarely more than 10 per cent, on any one night, give reliable values of reflection height and source distance. The reflection height is usually in the range 75–95 km. Photographs showing waveforms of the different types of atmospheric are reproduced.

LOCATION AND AIDS TO NAVIGATION

621.396.9   **2810**
Distance Measurement by Radar Technique—W. Messerschmidt. (*Arch. Tech. Messen*, no. 194, pp. 49–52; March, 1952.) Short descriptions of various methods, with discussion of the accuracy attainable.

621.396.9   **2811**
Circular Polarisation for C. W. Radar—J. F. Ramsay. (*Marconi Rev.*, vol. 15, pp. 71–89; 2nd Quarter 1952.) The simple properties of linearly, circularly and elliptically polarized waves are reviewed. A linearly polarized wave can be converted into a circularly polarized wave by resolving it into two equal orthogonal components and shifting the phase of one by $90°$. A phase shift of $180°$ simply rotates the plane of polarization. Practical devices for these purposes are called "circularizers" and "rotators" respectively. Several types are described and illustrated, and their application to microwave antennas discussed. Common T-R operation is achieved by incorporating a circularizer in the antenna and using orthogonal polarizations for transmitter and receiver. Mismatch in a system for circular polarization can result in elliptical polarization. Details of an experimental 10-cm cw radar antenna for common T-R operation are given.

621.396.93:621.396.671   **2812**
An Investigation of Polarization Errors in an H-Adcock Direction-Finder—F. Horner. (*Proc. IEE* (London), Part IV, vol. 99, pp. 229–240; July, 1952. Summary, *ibid.*, Part III, vol. 99, pp. 223–225; July, 1952.) The polarization-error characteristics of H-type Adcock direction finders for the vhf band and the upper part of the hf band are controlled to a large extent by resonance phenomena. Investigation of a vhf rotating system shows that the frequencies at which large errors are liable to occur, and the magnitudes of the maximum errors can be calculated from the physical dimensions of the antenna system. The lowest resonance frequency is that for which the product of antenna length and spacing is $\lambda^2/20$. A second resonance occurs when the spacing is $\lambda/2$, and, if spacing and antenna length are about equal, a third resonance occurs for a spacing of $0.8\ \lambda$. Three major causes of polarization errors are discussed and various methods of minimizing errors are considered.

621.396.933   **2813**
A Multichannel Distance Measuring Equipment for Aircraft—E. B. Mulholland. (*Proc. IRE*, (*Australia*), vol. 13, pp. 47–58; February, 1952.) Description of radar equipment operating in conjunction with responder beacons at known locations on the ground. The aircraft set transmits pairs of pulses with a selected time interval; each beacon contains a discriminator and responds only to pulses with a certain interval. A 12-channel system operating in the 200–230-mc band has been adopted as a standard for use on Australian airlines.

621.396.933   **2814**
The Civil Aeronautics Administration V.H. F. Omnirange—H. C. Hurley, S. R. Anderson and H. F. Keary. (*Proc. I.R.E.*, vol. 40, p. 860; July, 1952.) Correction to paper noted in 1000 of May.

621.396.933   **2815**
Anticollision Radar for Commercial Flights —M. Hobbs. (*Electronics*, vol. 25, pp. 110–113;

June, 1952.) Description of British equipment installed in high-speed jet aircraft for detecting thunderclouds and mountains. It operates on a 3-cm wavelength with a peak power of 10 kw. The antenna is an 18-inch paraboloid, stabilized gyroscopically, with beam angle $6°$, angle of scan on each side of the aircraft is $75°$. Tests of the apparatus are described and the principle of the safety-circle technique is illustrated. Details of the receiver circuits are noted.

621.396.9   **2816**
Leitfaden der Funkortung (Manual of Radio Direction Finding.) [Book Review]—W. Stanner and collaborators. Publishers: Elektron-Verlag, Garmisch-Partenkirchen, 1952, 164 pp., 8 DM. (*Fernmeldetech. Z.*, vol. 5, p. 197; April, 1952.) A general treatment including, in the longest and most interesting section of the book, an account of df developments in Germany and of British and American radar technique.

MATERIALS AND SUBSIDIARY TECHNIQUES

531.788.13   **2817**
An Automatically Controlled Knudsen-Type Vacuum Gauge—C. N. W. Litting and W. K. Taylor. (*Proc. IEE* (London), Part IV, vol. 99, pp. 241–249; July, 1952.) Description of a gauge in which the pressure-dependent radiometric force is balanced by an es force, the pressure being calculated from a measured voltage and the gauge dimensions. The various units of the instrument are described and also the method of calibration. Experimental results for N and Ar are in close agreement with theoretical values.

534.321.9:534.22:546.74   **2818**
Magnetically Induced Ultrasonic Velocity Changes in Polycrystalline Nickel—S. J. Johnson and T. F. Rogers. (*Jour. Appl. Phys.*, vol. 23, pp. 574–577; May, 1952.)

537.226:537.228.2   **2819**
Electrostriction Phenomena in Ferroelectric Ceramic Materials—G. A. Smolenski. (*Zh. Tekh. Fiz.*, vol. 21, pp. 1045–1049; September 1951.)

537.311.33   **2820**
Industrial Applications of Semiconductors: Part 1—Semiconductors—H. K. Henisch. (*Research* (London), vol. 5, pp. 101–107; March, 1952.) A simple account of the characteristic properties of these materials, with special reference to the physics of transistor action.

537.311.33:546.289   **2821**
Measurement of Diffusion in Semiconductors by a Capacitance Method—K. B. McAfee, W. Shockley and M. Sparks. (*Phys. Rev.*, vol. 86, pp. 137–138; April 1, 1952.)

537.311.33:546.289   **2822**
Diffusion of Donor and Acceptor Elements into Germanium—C. S. Fuller. (*Phys. Rev.*, vol. 86, pp. 136–137; April 1, 1952.)

538.221   **2823**
The Magnetic and Electrical Properties of Ferroxcube Materials—J. J. Went and E. W. Gorter. (*Philips Tech. Rev.*, vol. 13, pp. 181–193; January, 1952.) Aspects discussed include the chemical composition and crystallographic structure in relation to the saturation magnetization; the permeability and magnetic losses in weak and in strong fields; and the remarkable dielectric properties.

538.221   **2824**
Ferroxdure, a Class of New Permanent Magnet Materials—J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout. (*Philips Tech. Rev.*, vol. 13, pp. 194–208; January, 1952.) Ferroxdure is the name given to a class of magnetically hard oxidic ceramics, and especially to one consisting mainly of

$BaFe_{12}O_{19}$. This has hexagonal crystal structure with one axis of easy magnetization, resulting in a high value of coercive force together with rather low remanence and saturation magnetization; the relation between structure and magnetic properties is explained. The material is suitable for applications where opposition to demagnetization is required; several such applications are indicated. (See also *Phys. Rev.*, vol. 86, pp. 424–425; May 1, 1952.)

538.632:537.311.1   **2825**

Measurements of the Hall Effect in Zinc Oxide—K. Intemann and F. Stöckmann. (*Z. Phys.*, vol. 131, pp. 10–16; December, 1951.) Values of the order of 10 cm per V/cm are obtained for the mobility of electrons in evaporated films of ZnO; this is in satisfactory agreement with values found previously for compact ZnO.

539.234:537.29   **2826**

Effect of Electric Field on the Development of Thin Films—M. Perrot and J. P. David. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1753–1755; April 28, 1952.) Investigation of thin Ag and Al films evaporated in a high vacuum shows that different resistivities are obtained according to the value of the field applied during deposition.

539.234:546.26:537.311.32   **2827**

Conductivity and Flicker Effect of Very Thin Carbon Films—N. Nifontoff. (*Compt. Rend. Acad. Sci.* (Paris), vol. 234, pp. 1755–1757; April 28, 1952.) Continuation of work previously reported (171 of February and 341 of March), using more sensitive equipment.

539.234:546.72   **2828**

Measurements on the Current Sensitivity of the Electrical Resistance of Condensed Iron Films at Low Temperatures—A. van Itterbeek, R. Lambeir, B. Franken, G. J. van den Berg and D. A. Lockhorst. (*Physica*, vol. 18, pp. 137–144; March, 1952.) Measurements were made at temperatures down to those of liquid N_2, H_2 and He of the variation of resistance (*a*) with current, (*b*) in the presence of a magnetic field; the results are related to Kittel's theory (1129 of 1950). The resistance decrease with increasing current is rather less than for sputtered films.

546.811:621.3.011   **2829**

On the Electrical Properties of Thin Layers containing Grey Tin—N. A. Goryunove, I. D. Konozenko and A. P. Obukhov. (*Zh. Tekh. Fiz.*, vol. 21, pp. 814–817; July, 1951.) The specific resistance of thin layers of white tin containing small amounts of grey tin varies between 10^3 and 10^4 ohm cm. The thermal coefficient of resistance is negative and of the order of 3–5 per cent per 1°C.

546.817.221 + 546.817.231 + 546.817.241]:548.55   **2830**

Oxygen-Free Single Crystals of Lead Telluride, Selenide, and Sulfide—W. D. Lawson. (*Jour. Appl. Phys.*, vol. 23, pp. 495–496; April, 1952.) Anomalies in the type of conductivity of crystals of PbSe and PbTe have been found to result from the presence of oxygen. A method of preparation in an atmosphere of hydrogen is described which gives oxygen-free crystals. Single crystals of PbS have also been produced by this method.

