VOLUME 5 □ NUMBER 1 JANUARY 1968

IEEE spectrum

features

Departments: *please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

the cover

The seemingly unrelated discoveries of negative-conductance phenomena in bulk semiconductors during recent years have been found to represent an extremely sophisticated theory. Our cover illustrates the phase-shift characteristic that causes negative conductance in avalanche transit-time devices. Turn to the article on page 47 to understand why these phenomena promise to revolutionize the fields of microwaves and solid-state physics.

The Venn diagram[1] or Karnaugh map[2] is a useful tool for analyzing and simplifying Boolean or logic functions; however, many find the method of constructing the map difficult to remember and "reading" the map prone to error. This article by Paul Kintner (from his book, "Electronic Digital Techniques," to be published by McGraw-Hill) describes a new form of the logic map which is based on an easily remembered entry rule and uses the pattern recognition capabilities of the eye to avoid reading errors. Deriving its name from the entry path, it is called a spiral map.

The Spiral Map: A New Form Of Karnaugh Map

The spiral map is formed directly from the ordered (truth) table of combinations known as a logical basis[3]. The table is based on 2^N combinatorial values listed horizontally for N Boolean variables. An example for N = 3 is given:

	0123	4567
A	0101	0101
B	0011	0011
C	0000	1111

As seen, successive columns of the "input" values can be considered to be binary numbers with A, the least significant digit; B, the next more significant digit; etc. These numbers are placed over the combinations (columns of the table) as shown.

The table is completed by listing the output values for each input value. For example, the following defines the "Exclusive Or" function for A,B:

	0123
A	0101
B	0011
Z	0110

The spiral map is constructed by entering values from the logical basis or table in groups of four going from left to right. Each group is entered onto a rectangular array as shown in Figure 1. Note how the order of entry can be imagined as a spiral pattern.

A unique feature of the spiral map is that the input values, as well as the output values, are entered. Actually, several maps are formed: one for each of the variables, and one for the function itself. The maps based on the input variables are, of course, fixed and are called *reference maps*. Figure 2 shows the three maps formed for the "Exclusive Or" function. The reference maps make it possible to quickly find the logic value for any square of the output map: it is only necessary to glance at the corresponding squares of the reference maps. For example, the square at the right is seen to be defined by A = 1, B = 0, or the Boolean value $A \cdot \overline{B}$.

A two-group map is formed as shown in Figure 3 with the second group placed below the first. Figure 4 shows the four maps (three reference and one output) for the function given by the output list Z = 0001

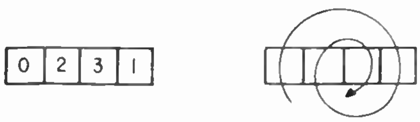


Figure 1. Basic Spiral Entry Pattern

0111, which is recognized to be the logic behavior of a binary adder.

A four-group map (based on N = 4) has the group rectangles arranged vertically according to the *same clockwise spiral pattern by which the values were entered into the groups*. Figure 5 illustrates this for four groups labeled a, b, c, and d.

An eight-group map (for N = 5) is formed by placing two four-group maps side by side, except that the right-hand section has group values entered according to a *counterclockwise* spiral. Figure 6 (A) shows the two spiral patterns; Figure 6 (B) gives the entry in numerical order, which the reader can follow to verify the entry pattern.

Figure 7 shows the spiral patterns for the four sections of a 16-group map (for N = 6). Note how the spiral patterns of the various sections are all mirror images around the central axes of the map.

A

0	0	1	1

B

0	1	1	0

Z

0	1	0	1

Figure 2. Illustration of Maps for $Z = A\,\overline{B} + \overline{A}\,B$. A and B are Reference Maps.

Space does not permit a full discussion of the methods for using the map to simplify functions, except to say that any rectangular or square area which is adjacent or reflected around the axes (both major and minor) defines a logic function which can be simplified. A verification of a valid area can be made as follows: M-variables are eliminated from a function by an area of 2^M-squares if the values of the (N-M) variables not eliminated are *constant* in the area. The

0	2	3	1
4	6	7	5

Figure 3. Entry For Two-Group Map (N = 3)

validity test is readily carried out through an inspection of the reference maps, checking the area in question on each map.