549.514.51:537.226.2.096   **2831**

Temperature Variation of the Dielectric Constant of Quartz at H.F. between 20° and 700°C. $\alpha\beta$ Transformation. Thermal Twinning and Mechanical Detwinning—J. P. Pérez. (*Ann. Phys.* (Paris), vol. 7, pp. 238–282; March/April, 1952.) The quartz plates examined formed the dielectric of capacitors whose capacitance changes were determined from the readings of a calibrated variable capacitor. For plates cut parallel to the optic axis, the capacitance increases continuously with temperature

except at the $\alpha\beta$ transformation point, where there is usually a discontinuity. Plates cut perpendicular to the optic axis could not be studied beyond about 300°C, the conductivity becoming too great. The results indicate that the $\alpha\beta$ transformation is quasi-instantaneous and is characterized by slip at twinning surfaces. 68 references.

621.314.634   **2832**

Positive Current Creep in Selenium Rectifiers—R. Cooper and J. Harrington. (*Proc. Phys. Soc.* (London), vol. 65, pp. 303–304; April 1, 1952.) Experiments are described which indicate that although high power dissipation in the barrier layer may be a factor which favors positive current creep, as suggested by Henisch and Ewels (926 of 1951), it is not an essential condition and some other factor must be involved.

621.315.612.6   **2833**

The Electrical Conductivity and Breakdown Strength of Thin Glass Films—G. Glaser. (*Z. angew. Phys.*, vol. 4, pp. 12–16; January, 1952.) Experiments on glass films of thicknesses down to 0.07 μ indicate that there is only an indirect relation between electrical breakdown strength and conductivity. Breakdown strength increases as temperature decreases.

666.22:546.244-31   **2834**

Tellurite Glasses—J. E. Stanworth. (*Nature* (London), vol. 169, pp. 581–582; April 5, 1952.) A short account of the properties of various glasses with TeO_2 as the principal constituent, the second being PbO or BaO and the third Li_2O, Na_2O, B_2O, P_2O, GeO_2, V_2O_5, Cb_2O_5 or TiO_2. The expansion coefficients are unusually high for oxide glasses ($10-20 \times 10^{-6}$ per 1°C), refractive indices of the order 1.8–2.2, permittivities about 27 and loss factors about 0.003 from 50 cps to 1–2 mc. Some of the glasses have very low absorption at any wavelength up to 5.5μ.

MATHEMATICS

517.564.3:621.396.619.13   **2835**

Spectrum of a Frequency-Modulated Wave – Vaughan. (See 2892.)

517.93   **2836**

A Graphical Analysis for Non-Linear Systems—Pei-Su Hsia. (*Proc. IEE* (London), Part II, vol. 99, pp. 125–131; April, 1952. Discussion, pp. 132–134.) Full paper, summary noted in 1945 of August.

519.21   **2837**

Remarks on an Interpolation Theorem—A. Blanc-Lapierre. (*Compt. Rend. Sci.* (Paris), vol. 234, pp. 1733–1735; April 28, 1952.) Discussion of some properties of a class of stationary aleatory functions $X(t)$ such that the values of the function for any values of t are determined from knowledge of the values of $X(t_0 + nA)$ taken by the function $X(t)$ at uniformly distributed intervals on the t axis, n being integral. The theory is related to the theory of communication.

681.142:621.318.572   **2838**

Universal High-Speed Digital Computers: Serial Computing Circuits—F. C. Williams, A. A. Robinson and T. Kilburn. (*Proc. IEE* (London), Part II, vol. 99, pp. 107–120; April, 1952. Discussion, pp. 120–123.) Circuits designed for the Manchester University computer are described.

681.142:[621.392.26 + 621.396.611.4   **2839**

The Solution of Waveguide and Cavity-Resonator Problems with the Resistance-Network Analogue—G. Liebmann. (*Proc. IEE* (London), Part IV, vol. 99, pp. 260–272; July, 1952.) Description of a resistance-network analogue method which can be considered as an experimental counterpart of Southwell's relaxation technique. The network is the same

as that previously described for the solution of certain equations (1954 of 1950), with modifications necessary for carrying out the iteration process.

681.142:621.395.625.3   **2840**

Universal High-Speed Digital Computers: a Magnetic Store—F. C. Williams, T. Kilburn and G. E. Thomas. (*Proc. IEE*, Part II, vol. 99, pp. 94–106; April, 1952. Discussion, pp. 120–123.) An experimental intermediate store is described, comprising a drum coated with a layer of pure Ni as recording medium and rotated in synchronism with the fundamental frequency of the computer. Such a store, with a capacity of 3,000 numbers, each of 40 binary digits, is incorporated in the Manchester University computer.

MEASUREMENTS AND TEST GEAR

531.76   **2841**

A Direct-Reading Mains-Frequency-Cycle Counter—P. Huggins. (*Electronic Eng.*, vol. 24, pp. 276–277; June, 1952.) The counter includes three dekatrons [2066 of 1950 (Bacon and Pollard)] to indicate respectively units, tens and hundreds of cycles; a time interval can be read, in terms of mains-frequency cycles, to within 0.02 second.

531.76:621.389   **2842**

A Millisecond Timing Unit—J. E. Haworth. (*P.O. Elec. Engrs' Jour.*, vol. 45, part 1, pp. 25–29; April, 1952.) Description of electronic equipment for timing intervals up to 10 seconds in 0.5-ms units. The recording meters return to zero at the end of 10 seconds and then continue counting the number of cycles of a timing wave that occur in the interval to be measured. The equipment is particularly suitable for timing switching operations.

621.317.3:519.271:621.396.822   **2843**

Statistical Errors in Measurements on Random Time Functions—W. B. Davenport, Jr., R. A. Johnson and D. Middleton. (*Jour. Appl. Phys.*, vol. 23, pp. 377–388; April, 1952.) Analysis of the dependence of the statistical errors on the averaging interval, the method of averaging, and the statistics of the particular time function considered, with detailed discussion of the measurement of the power of a random-noise voltage of specified bandwidth.

621.317.3:537.228.1   **2844**

A New and Quick Method for Detection of Piezoelectricity and Measurement of the Piezoelectric Constants—K. G. Srivastava. (*Indian Jour. Phys.*, vol. 25, pp. 33–34; January, 1951.) Description of a simple method using a differential transformer.

621.317.321   **2845**

The Measurement of High Alternating Voltages by Means of Capacitive Voltage Dividers (C-Measurement): Part 2—P. Böning. (*Arch. Tech. Messen*, pp. 61–62; March, 1952.) An impedance diagram is constructed for the arrangement previously considered (440 of March), the resistive load being replaced by an instrument transformer. This circuit may be operated so that the voltage cross the transformer is in phase with the voltage to be measured.

621.317.35:621.397.5   **2846**

A Line Strobe Monitor for Investigating Television Waveforms—E. Davies. (*Jour. Telev. Soc.*, vol. 6, pp. 336–342; January/March, 1952.) Problems in the design of an oscilloscope for television waveforms are discussed generally. Requirements in respect of timebase range, display cro and associated video amplifier are specified, and diagrams and description are given of a Marconi instrument satisfying these requirements.

621.317.352:621.392.52   **2847**

Null-Balance Technique for Filter Measurement—E. R. Wigan. (*Wireless Eng.*, vol. 29,

pp. 195–196; July, 1952.) Description of a simple and accurate method of determining points on the frequency response curve of a filter, using only laboratory apparatus usually available.

621.317.374   2848

An Approximate Method for Deducing Dielectric Loss Factor from Direct-Current Measurements—B. V. Hamon. (*Proc. IEE* (London), Part IV, vol. 99, p. 151–155; July, 1952. Summary, *ibid.*, Part II, vol. 99, pp. 291–293; June, 1952.) Description of a method for deducing the approximate loss factor of a solid dielectric from the charging current that flows after the sudden application of a direct voltage. Under certain conditions, the loss factor at a frequency of f cps can be found from the charging current at a time $0.1/f$ second after application of the voltage. The useful range of the method is discussed and results obtained are compared with bridge measurements.

621.317.382.029.6   2849

Absolute Power Measurement at Microwave Frequencies—A. L. Cullen. (*Proc. IEE* (London), Part IV, Monograph No. 23; April, 1952; *ibid.*, Part II, vol. 99, pp. 183–186; April, 1952; summary only.) See 1052 of May.

621.317.382.029.6   2850

A General Method for the Absolute Measurement of Microwave Power—A. L. Cullen. (*Proc. IEE* (London), Part IV, Monograph No. 24; April, 1952; *ibid.*, Part II, vol. 99, pp. 186–188; April, 1952; summary only.) Presents the theoretical basis of the method noted in 2849 above, taking as a starting point Maclean's "resonator action" theory (*Quart. Jour. Appl. Math.*, vol. 2, p. 329; 1944). Any loss-free cavity with input and output wave-guides, and containing a movable member, can be used; the force on the movable member can be related to the power flow through the system by subsidiary experiments involving measurements of length only.

621.317.715:537.223.1   2851

Short Note on a Piezoelectric Vibration Galvanometer—H. Lueg. (*Z. Angew. Phys.*, vol. 4, pp. 16–17; January, 1952.) The instrument described is based on Golay's electrometer (3432 of 1937), with a three-stage amplifier incorporated; it will stand heavy overloading.

621.317.729:537.291   2852

Factors affecting the Design of an Automatic Electron-Trajectory Tracer—K. F. Sander, C. W. Oatley and J. G. Yates. (*Proc. IEE* (London,) Part III, vol. 99, pp. 169–176; July 1952. Discussion pp. 177–179.) Discussion of a method in which information derived from measurements on a model system in an electrolyte tank is passed automatically and continuously to a differential analyzer, where the equations of motion of the electron are integrated and its trajectory traced on an output table. Results obtained indicate that a digital computer would be preferable to the differential analyzer.

621.317.734   2853

Two Electronic Resistance or Conductance Meters—L. B. Turner. (*Proc. IEE* (London), Part II, vol. 99, pp. 209–216; June, 1952. Discussion, pp. 216–219.) Description of mains-fed instruments with respective ranges of 10^4–$1.2 \times 10^{12}\Omega$ and 260–$3.3 \times 10^{9}\Omega$. Scale accuracy is not affected by changes of tubes.

621.317.755:621.385.3/.5   2854

A Simple Valve Comparator—B. C. Foster. (*Electronic Eng.*, vol. 24, pp. 220–223; May, 1952.) Description of cro equipment using eight tubes and providing simultaneous display of the I_a/V_a characteristics of two tubes at a series of grid voltages.

621.317.761+621.396.621.54   2855

Direct-reading Frequency Measuring Equipment for the Range of 30 CPS to 30 MC—L.R.M. Vos de Wael. (*Proc. I.R.E.*, vol. 40, pp. 807–813; July, 1952.) English version of paper noted in 743 of April, supplemented by a brief description of a double-heterodyne receiver adapted from that described previously by van der Wijck (2041 of 1949) for use when measuring the frequency of a remote transmitter.

621.317.761   2856

Frequency-Deviation Meter plots Drift—N. C. Hekimian. (*Electronics*, vol. 25, pp. 134–135; June, 1952.) The circuit uses two type-6BN6 gated-beam tubes [1292 of 1950 (Adler)] in a limiter-discriminator arrangement. Output pulses of variable width are integrated and applied to a bridge circuit containing a microammeter calibrated in kc.

621.317.772   2857

A Versatile Phase-Angle Meter—G. N. Patchett. (*Electronic Eng.*, vol. 24, pp. 224–229; May, 1952.) A direct-reading cro apparatus with frequency range 20 cps–20 kc, input-voltage range 1–500 v and phase-angle range 90° lag to 90° lead, the accuracy being to within ± 2 per cent.

621.317.772   2858

Versatile Phase-Angle Meter—D. S. Gordon: and G. N. Patchett. (*Electronic Eng.*, vol. 24, p. 343; July, 1952.) Comment on 2857 above and author's reply.

621.317.79:621.396.619   2859

A New Modulation Meter—E. Rülker. (*Nachrichtentechnik*, vol. 2., pp. 47–51; February, 1952.) Description of an instrument with a logarithmic scale from -50 db to $+5$ db, suitable for the frequency range 40 cps–15 kc. The logarithmic scale is achieved by using the interelectrode path of a diode, operated in the initial-current region, as one section of the amplitude-dependent voltage divider.

621.317.79:621.396.62   2860

An Automatic Circuit Checker for Radio Receivers—V. J. Cox. (*Electronic Eng.*, vol. 24, pp. 258–263; June, 1952.) Description of test gear for use on a production line. Receivers with incorrect wiring or faulty components are rejected and the location of faults indicated, so that they can be diagnosed and corrected by comparatively unskilled workers. The method is based on comparison with a standard receiver.

621.319.43.089.6   2861

Equipment for Alignment of Ganged Variable Capacitors—J. Schmidt. (*Nachrichtentechnik*, vol. 2, pp. 55–57; February, 1952.) Description of apparatus for rapid comparison of the capacitance/angle curve of a variable capacitor with that of a standard or of a ganged unit, with extension to the alignment of three units.

621.396.615.14.029.63/.64   2862

A Generator for Measurements at U.H.F.—P. Clostre and R. Wallauschek. (*Ann. Télécommun.*, vol. 7, pp. 196–204; April, 1952.) A description is given of equipment for making all the measurements necessary for determining the characteristics of traveling-wave tubes and for various measurements on waves in the range 5–12 cm. The equipment comprises (*a*) a klystron oscillator in a perfectly screened enclosure, (*b*) screened aperiodic detector, (*c*) screened wavemeter with a coupling system practically aperiodic throughout the range of the klystron, (*d*) direct output with variable coupling, (*e*) calibrated attenuator with input and output impedances matched to external lines, (*f*) modulator. The total range is covered in three sections with adequate overlap, using different klystrons. The direct output power is of the order of 50–100 mw, the attenuator giving outputs variable from 1mw to 10^{-16}w. Theory of the attenuator is given in an appendix.

OTHER APPLICATIONS OF RADIO AND ELECTRONICS

621.38.001.8:543/545   2863

Recent Developments in Electronic Instrumentation for Chemical Laboratories—F. Gutmann. (*Proc. I.R.E.* (*Australia*), vol. 13, pp. 11–26; January, 1952.) See 2565 of October.

621.383.2   2864

Some New Image-Convertor Tubes and their Applications—J. A. Jenkins and R. A. Chippendale. (*Electronic Eng.*, vol. 24, pp. 302–307; July, 1952.) Description of the special features of the Mullard tubes, types ME 1200, ME 1201 and ME 1202, and their application in infrared pyrometry and photography, in high-speed photography, in process control for light-sensitive materials, and for brightness intensification in X-ray radiography.

621.385.833   2865

The Axial Potential of the Three-Electrode electron] Lens—P. Grivet and M. Bernard. (*Jour. Phys. Radium*, vol. 13, pp. 47–48; January, 1952.) See also 2000 of August.

621.385.833   2866

The Field Ion Microscope—E. W. Müller. (*Z. Phys.*, vol. 131, pp. 136–142; December, 1951.) By reversing the polarity of the voltage in the field electron microscope, adsorbed atoms are torn from the tip of the object as positive ions and form an image of its surface on the luminescent screen. Fields up to 300 MV/cm have been used. Resolution is better than that of the electron-emission microscope because the tangential velocities are lower. The required image intensity is attained by continual reinforcement of the desorption process.

PROPAGATION OF WAVES

621.396.11   2867

Polarization of Radio Waves reflected from the Ionosphere—B. Landmark. (*Jour. Atmos. Terr. Phys.*, vol. 2, pp. 254–255; 1952.) Results of vertical-incidence measurements in Norway are shown. The third magneto-ionic component or Z trace was polarized in the same way as the ordinary wave [See also 3090 of 1951 (Hogarth)]. Correlation was found between mean values of absorption at 3.6 mc and the axis ratio ϵ of the polarization ellipse. During magnetic storms and auroral displays the value of ϵ varied considerably.

621.396.11:535.515   2868

Comparison of Double Refraction in Crystals and in the Ionosphere—G. Lange-Hesse. (*Arch. Elekt. Übertragung*, vol. 6, pp. 149–158; April, 1952.) Differences between the double-refraction phenomena of light in crystals and of em waves in the ionosphere are investigated using tensor algebra. Whereas the permittivity tensor for the crystal in the direction of the optic axis is symmetrical, that for the ionosphere in the direction of the earth's magnetic field includes an antisymmetrical term, giving rise to double refraction. Using the concept of normal surfaces, it is shown that there is no direction of propagation in the ionosphere for which double refraction disappears.

621.396.11:551.510.535   2869

Lunar Variations in F$_2$-Region Critical Frequency at Singapore—B. W. Osborne. (*Nature* (London), vol. 169, pp. 661–662; April 19, 1952.) Results of measurements of f_0F_2 from November 1948 to June 1951 inclusive are analyzed and show a definite correlation with the moon's phase, with well-marked maxima at the first and last quarter.

621.396.11.029.45   2870

The Ionospheric Propagation of Radio Waves of Frequency 16 kc/s over Distances of

about 200 km—R. N. Bracewell. (*Proc. IEE* (London), Part IV, vol. 99, pp. 217–228; July, 1952. Summary, *ibid.*, Part III, vol. 99, pp. 217–221; July, 1952.) Measurement of phase, the essential feature of the investigations, was effected by comparison with a reference signal sent from the transmitter by land line. The field variations at 200-km range are explicable in terms of a constant ground-wave and multiple reflections from the ionosphere, which are more evident than at 90-km range. The apparent height of reflection varies approximately as $5.5 \log_e$ seconds χ + constant, where χ is the sun's zenith distance. The reflection height for summer noon is 65 km and for the night 82 km.

621.396.11.029.45   2871
The Ionospheric Propagation of Radio Waves of Frequency 16 kc/s over Distances of about 540 km—W. C. Bain, R. N. Bracewell, T. W. Straker and C. H. Westcott. (*Proc. IEE* (London), Part IV, vol 99, pp. 250–259; July, 1952. Summary, *ibid.*, Part III, vol. 99, pp. 226–228; July, 1952.) An account of investigations, carried out at Aberdeen during 1940–1944 and May–October 1949, of the amplitude and phase of signals from the Rugby transmitter. Diurnal and seasonal effects are noted. The results show that reflection takes place at an apparent height of about 74 km by day and 92 km at night, with effective reflection coefficients of about 0.27 and 0.55 respectively in the summer of 1949. At shorter ranges the downcoming wave is approximately circularly polarized, whereas at 535 km the polarization was almost linear during morning twilight. The results confirm the conclusion of Weekes (1491 of 1950) that there is a marked change in the propagation characteristics of 16-kc waves in passing from 300 to 500 km.

621.396.11.029.45:551.510.535   2872
An Explanation of Radio Propagation at 16 kc/s in Terms of Two Layers below E Layer—Bracewell and Bain. (See 2804.)