Figure 8 illustrates the elimination of two variables from a four-variable function. Note that the area consists of four squares

A

0	0	1	1
0	0	1	1

B

0	1	1	0
0	1	1	0

C

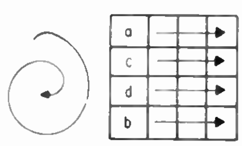
0	0	0	0
1	1	1	1

Z

0	0	1	0
0	1	1	1

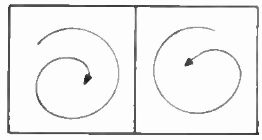
Figure 4. Maps For Z = 00010111

(M = 2, 2^M = 4), is symmetrical around the vertical axis of the map, and the values for two variables (N-M = 2) B,C are constant in the area. Accordingly, the variables A,D are eliminated. A further check as to the validity of the operation is that the *eliminated* variables assume all possible logic values (in this case 00, 01, 10, 11) in the area. The simplified logic expression is then obtained from the reference maps according to the constant values of the uneliminated variables: for the example, this is seen to be $\overline{B} \cdot \overline{C}$.



0	2	3	1
8	10	11	9
12	14	15	13
4	6	7	5

Figure 5. Arrangement of Four-Group Map (N = 4)



PART A

0	2	3	1	17	19	18	16
8	10	11	9	25	27	26	24
12	14	15	13	29	31	30	28
4	6	7	5	21	23	22	20

PART B

Figure 6. Spiral Patterns and Numerical Entry Order For Eight-Group Map (N = 5)

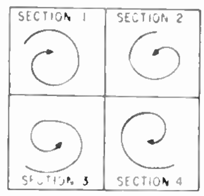


Figure 7. Spiral Patterns for Sixteen-Group Map (N = 6)

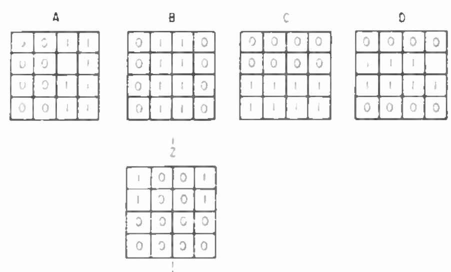


Figure 8. Example of Elimination of Two Variables. $Z = \overline{B} \cdot \overline{C}$ is simplified value.

References

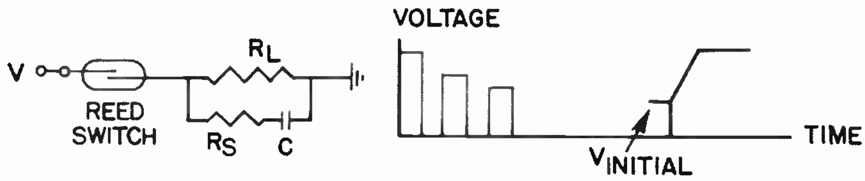
1. E. W. Veitch, "A Chart Method for Simplifying Truth Functions," Proc. Association for Computing Machinery Conference, May 2–3, 1952, p. 127–133.

2. M. Karnaugh, "The Map Method for Synthesis of Combinational Logic Circuits," Communications and Electronics, November 1953, p. 593–599.

3. R. Ledley, "Digital Computer and Control Engineering," McGraw-Hill Book Co., Inc., New York, 1960, p. 321.

IEEE spectrum

ing rise to a constriction resistance that serves to heat the area. This heating effect first causes the metal to melt forming a bridge; then with further lever separation, to boil causing the bridge to rupture by vaporization.



Typical R-C suppression network; voltage contact suppressed.

Another major process which causes material transfer in reed switches is arcing.

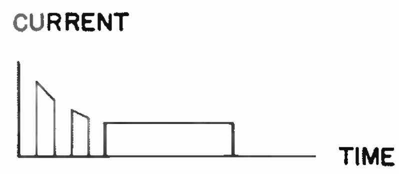
Arcing can occur both at the time the levers are first closing and at the time of the initial lever separation.

The arc on break is essentially the result of thermionic emission. It is caused by the heating of the levers, which in turn causes increasing constriction resistance at the time of lever separation.