621.396.11.029.45:551.510.535   2873
The Oblique Reflexion of Long Wireless Waves from the Ionosphere at Places where the Earth's Magnetic Field is regarded as Vertical—J. Heading and R. T. P. Whipple. (*Philos. Trans. A.*, vol. 244, pp. 469–503; April 3, 1952.) Wilkes's model of the ionosphere, with two distinct regions (2548 of 1947), is used and wave-propagation equations in relatively simple form are derived. For a model in which ion density increases exponentially with height and collision frequency is constant over the range of height of the reflecting region, exact solutions are found for these equations, yielding exact expressions in terms of factorial functions for the reflection coefficients of the two regions separately; the total effect of the ionosphere on an incident wave is obtained by properly combining these coefficients. Apparent height of reflection is defined in terms of the phase of the reflected wave. The results of the theory are presented in graphical form for a model approximating to the "tail" of a Chapman region, and a comparison is made with experimental observations.

621.396.11.029.51   2874
Polarization Measurements of Low Frequency Echoes—E. L. Kilpatrick. (*Jour. Geophys. Res.*, vol. 57, pp. 221–226; June, 1952.) Report of National Bureau of Standards investigations at Sterling, Va. Plane-polarized 160-kc pulses of about 80-μs duration were transmitted vertically from a rectangular loop in the E-W vertical plane. The elliptical pattern of the polarization of the reflected signals under stable conditions was oriented 60°–70° east of magnetic north, the axis ratio varying from 3 to 5, with left-hand rotation of the polar vector. During unstable periods the duration of the echoes usually increased or the echoes split into separate portions with different polarizations.

621.396.11.029.51   2875
A Note on the Polarization of Low Frequency Ionosphere Echoes—J. M. Watts. (*Jour. Geophys. Res.*, vol. 57, pp. 287–289; June, 1952.) Report of National Bureau of Standards investigations at Sterling, Va., using circularly polarized 160-kc pulses. Continuous records of the height at which reflection occurred were obtained on slowly moving film. Daytime observations confirmed the results obtained with plane-polarized waves [2874 above (Kirkpatrick) and 3037 of 1950 (Benner et al.)]. Effects at night were more complex.

621.396.11.029.6:551.594.5   2876
The Fading Rate of Ionospheric Reflections from the Aurora Borealis at 50 Mc/s—K. Bowles. (*Jour. Geophys. Res.*, vol. 57, pp. 191–196; June, 1952.) Measurements have been made of the rapid flutter type of fading associated with auroral propagation (2002 of 1951). The observed fading rates are about ten times greater than might be expected from normal ionospheric drift velocities. For communication by auroral reflection to be effected, in most cases both transmitting and receiving antennas must have the same polarization and be pointed approximately in the direction of the visible aurora. Speech transmission on AM 50-mc waves is occasionally possible.

RECEPTION

621.396.621   2877
Twelve Years' Progress in the Design of Domestic Broadcast Sound Receivers—F. T. Lett. (*Jour. Brit. I.R.E.* vol. 12, pp. 254–265; April, 1952.) A review of receiver-design trends, particularly in the post-war period, and of progress in the design of components resulting in the production of small portable receivers with good performance.

621.396.621   2878
The Marconi Single-Sideband Receivers—Types HR92 and HR93—C. P. Beanland and F. I. Rickaby. (*Marconi Rev.*, vol. 15, pp. 60–70; 2nd Quarter 1952.) Special features and performance data for (a) Type-HR92, a double-diversity, dual-channel ssb receiver pre-tuned to any three spot frequencies in the range 3–27.5 mc, (b) Type-HR93, a single-channel version with six spot frequencies.

621.396.621:517.432.1   2879
Theory of the Impulse Response of Receivers. Application of Heaviside Operators and the Duhamel Integral—R. Kitai. (*Proc. IEE* (London), Part III, vol. 99, pp. 221–222; July, 1952.) Summary only. The response of a simple superheterodyne receiver to a step-voltage input is analyzed by application of operational analysis, the voltage/time responses being determined at the grid of the converter valve, at the grid of the IF amplifier tube, and at the detector output. The indicial-response equations given can also be used, together with the Duhamel integral, when the input differs from a step voltage. The theory was verified by measurements on a receiver.

621.396.621:621.317.79   2880
An Automatic Circuit Checker for Radio Receivers—Cox. (See 2860.)

621.396.621.018.12:621.3.018.783   2881
Low-Frequency Distortion due to H. F. Phase Shift—J. Zakheim. (*Radio franç.*, no. 4, pp. 22–24; April, 1952.) If the oscillating circuit of a receiver is exactly tuned to the frequency of a modulated carrier, the phase advance of the modulation voltage, resulting from the equal and opposite phase shifts of the two hf sidebands, produces no nonlinear distortion and is of secondary importance in the transmission of speech or music. If, however, the receiver circuit is mistuned, the phase shifts of the two sidebands are not equal and distortion is produced which depends on the difference (τ) between the phase shifts of the side-

bands and on the depth of modulation (m). Curves are given showing the distortion as a function of τ for $m = 30$, 60 and 90 per cent. For receivers with more than one tuned circuit the phase shift introduced by each circuit may result in a total distortion too great to be tolerable.

621.396.621.54 + 621.317.761   2882
Direct-Reading Frequency Measuring Equipment for the Range of 30 c/s to 30 Mc/s—Vos de Wael.—(See 2855).

621.396.622 + 621.396.621:621.396.619.11   2883
The Synchrodyne and Coherent Detectors—D. G. Tucker. (*Wireless Eng.*, vol. 29, pp. 184–188; July, 1952.) The performance of the synchrodyne and the coherent detector with respect to signal/noise ratio, when the input signal is accompanied by noise in its own allotted frequency band, is compared with that of the ordinary so-called "linear" detector in which the received signal is subjected to plain rectification. There is a very considerable difference in the output noise spectra when the signal is absent or very small, and in the "linear" detector the noise produces a dc output which is absent in the case of the coherent detector. Thus, although there is little difference in performance on continuous envelope-modulated signals, the coherent detector is much better on pulse or intermittent signals. The synchrodyne has basically the same performance in these respects as the coherent detector, but may be slightly inferior, due to noise appearing in the output of its local oscillator.

621.396.622.7:621.29.64:621.396.615.142   2884
Velocity-Modulated Detector—F. N. H. Robinson. (*Wireless Eng.*, vol. 29, pp. 200–202; August, 1952.) An explanation is given of the action of a microwave detector in which the rf signal is fed to a helix through which an electron beam passes. If the beam velocity is approximately the same as the wave propagation velocity along the helix, the beam emerges from the helix with vm and the presence of the signal results in a change of the current picked up by a collector on which the beam impinges. The sensitivity of such a device is estimated, taking account of possible noise sources, and is found to be some hundred times better than that for a video crystal. Experiments at wavelengths of 10 cm and 3 cm were carried out with a simple type of structure having an inefficient collector. Sensitivities of 10^{-9} w at 10 cm and 10^{-11} w at 3 cm with a video bandwidth of 1 mc were obtained, and 10^{-11} w at 10 cm, using a higher load resistance, with a video bandwidth of 1 kc. The sensitivity could be greatly improved by a better design of collector.

621.396.81   2885
On the Theory of Measurement of Weak Signals having a Continuous Spectrum—V. S. Troitski. (*Zh. Tekh. Fiz.*, vol. 21, pp. 994–1003; August, 1951.) Weak signals are defined as signals the spectral density of which is lower than that of the noise fluctuations in the measuring apparatus. Two methods are available for measuring such signals, viz. the compensation method and the modulation method. The theory of these methods is discussed and a comparison between the two is made. The following main conclusions are reached: (a) the sensitivity of both methods can be made as high as desired by narrowing the bandwidth of the filter; (b) the advantage of the modulation method is that it eliminates the effect of slow random modulation of the noise level.

621.396.82:621.396.619.13   2886
Interference in F.M. Reception caused by Strong Interfering Transmitters—M. Kulp. (*Arch. elekt. Übertragung*, vol. 6, pp. 143–148; April, 1952.) The intensity of the interference at the receiver output depends on the ratio q between the amplitudes of the unwanted and wanted signals at the limiter input, and changes

suddenly as this ratio vars through unity. Formulas are derived and tabulated for the interference caused by an uvanted FM signal of frequency equal to or dferent from that of the wanted signal, and r the interference caused by an unwanted AMsignal of different frequency, for $q < 1$. The ndifications to the formulas required to meet the case $q > 1$ are indicated.

STATIONS AND COMUNICATION SYSTEM

621.39.001.11:519.272   **2887**
The Mathematical Treatment of Random Phenomena in the Stuy of Oscillations (Norbert Wiener's Theory)—F. A. Fischer. *(Fernmeldetech. Z.* vol. 5, p. 151–158; April, 1952.) An outline of the esentils of Wiener's theory of harmonic analys, wih examples of its application.

621.39.001.11:621.394.14.04.15   **2888**
Efficient Coding—B. M. Oliver. *(Bell Sys. Tech. Jour.,* vol. 31, pp. 24 750; July, 1952.) "This paper reviews briefly a few of the simpler aspects of communication theory, especially those arts which relate to the information rate and channel capacity required for samp d, quatizd messages. Two methods are then discussed whereby such messages can be onverted to a 'reduced' form in which the successive samp are more nearly independent an for which the simple amplitude distributic is more ked than in the original message This redu d signal can then be encoded into binary dus with good effici ney using a Sannon-Fano ode on a symbol-ty-symbol (or airly-pai basis. The usual in ficiency whi results tr ignoring the correlation betwee messige ments is lessened because this co relat on is k s in the reduced message."