The arc on make is caused solely by field emission and does not necessarily entail lever preheating. With open circuit voltages as low as 15 volts an arc can occur.

IBM studies to date indicate that major causes of contact degradation are a function of load conditions. It is a fallacy to rely exclusively on volt-ampere ratings in estimating contact life. For any particular load condition, contact life can be considerably extended through the use of suppressive techniques.

On request, we will send you a detailed report on the design of typical R-C suppression circuits which can maximize the life of your reed switches. Proper suppression techniques create new possibilities for circuit designers where speed, size and power are increasingly important.



Current through A—contact suppressed.

Send in the coupon. IBM Industrial Products Marketing, 1000 Westchester Avenue, White Plains, New York 10604

IBM Industrial Products Marketing Dept. T2 IEEES
1000 Westchester Avenue
White Plains, New York 10604

☐ **IBM reed switch specifications**

☐ **Suppression circuits extend reed switch life**

name ____________________

position ____________________

company ____________________

address ____________________

city ______________ state ______________ zip ______________

IBM
INDUSTRIAL PRODUCTS

IEEE forum

Correspondence relating to activities and policies of the IEEE

Duty to dissent

Your editorial "Duty to Dissent" in the June IEEE SPECTRUM, page 47, is as important to the future of the United States as was Paul Revere's cry of alarm "on the eighteenth of April in seventy-five."

No thoughtful person is likely to deny that in the evolutionary development of American physical environment, it has been the engineering profession more than any other agency that has steadfastly protected the population against exploitation and fraud by "vested interests."

Likewise, no thoughtful professional engineer is likely to deny that since the advent of the TVA and the projection of Government into the water and power utility fields in a massive way, the engineering societies have progressively assumed a greater and greater degree of noncontroversial technical aloofness in their publication and management policies. The result has been the effective muzzling of good, honest debate on matters vital to the continuation of the U.S. way of life in the new electronic–nuclear age that is already upon us.

You are to be congratulated on what you have done. If you now actively pursue the cause you have championed, you will be credited with having led the renaissance of the U.S. engineering profession. You and the profession will both be blessed and the public will have regained its greatest protector.

C. Ken Weidner
Washington, D.C.

You put the matter somewhat stronger than I would, but there is much wisdom in what you say. The difficulty is that to take up the task of dissent effectively, one must be both knowledgeable and boldly self-confident—unfortunately a rare combination. However, let's do what we can!

C. C. Cutler
Editor

New Life Members correspond

I was indeed pleased to receive your letter of October 18 advising me of my transfer to Life Membership, an honor that I deeply appreciate.

Having been retired from my previous position, the project up here is my swan song, after which I shall hang out the "Gone Fishing" sign. It is pleasing to know that I shall then be able to keep in touch through my continued membership.

Congratulations on IEEE SPECTRUM. In my view it is the ideal type of publication for general circulation to members. With the backup of the Group Transactions and other publications, I feel that the field is now well covered.

As you no doubt realize, this is most important to overseas members, most of whom have only infrequent opportunities to participate in other Institute activities.

T. R. W. Bushby
Connellan Airways Pty. Ltd.
Alice Springs, N.T., Australia

Please allow me to send my most thankful acknowledgment for your kind letter of October 18.

Your encouraging thoughts about my completion of 35 years of membership in the IEEE will remain as a source of strength during the rest of my life.

I wish to say, with simple candor, that without the many benefits received as a member of IEEE, especially those providing convenient opportunities to update technical knowledge, and exposure to new and advanced engineering concepts, it would have been very difficult for me to function as a useful engineer, beyond the first 20 years after graduation.

My advice to the young ones starting a career in electrical or electronics engineering is to insure against professional obsolescence by taking full advantage of the splendid opportunities IEEE provides to keep its members abreast of the times.

John J. Rivera
New York, N.Y.

Contributions to this department should be addressed as follows: IEEE Forum, IEEE Spectrum, 345 East 47 Street, New York, N.Y. 10017. It will be assumed that letters so addressed are intended for publication.

An authors' guide to IEEE publications

To assist authors in submitting manuscripts to IEEE publications, we will publish in IEEE SPECTRUM each January a listing such as that which follows. It gives the names of the principal IEEE publications, their frequency of publication, their fields of interest, and the name and address of their editor, to whom manuscripts should be sent.