621.391:621.396.619 16   **2889**
A Recent Development i Communication Technique—C. W. arp. *(Proc. IEE* (London), Part III, vol. 99, pp. 181 86; July, 1952.) A description is given of a proposed new system of communication, tntaively termed the "ambiguous-index system," which has been evolved as a logial deelopment of earlier systems. The system ha certain practical advantages, and it is sugg ted that one variant, which is very nearly eqivdent to the theoretical 2-digit multiple-level pem system, could be of great value when the available frequency bandwidth is insufficient for binary coding by as many as six dicts. The new system permits a satisfactory classificat on of the system commonly known as "puls d FM," which represents a special case of the "ambiguous-index system," which is very losely related to signal coding systems. A modication of the "pulsed FM" system probabl represents an almost ideal system of transmission for the case where the available bandwidth is only about twice that required for ssb working.

621.396.619.018.782.4   **2890**
Harmonic Distortion of Modulation—E. G. Hamer. *(Wireless Eng.,* vol. 29, pp. 212–216; August, 1952.) Discussion of distortion due to echoes in radio systems, in particular, that due to echoes in antenna feeders. The distortion of AM and of FM signals is calculated and the results are presented in the form of abacs from which the amount of distortion can be readily determined for a wide range of modulation frequencies and depths of modulation. The equations and abacs show that a FM system is much more vulnerable to the effects of echo signal than an AM system.

621.396.619.11/.14:621.396.822   **2891**
On the Distribution of Energy in Noise- and Signal-Modulated Waves: Part 2—Simultaneous Amplitude and Angle Modulation— D. Middleton. *(Quart. Appl. Math.,* vol. 10, pp. 35–36; April, 1952.) The energy distribution is determined for a carrier modulated simultaneously in amplitude and phase (or frequency) by a pair of random-noise waves, one of which is delayed relatively to the other, coherence between the two modulations being assumed. Mean power and spectral distribution are derived from the autocorrelation functions and shown as graphs. The angle modulation results in a redistribution of energy without changing the total power. Cross-correlation terms are responsible for the spectral asymmetry not previously found. As the relative delay increases, spectral maxima oscillate about the carrier frequency in decreasing swings, so that symmetrical spectra are still obtained for particular delay values. Clipping due to over-modulation causes little spreading of the spectra, the shapes of which are mainly controlled by the degree of angle modulation. Special cases considered are those of no coherence and of coherence without over-modulation. The latter has applications to the noise output of magnetrons. Part 1: 2333 of August.

621.396.619.13:517.564.3   **2892**
Spectrum of a Frequency-Modulated Wave —W. C. Vaughan. *(Wireless Eng.,* vol. 29, pp. 217–222; August, 1952.) Expressions involving Bessel functions are derived for the frequency spectrum of a FM wave and for the side-band pattern. Simple arithmetical methods of evaluating Bessel functions to an accuracy adequate for most practical requirements are described in detail. A graphical construction showing the relation between $J_{n-1}(m)$, $J_n(m)$ and $J_{n+1}(m)$ forms the basis of a method for determining a limiting value for the ratio between two coefficients of adjacent orders. This is illustrated numerically for $J_4(5)/J_5(5)$, and the values of $J_n(5)$ are then computed for values of n from 1 to 8 and compared with the values given in standard tables. A second method, with numerical example, for evaluating Bessel coefficients is also given, and the spectrum of a FM wave with a modulation index of 5 is shown on a graph which also demonstrates the geometrical relation between the magnitudes of alternate components.

621.396.619.16:621.394.3   **2893**
A System for the Transmission of Teleprinter Signals through Narrow Frequency Bands—A. F. Boff. *(Marconi Rev.,* vol. 15, pp. 49–59; 2nd Quarter 1952.) Modern communication theory is applied in the experimental equipment described. The use of multi-level pulse coding in place of the conventional binary digits of the telegraph code reduces the working bandwidth to only 10 cps per channel, but the signal/noise ratio for satisfactory operation is higher than that for normal telegraphy.

621.396.65:621.385.029.6   **2894**
Travelling-Wave Valves in Radio Beam Links—G. Goudet. *(Ann. Télécommun.,* vol. 7, pp. 152–154; April, 1952.) The use of travelling-wave tubes enables amplification to be effected directly at relay stations, without recourse to an IF. Bandwidth and gain requirements can easily be met and a relay station for a 200-channel telephony system could have only three tubes, the first a low-noise type with a gain of about 20 db, receiving a signal of 10^{-8} w and furnishing 10^{-6} w, the second with a gain of 35 db giving an output signal of 3 mw, at which level a crystal mixer could be used for frequency changing, the third tube, also with a gain of 35 db, giving a final output of 1 w. Suitable tubes have been designed in the Centre National d'Études des Télécommunications (see 2936–2938 below).

621.396.65:621.396.615.142.2   **2895**
Operating Klystrons in F.M. Microwave Links—J. Cohn. *(Electronics,* vol. 25, pp. 124–127; June, 1952.) Discussion of methods of evaluating and correcting the distortion due to long transmission lines and mismatches. The analysis is based on graphical solutions of the equation derived from consideration of the equivalent circuit of a reflex klystron and its load.

621.396.65.029.63/.64   **2896**
Decimetre- and Centimetre-Wave Beam Links, with Reference to C.C.I.F. Requirements—K. O. Schmidt. *(Funk u. Ton.* vol. 6, pp. 176–190; April, 1952.) Discussion in particular of noise, signal level, and power requirements for telephony and television multichannel links, with a few technical data for (a) a p.phm 24-channel telephony system using a wavelength of 15 cm, (b) a FM wide-band (6 mc) television system, also on 15 cm, (c) a FM multichannel telephony system providing 480 channels of 15 cm, (d) a similar system providing 1,440 channels on 7.5 cm.

621.396.712.3   **2897**
The New Broadcasting Studies for the Laibach Transmitter—V. Dušan. *(Radio Tech.* (Vienna), vol. 28, pp. 173–176; April, 1952.) General description of the arrangement and acoustic properties of the various studios. An illustration shows the special treatment of the walls of the large studio.

621.396.721   **2898**
Portable Field Radio Equipment SUF-21K —K. Behr and H. Norrby. *(Ericsson Rev.,* no. 1, pp. 22–28; 1952.) Description of Swedish-made telephony apparatus mainly designed for military use. The frequency range is about 8 mc in the band 30–50 mc; FM is used. Power ransmitted is up to 4 w, giving an effective range of 15–25 km.

621.396.97.029.62:621.396.619.13   **2899**
V.H.F. Broadcasting—J. R. Brinkley. *(Wireless World,* vol. 58, p. 279; July, 1952.) Brief report based on a survey carried out in the U.S.A. early in 1952, and relevant to current British plans. Fifteen years after the establishment of the first FM stations, only 5 per cent of radio sets and 2 per cent of television sets are for FM sound broadcasting. FM receivers cost much more than AM receivers and many manufacturers have given up FM models; the number of FM stations has also dropped. FM seems likely to be successful for filling gaps in the AM service rather than for supplanting it.

SUBSIDIARY APPARATUS

621-526   **2900**
A System utilizing Coarse and Fine Position-Measuring Elements Simultaneously in Remote-Position-Control Servo Mechanisms —J. C. West. *(Proc. IEE* (London), Part II, vol. 99, pp. 135–141; April, 1952.)

621.316.722.078.3   **2901**
The Design of Series-Parallel Voltage Stabilizers—S. N. Pocock and F. A. Benson. *(Electronic Eng.,* vol. 24, p. 343; July, 1952.) Comment on 2045 of August and author's reply.

TELEVISION AND PHOTOTELEGRAPHY

621.397.24/.26   **2902**
Television from Paris—*(Wireless World,* vol. 58, pp. 298–300; August, 1952.) Outline description of the recent Paris-London relay. From Lille, the end of the existing French relay system, the ordinary broadcast and a second transmission of the program by a 9-kmc portable set were received at Cassel, 28 miles away. Here a 405-line image orthicon was focused on the long-persistence screen of a 15-inch cr tube presenting the 819-line picture. A special land line conveyed the 50-cps mains frequency of the British grid to Cassel for frame-scan locking. Subsequent links were Cassel-Alembon, 7kmc; thence Swingate-Wrotham-London, 4.5 kmc. A vhf communication system linked repeater stations from Cassel to London. A land line direct from Paris to London transmitted the program sound.

621.397.242   **2903**

Carrier-Frequency Transmission on Television Cables—R. Hoffmann and J. Müller. (*Fernmeldetech. Z.*, vol. 5, pp. 173–177; April, 1952.) Various possible methods were considered for the transmission of television signals over distances up to 25 km between studio and transmitter or points in the telephone network. For 625-line signals, 21-mc carrier-frequency transmission on coaxial cables was adopted, equalizers (range 15–27 mc) and amplifiers being fitted at intervals of 6–7 km. Equipment used on a 14-km link between Berlin-Templehof and the exhibition hall at Witzleben is described, the quality of the transmission being illustrated by oscillograms of square-wave signals with frequencies of 0.3 and 0.5 mc. A diagram shows the over-all amplitude and phase characteristics of the system.

621.397.5:519.24   **2904**

Statistics of Television Signals—E. R. Kretzmer. (*Bell Sys. Tech. Jour.*, vol. 31, pp. 751–763; July, 1952.) Measurements have been made of some basic statistical quantities characterizing picture signals. These include various amplitude distributions, auto-correlation, and correlation among successive frames. The methods of measurement are described, and the results are used to estimate the amount by which the channel capacity required for television transmission may be reduced through exploitation of the statistics measured.

621.397.5:519.272.1   **2905**

Experiments with Linear Prediction in Television—C. W. Harrison. (*Bell Sys. Tech. Jour.*, vol. 31, pp. 764–783; July, 1952.) The correlation between the elements of a signal makes possible the prediction of the future of the signal in terms of the past and present. A method of prediction is described which does not make full use of the past, but is effective with certain signals and is relatively simple. In this method the prediction for the next signal sample is simply the sum of previous signal samples each multiplied by an appropriate weighting factor which depends on the statistics of the signal. The relatively simple apparatus required for experiments on linear prediction of television signals is described and results obtained with it are illustrated and discussed.

621.397.5:535.623/.624   **2906**

Progress in Colour Television—R. L. Smith-Rose. (*Nature* (London), vol. 169, pp. 563–566; April 5, 1952.) A review of the development of the art, principally in the U.S.A.

621.397.5:535.623   **2907**

Image-Orthicon Color-Television Camera Optical System—L. T. Sachtleben, D. J. Parker, G. L. Allee and E. Kornstein. (*RCA Rev.*, vol. 13, pp. 27–33; March, 1952.) A relay lens system and cross-shaped dichroic image divider provide three separated images. The green light is focused directly and the red and blue beams are reflected on to their respective image orthicons. Astigmatism introduced by the divider is corrected by means of two mutually perpendicular plates between the field lens and the relay lens system. The aperture of the system may be remotely controlled by a motor-driven iris.

621.397.5:535.623   **2908**

The R.C.A. Color-Television Camera Chain—J. D. Spradlin. (*RCA Rev.*, vol. 13, pp. 11–26; March, 1952.) Description of features of the system developed for commercial operation, particularly the construction, operation and adjustments of the camera, the target-scanning circuit, the three channel amplifiers and the monochrome and color monitoring arrangements.

621.397.5:535.623:621.317.2   **2909**

The N.B.C. New York Color-Television

Field-Test Studio—J. R. DeBaun, R. A. Monfort and A. A. Walsh. (*RCA Rev.*, vol. 13, pp. 107–124; March, 1952.) Description of the studio arrangement, apparatus and operation for standard tests and adjustment of equipment.

621.397.5:621.317.35   **2910**

A Line Strobe Monitor for Investigating Television Waveforms—Davies. (See 2846.)

621.397.5:621.396.73   **2911**

A Developmental Portable Television Pickup Station—L. E. Flory, W. S. Pike, J. E. Dilley and J. M. Morgan. (*RCA Rev.*, vol. 13, pp. 58–70; March, 1952.) Illustrated description of the unit which consists of a 50-lb pack and an 8-lb vidicon camera. Video and sound channels operate on a common uhf carrier; the range is up to $\frac{1}{2}$ mile. See also *Electronics*, vol. 25, pp. 98–101; June, 1952.

621.397.5(083.74)   **2912**

International Television Standards—(*Wireless World*, vol. 58, pp. 296–297; August, 1952.) Details of the 405-, 525-, 625- and 819-line systems approved by C.C.I.R. in 1951.

621.397.61:621.3.018.41.016.352   **2913**

Frequency Stability for Television Offset-Carrier Operation—P. J. Herbst and E. M. Washburn. (*RCA Rev.*, vol. 13, pp. 95–106; March, 1952.) For successful offset-carrier operation at uhf (1798 and 2653 of 1950) the stability requirements of a local monitor are such that single checks against a standard-frequency transmission may be inadequate. A highly stable crystal oscillator is described which can safely be used in areas where a primary standard is not readily available. Details are given of the treatment and mounting of the crystal, and of the performance of the unit on protracted tests. It meets the monitor tolerance of a frequency deviation $< \pm 5$ parts in 10^7 in a 30-day period.

621.397.611.2   **2914**

Performance of the Vidicon, a Small Developmental Television Camera Tube—B. H. Vine, R. B. Janes and F. S. Veith. (*RCA Rev.*, vol. 13, pp. 3–10; March, 1952.) The vidicon developed is 1 inch in diameter and about $6\frac{1}{4}$ inches long. Gamma, sensitivity, spectral response, persistence and life of three types of photoconductive layer, viz., amorphous Se, Sb_2S_3 and a modification of Sb_2S_3, are discussed. Sensitivities in the three cases are respectively 100, 300 and 50 ma per lumen.

621.397.62   **2915**

The First British Multi-Channel Television Receiver—W. D. Asbury, K. M. B. Wright and W. M. Lloyd. (*Jour. Telev. Soc.*, vol. 6, pp. 343–351; January/March, 1952.) A full description of the design is given, with circuit details and response curves. The receiver is a superheterodyne with the oscillator frequency on the high side and intermediate frequencies of 19.5 mc for sound and 16.0 mc for the picture. A single rf stage is used, and the pentode oscillator-mixer is designed to eliminate interaction between signal and oscillator circuits. The user can tune quickly to any one of the five channels.

621.397.621   **2916**

Simple Line-Scan Circuit—W. T. Cocking. (*Wireless World*, vol. 58, pp. 305–309; August, 1952.) The directly fed deflector-coil circuit (527 of March) can be adapted to provide hv by including an extra inductance which compensates the energy loss in the deflector coil during flyback. The operation and design of the circuit, including the waveform-correction element, are discussed. The complete circuit has proved noncritical in adjustment and satisfactory for operating at 53° tube at 10 kv.

621.397.621   **2917**

The Required Figure of Merit of Frame Deflector Coils—E. Emms. (*Electronic Eng.*, vol. 24, pp. 238–23 May, 1952.) A method is indicated for estimating the required "electrical goodness" for frame deflector coils to be used with a particular output pentode operated at given line voltage to scan a cr tube operated at given high voltage. Worked examples make use of results obtained previously (2060 of August).

621.397.621   **2918**

Faulty Interlacing—G. N. Patchett. (*Wireless World*, vol. 58, pp. 250–254 and 315–319; July and August, 1952.) Experiments are described using integrator- and differentiator-type frame-synchronizing-signal separator circuits in conjunction with blocking-oscillator and thyratron timebase circuits to investigate the influence of the shape of the synchronizing pulse on the accuracy of interlacing. For correct interlacing, the pulse from the separator must have a sharp leading edge, with a fixed time delay with respect to the first frame pulse, and must be identical in all respects for odd and even frames. A separator circuit satisfying these requirements is described, with illustrations of waveforms obtained; a 6F32 or 6F33 tube is used.

621.397.621.2   **2919**

Self-Focusing Picture Tube—A. Y. Bentley, K. A. Hoagland and H. W. Grossbohlin. (*Electronics*, vol. 25, pp. 107–109; June, 1952.) Description of an electrostatically focused electron gun designed for mass production. The focusing electrode is a cylinder at cathode potential and of larger diameter than the adjacent high-potential cylindrical cups. Supply-voltage variations have no significant effect on the focusing. The guns are manufactured as two separate units which are welded together after checking.

621.397.8   **2920**

The Evaluation of Picture Quality with Special Reference to Television Systems—L. C. Jesty and N. R. Phelp. (*Jour. Brit. IRE*, vol. 12, pp. 211–253; April, 1952.) Reprint. See 275 of February and 1127 of May.

621.397.8   **2921**

A Method of Measuring Television Picture Detail—G. G. Gouriet. (*Electronic Eng.*, vol. 24, pp. 308–311; July, 1952.) The "picture detail" over a given length of scan is defined as "the modulus slope of the brightness variation integrated over that length." Experimental equipment for measuring this quantity is described, and its applications to service monitoring and for comparison of the performances of picture-producing equipment are discussed.

TRANSMISSION

621.396.61   **2922**

A Multichannel Single-Sideband Radio Transmitter—L. M. Klenk, A. J. Munn and J. Nedelka. (*Proc. I.R.E.*, vol. 40, pp. 783–790; July, 1952.) Description of a transmitter designed for transoceanic communications and operating over the frequency band 4–23 mc; four telephone channels are available. The main feature is the use of a servo system permitting push-button tuning to any one of ten preselected operating frequencies in about 15 seconds.

621.396.61:621.396.645   **2923**

Amplifiers for Multichannel Single-Sideband Radio Transmitters—N. Lund, C. F. P. Rose and L. G. Young. (*Proc. I.R.E.*, vol. 40, pp. 790–796; July, 1952.) Design requirements for hf amplifiers with low interchannel modulation and adjacent-band radiation are discussed. The speech rating of the amplifier is most significant, but the tone rating is the basis of tube selection; a method for determin-

ing the relation between the two ratings is presented. Unbalanced circuits are preferred for low- and medium-power amplifiers, while balanced circuits may have advantage. for some high-power amplifiers. See also 2922 above.

621.396.61:621.396.97   **2924**
Remote Control of High-Power Transmitters—(*Jour. Brit. IRE*, vol. 12, p. 268; April, 1952.) The new 1 0-kw transmitter at Daventry for the BBC third program has been working unattended since 13th January 1952. The transmitter comprises two identical units, whose outputs are combined in a circuit that insures no transfer of power between the two sections. The switching on of the various power supplies in the correct sequence and at the correct time intervals is effected automatically by a system of interlocked relays. Automatic monitors [1491 of 19 1 Rantzen et al.)] continuously monitor the outputs from the two sections and shut down either half which develops a fault, special arrangements preventing complete stoppage of transmissions as a result of a comparatively minor fault.

621.396.619.24   **2925**
Single-Sideband Transmission by Envelope Elimination and Restoration—L. R. Kahn. (*Proc. I.R.E.*, vol. 40, pp. 803–806; July, 1952.) The system described eliminates the need for costly linear rf amplifiers in the transmitter. The psm component of the ssb signal is amplified by means of class-C amplifiers, while the af envelope is separately detected and amplified and recombined with the rf signal at the final stage. Experiments indicate that the performance of the system is equal to or better than that of the conventional transmitter with linear rf amplifier.

621.396.619.24   **2926**
Design of Modulation Equipment for Modern Single-Sideband Transmitters—A. E. Kerwien. (*Proc. I.R.E.*, vol. 40, pp. 797–803; July, 1952.) The transmitters considered are of the type using filters to suppress the unwanted sideband. Factors discussed include balance requirements, frequency stability, choise of IF, and methods of avoiding transmission of spurious signals. See also 2922 above.

TUBES AND THERMIONICS

621.314.65   **2927**
A High-Voltage, Cold-Cathode Rectifier—E. G. Linder, J. H Coleman and E. G. Apgar. (*Proc. I.R.E.*, vol. 40, pp. 818–828; July, 1952.) Description and theory are given of a mercury-vapor glow-discharge rectifier operating with an external magnetic field; the electron paths are similar to those in a cylindrical magnetron, but are modified by the use of end plates on the cathode. This rectifier has withstood peak inverse voltages up to 40 kv, and has been run at a peak inverse voltage of 30 kv for 3,000 hours. Operation at temperatures between 20° and 80°C has been satisfactory. The upper frequency limit lies between 120 and 240 kc. See also *Proc. Nat. Electronics Conf.* (Chicago) vol. 7, pp. 64–72; 1951.

621.314.7:621.396.822   **2928**
Barrier-Layer Interaction and Statistical Fluctuations in the Three-Electrode Crystal—H. F. Mataré. (*Z. Phys.*, vol. 131, pp. 82–97; December, 1951.) Noise in semiconductors is discussed. A particular problem arising in transistor operation is the increase of noise resulting from the control mechanism. This is effective over the whole frequency spectrum but is conveniently measured at hf. The interaction of the streams of charge carriers originating at the emitter and collector barrier layers leads to a large increase of the statistical fluctuations in the effective resistances. A theoretical treatment similar to Richardson's (1391 of 1950) is used. Interaction phenomena of a

similar nature are probably present in interstellar noise sources.

621.383.5:546.28   **2929**
Photoelectric Properties of Ionically Bombarded Silicon—E. F. Kingsbury and R. S. Ohl. (*Bell Sys. Tech. Jour.*, vol. 31, pp. 802–815; July, 1952.) The first Si photocells were cut from bulk Si containing a natural potential barrier between *n*-type and *p*-type material. Later cells were produced by subjecting the polished face of wafers about 0.025 inch thick to bombardment by positive ions of He of energy from 100 ev to 30 kev, the Si being kept at about 395°C. Typical spectral-response curves are illustrated and the photon efficiency as a function of wavelength is shown for a particular cell.

621.384.5:621.316.722:621.396.822   **2930**
Peak-Noise Characteristics of Glow-Discharge Voltage-Regulator Tubes—H. Bache and F. A. Benson. (*Electronic Eng.*, vol. 24, pp. 278–279; June, 1952.) Measurements have been made on tubes of 15 different types to obtain information about the dependence of noise on design and operational factors. Results are tabulated and shown in graphs. In all cases noise increases as tube current decreases. Tube specifications should state approximate limits for both peak and mean values of noise voltage throughout the current range.

621.384.5:621.316.722:621.396.822   **2931**
Mean-Noise Characteristics of Glow-Discharge Voltage-Regulator Tubes—H. Bache and F. A. Benson. (*Electronic Eng.*, vol. 24, pp. 328–329; July, 1952.) Results of measurements on six different types of voltage-regulator tube are shown graphically. The curves have the same general shape as in the case of peak-noise measurements (2930 above), the noise increasing with decreasing tube current, the increase becoming very rapid as the region of minimum current is approached. The relation between peak noise and mean noise is discussed briefly.

621.384.5:621.385.5:621.318.572   **2932**
The Development of a Multi-Cathode Decade Gas-Tube Counter—G. H. Hough. (*Proc. IEE* (London), Part IV, vol. 99, pp. 177–186; July, 1952.) Full paper. See 2657 of October.

621.384.5:621.385.5:621.318.572   **2933**
Some Recently Developed Cold-Cathode Glow-Discharge Tubes and Associated Circuits—G. H. Hough and D. S. Ridler. (*Electronic Eng.*, vol. 24, pp. 152–157, 230–235 and 272–277; April–June, 1952.) Construction and applications are described of the G1/370 K high-speed trigger tube and the multielectrode G10/240 E counting tube. See also 784 of 1951 and 2932 above.

621.385.029.6   **2934**
Effect of Thermal-Velocity Spread on the Noise Figure in Traveling-Wave Tubes—P. Parzen. (*Jour. Appl. Phys.*, vol. 23, pp. 394–406; April, 1952.) A method of Hahn (3521 of 1939 and 537 of 1940) is extended by working out a new theory for the diode drift tube and traveling-wave tube to include the effect of the thermal-velocity spread at vhf. Some computations of gain and noise figure for a traveling-wave tube amplifier are given for conditions met in practice.

621.385.029.6:621.396.65   **2935**
Two Traveling-Wave Valves for Radio Beam Links—(*Ann. Télécommun.*, vol. 7, pp. 150–204; April, 1952.) A symposium of seven papers, abstracts of which are given in 2936–2940 below, 2892 and 2894 above. 40 references.

621.385.029.6:621.396.65   **2936**
Two Traveling-Wave Valves for Radio Beam Links: General Results and Technol-

ogy—M. Kuhner and P. Lapostolle. (*Ann. Télécommun.*, vol. 7, pp. 155–168; April, 1952). A detailed account is given of the investigations resulting in the production of Type-M8 and Type-M11 tubes, starting from the experimental tubes previously described [2060 of 1950 (Blanc-Lapierre and Kuhner)]. The two tubes are of similar construction, but the envelope of the Type-M8 is provided with an extra seal assuring a direct contact between the output cone and the external circuit. The values of the gain and the pass band of both tubes are particularly high, 35–40 db and 800–1,000 mc for the M8 (8 cm wavelength) and 30 db and 150–200 mc for the M11 (11 cm wavelength), the noise factors being 20 db and 17 db respectively. Simple mounting and matching arrangements facilitate tube changing.

621.385.029.6:621.396.65   **2937**
Two Travelling-Wave Valves for Radio Beam Links: Measurements and Characteristics—P. Clostre and R. Wallauschek. (*Ann. Télécommun.*, vol. 7, pp. 169–172; April, 1952.) Results of measurements of gain, output power and bandwidth of Type-M8 and Type-M11 tubes are presented graphically. Measurements of the phase difference between input and output are described and curves are given for a Type-M11 tube showing gain and phase shift as functions of (*a*) helix voltage (target current constant), (*b*) target current (helix voltage constant). Normal operating voltages and currents are tabulated.

621.385.029.6:621.396.65   **2938**
Electron Guns for Travelling-Wave Valves [Types M8 and M11]—K. E. Picquendar. (*Ann. Télécommun.*, vol. 7, pp. 173–180; April, 1952.) Electron guns giving a cylindrical beam were first studied. In order to improve their efficiency, an investigation was made of the effect of the axial magnetic field applied; a simple method was developed for tracing the electron trajectories under the influence of a variable electric field and of a constant axial magnetic field. Magnetic screening was found necessary to reduce the deleterious effect of the magnetic field. With parallel-beam guns, the very high current density required at the cathode resulted in short tube life. To remedy this, electron guns giving convergent beams were adopted and means provided for preventing ion bombardment of the cathode. A method was devised for tracing approximately the electron trajectories for convergent-beam guns taking account of space-charge effects. Tests of the M86 electron gun in a series of tubes gave efficiencies of the order of 98 per cent. The efficiency of a later type, M88, was satisfactory for a wide range of currents (1–9 ma).

621.385.029.6:621.396.65:621.392.43   **2939**
Impedance Matching of Travelling-Wave Valves [Types M8 and M11] to Coaxial Lines—P. Clostre and R. Wallauschek. (*Ann. Télécommun.*, vol. 7, pp. 181–190; April, 1952.) An account is given of investigations of coupling arrangements, with a detailed description of the capacitive type of mashing unit finally adopted, which, in the case of the M8 tube, has a bandwidth of the order of 1–1.25 kmc for a swr < 1.5. Coupling of the 70-Ω coaxial cables to the cylindro-conical fittings soldered to the two ends of the helix is effected for the M11 tube by rings surrounding the glass envelope. For the M8 tube a series-parallel type of capacitive coupling is used. For a distance of a few centimeters from the ends, the pitch of the helix is progressively decreased to the value selected for the central section. Sectional diagrams of the coupling units are given and also input-impedance diagrams for both types of tube.

621.385.029.6:621.396.65:621.392.43   **2940**
Impedance Matching of a Travelling-Wave Valve [Type M8] to a Waveguide—P. Chavance

and L. Moutte. (*Ann. Télécommun.*, vol. 7, pp. 191–195; April, 1952.) Use is made of the properties of ridge waveguides in a matching unit with bandwidth >600 mc for a swr <1.7. A second arrangement permits accurate matching of a series of tubes over two bands, each 20 mc wide, centered on 3,640 and 3,920 mc, the swr being <1.1, a variable matching unit in conjunction with specially designed fittings for locating the tube in the fixed matching unit enabling the required swr to be quickly achieved when one tube is replaced by another.

621.385.032.24:621.94   2941
Grid-Winding Lathe—(*Electronics*, vol. 25, pp. 238–242; June, 1952.) Description of the process developed by Standard Telephones and Cables by which 1,000 tube grids are wound per hour on the new Brimar notch-and-swage lathe.

621.385.13:621.396.822   2942
Low-Frequency Noise Spectra of Hot-Filament Low-Pressure Discharge Tubes—H. Martin and H. A. Woods. (*Proc. Phys. Soc.* (London), vol. 65, pp. 281–286; April, 1952.) Measurements on hot-cathode discharges in Hg vapor confirm that, as previously reported by Cobine and Gallagher (2400 of 1947), the noise spectrum is continuous, with superposed peaks. The upper limit of the spectrum is in some cases above 10 mc. The peaks can be partly attributed to harmonic components of ionic relaxation oscillations. Most of the noise from the tubes investigated originated near the cathode.

621.385.2:546.289   2943
Temperature-Independent Crystal Diodes—R. Rost. (*Fernmeldetech. Z.*, vol. 5, pp. 177–178; April, 1952.) Tests carried out on Ge-W diodes, mass produced by a process involving temperatures of about 200°C, show that during cooling to room temperature the rectification factor in most cases becomes constant below about 80°C, this property depending on the use of a particular method of construction which results in diodes with rectification-factor variations <1 per cent over the range −20° to +70°C.

621.385.832   2944
Elementary Theory of the Generation of Electron Beams by means of Triode Systems: Part 2—Intensity and Structure of the Electron Beam—M. Ploke. (*Z. angew. Phys.*, vol. 4, pp. 1–12; January, 1952.) Langmuir's space-charge formula is used to develop formulas for the beam intensity and the current density at different cross-sections. Closed expressions are found relating aperture angle and crossover area with the dimensions and operating voltages of the system. Part 1: 2087 of August.

621.385.832   2945
Improved Cathode-Ray Tube for Application in Williams Memory System—W. E. Mutter. (*Elec. Eng., N. Y.*, vol. 71, pp. 352–356; April, 1952.) Description of the IBM-79, a 3-inch cr tube of storage type. See also 2258 of 1949 (Williams and Kilburn.)

621.396.615.14   2946
Generation of Oscillations at Centimetre Wavelengths—H. Severin. (*Tech. Mitt. schweiz. Telegr.-TelephVerw.*, vol. 29, pp. 466–476; December, 1951. In German.) Review of the physical principles and the performance of disk-seal, retarding-field, magnetron, klystron and traveling-wave tubes.

621.396.615.141   2947
Space-Charge Reactance Tube—L. E. Williams. (*Electronics*, vol. 25, pp. 166–182; June, 1952.) A method of measuring the variation of grid-cathode and grid-anode capacitance as a function of operating voltage is out-

lined. Applications based on the linear nature of this variation are described. These include frequency modulation of a Hartley oscillator by varying the grid bias of a reactance tube and electronic tuning of hf filters.

621.396.615.141.1   2948
A Wide-Range Oscillator in the Range from 8000 to 15000 Mc/s—R. W. Wilmarth and J. L. Moll. (PROC. I.R.E., vol. 40, pp. 813–817; July, 1952.) In retarding-field tubes of the general type described by Heil and Ebers (2677 of 1950), losses may be experienced in extracting the energy. Modifications are here described in which the energy is extracted by causing the repeller, which is part of the resonant circuit, to radiate into a waveguide. Either inductive or capacitive tuning may be used; in the latter case the efficiency of the oscillator is 2–4 per cent over the frequency range, with maximum efficiency occurring near the highest frequency. Inductive tuning appears to have certain intrinsic advantages which can be realized after certain mechanical and electrical difficulties have been overcome.

621.396.615.141.1   2949
Experimental Investigation of the Theory of Retarding-Field Oscillations—H. G. Unger. (*Frequenz*, vol. 6, pp. 89–98; April, 1952.) Measurements of the dynamic admittance of the grid-anode path of a cylindrical retarding-field tube were made by a resonance method, using external excitation at a frequency of about 500 mc, damping being applied to prevent self-excitation. The results are presented graphically and compared with the theory of Gundlach and Kleinsteuber (2389 of 1941 and 1866 of 1943). Discrepancies between the results and the theory for planar electrode systems are discussed.

621.396.615.141.2   2950
An Experimental Study of Low-Power C.W. Magnetrons having Few Segments—E. B. Callick. (PROC. I.R.E., vol. 40, pp. 836–843; July, 1952.) Measurements on two- and four-segment magnetrons are reported. The characteristics differ very little from those of multisegment tubes having the same ratio between cathode and anode diameters, except that the efficiency is lower and varies as a function of that ratio. High efficiency can be obtained if the cathode diameter is relatively small. Possible explanations are discussed.

621.396.615.141.2   2951
Note on the History of the Development of the Magnetron in Germany up to 1945—K. Fritz. (*Arch. elekt. Übertragung*, vol. 6, pp. 209–210; May, 1952.)

621.396.615.141.2:621.385.029.6   2952
Potentials and Electron Paths in Multisegment Magnetrons—K. Fritz. (*Arch. elekt. Übertragung*, vol. 6, pp. 211–215; May, 1952.) Theory is developed for the multisegment magnetron considered as a traveling-wave tube. A formula is derived for the threshold voltage, i.e. the lowest direct voltage for which electrons can reach a point in the interaction space in the absence of alternating voltage. Good agreement is obtained with the value found experimentally for the lowest anode voltage at which electrons reach the anode of a type-725A magnetron. The anode alternating voltage is expressed as a function of direct anode voltage and traveling-wave potential. The electrical and mechanical components of energy are evaluated, and a first-order second-degree differential equation is derived for the mean electron path; this can be solved in terms of a sine function.

621.396.615.142   2953
Velocity-Modulation Valves for 100 to 1000 Watts Continuous Output—B. B. van Iperen. (*Philips Tech. Rev.*, vol. 13, pp. 209–222; February, 1952.) In designing vm tubes for the

generation of cm waves, account must be taken of effects which may be disregarded in the elementary theory, such as the mutual repulsion of the beam electrons, damping due to finite transit time through the gaps, and deflection of electrons by the hf field at the input gap. Details are given of four experimental tubes incorporating guns, cavities and gaps designed to reduce these effects.

621.396.615.142:621.396.022.7.029.64   2954
Velocity-Modulated Detector—Robinson. (See 2884.)

621.385.032.216   2955
The Oxide-Coated Cathode. Vol. 1: Manufacture. [Book Review]—G. Herrmann and S. Wagener. Publishers: Chapman and Hall, London, Eng., 1951, 148 pp., 21s. (*Proc. Phys. Soc.* (London), vol. 65, p. 165; February 1, 1952.) "The book is well produced and illustrated and the information is authoritative and up-to-date."

621.385.032.216   2956
The Oxide-Coated Cathode. Vol. 2: Physics [Book Review]—G. Herrmann and S. Wagener. Publishers: Chapman and Hall, London, Eng., 1951, 302 pp., 42s. (*Proc. Phys. Soc.* (London), vol. 65, pp. 165–166; February 1, 1952.) "The publication of an extensive and lucid review which brings the subject back into the sphere of more general discussions ... will be generally appreciated. The present volume is based on work first published in Germany in 1944. It has since been brought up to date ... and the literature has been covered up to the beginning of 1950." See also 2603 of 1951. Vol. 1: 2672 above.

001.891:538.569.4.029.64   2957
University Research in Physics: Part 3—Research in Physics at Oxford University—J. A. Teegan. (*Beama Jour.*, vol. 59, pp. 81–87; March, 1952.) A short account of work on sw absorption specroscopy.

MISCELLANEOUS

025.45:621.3   2958
Work of the Electrical Engineering Commission for Universal Decimal Classification at the Rome Meeting—C. Frachebourg. (*Tech. Mitt. schweiz. Telegr.-TelephVerw.*, vol. 29, pp. 476–478; December, 1951.) A brief account of discussions on proposed modifications to various portions of the 621.3 classification, with emphasis on the proposals for a new main section 621.37 dealing with the technique of electric waves, oscillations and pulses.

621.38:001.891:359.4(41)   2959
Electronics in Naval Research—(*Electronic Eng.*, vol. 24, pp. 322–323; July, 1952.) A short account of the organization of the Royal Naval Scientific Service, with particular mention of the activities of the Admiralty Signal and Radar Establishment H.M. Underwater Detection Establishment, and the Services Electronics Research Laboratory.

621.396:061.3   2660
I.R.E.-U.R.S.I. Spring Meeting, Washington, D. C., April 21–24, 1952—(PROC. I.R.E., vol. 40, pp. 738–48; June, 1952.) Summaries are given of 71 technical papers presented at the meeting.

621.39.001.11:025.3/.4   2961
Information Theory and its Application to Taxonomy—D. K. C. MacDonald. (*Jour. Appl. Phys.*, vol. 23, pp. 529–531; May, 1952.) A possible source of confusion between the concept of information content and entropy in the theory of information is discussed and resolved. Information theory is then applied to taxonomy (classification of data) and several models are considered, representing various possible methods of filing, with the object of determining the optimum size of filing unit in relation to the available data.