If you are in doubt about where to send a manuscript, send it to E. K. Gannett, Director, Editorial Services, IEEE, 345 E. 47 St., New York, N.Y. 10017.

It will be helpful to any author planning to submit a manuscript to an IEEE publication to refer to three items of information that have been published in IEEE SPECTRUM. They are: "Information for IEEE authors," pp. 111–115, Aug. 1965; "IEEE Recommended Practice for Units in Published Scientific and Technical Work," pp. 169–173, Mar. 1966; and "A supplement to 'Information for IEEE authors,'" p. 91, May 1966. A reprint of all three is available from IEEE Editorial Department, 345 E. 47 St., New York, N.Y. 10017.

IEEE Spectrum
(Monthly)
Provides the primary link between the IEEE and all of its members, enabling them to keep in contact with technical progress in fields of interest of the IEEE other than their own specialty. Technical articles published are selected for their interest and readability to major numbers of members.
Ronald K. Jurgen, Managing Editor,
IEEE SPECTRUM,
345 E. 47 St., New York, N.Y. 10017

Proceedings of the IEEE
(Monthly)
Provides a link by which workers in the forefront of the many technical fields of the IEEE may communicate with technically active members in their own and other fields at an advanced technical level. The mission is to provide information services in new technical areas, in areas not served by a single Group, or to bring the progress in the field of a Group to a larger segment of the total members than that of the Group.
W. R. Crone, Managing Editor,
PROCEEDINGS OF THE IEEE,
345 E. 47 St., New York, N.Y. 10017

IEEE Student Journal
(Bimonthly starting with January)
Provides technical articles and other material to students in colleges and technical institutes which will give them greater familiarity with the profession they are about to enter, and, in particular, will assist them in choosing and pursuing a career in electrical and electronics engineering.
Editor, IEEE STUDENT JOURNAL,
345 E. 47 St., New York, N.Y. 10017

IEEE Electrolatina
(Quarterly)
A quarterly technical periodical best described as of, by, and for the Latin-American Region of the IEEE, where it is edited and produced. All material appears in Spanish or Portuguese and is selected for its interest to IEEE members in South and Central America, Mexico, and the Caribbean area. These members receive ELECTROLATINA automatically.
Ing. Erik Wallsten,
Monte Libano 875,
Mexico 10, D.F., Mexico

GROUP TRANSACTIONS AND JOURNALS provide the means by which the IEEE member keeps abreast of the advances in a specific field of specialization. Their mission is to record and disseminate new scientific and technical information for both present and future members of the profession.

IEEE Transactions on Aerospace and Electronic Systems (AES) *(Bimonthly starting with January)*
Equipment, procedures, and techniques applicable to the organization, installation, and operation of functional systems designed to meet high performance requirements of earth and space systems.
David Dobson
RCA, Box 588, Burlington, Mass. 01801

IEEE Transactions on Antennas and Propagation (AP)
(Bimonthly starting with January)
Theory and practice in antennas wave propagation, scattering and diffraction, electromagnetic theory, plasmas, and radio astronomy systems.
Dr. John Ruze
M.I.T. Lincoln Laboratory
Lexington, Mass. 02173

IEEE Transactions on Audio and Electroacoustics (AE)
(Quarterly starting with March)
Technology of communication at audio frequencies and of the audio-frequency portion of radio-frequency systems, including the acoustic terminations and room acoustics of such systems, and the recording and reproduction from recordings.
Frederick Van Ween
General Radio Company
West Concord, Mass. 01781

IEEE Transactions on Automatic Control (AC)
(Bimonthly starting with February)
The science and technology that underlies systems, and that can be represented through common mathematical symbolism. Topics of interest include: optimization, adaptation, modeling, components, self-organization, and application to physical systems.
George S. Axelby
Westinghouse Air Arm Division
Friendship International Airport
Baltimore, Md. 21203

IEEE Transactions on Biomedical Engineering (BME)
(Quarterly starting with January)
Theory and application to engineering problems in medicine and biology.
David B. Geselowitz
Moore School of Electrical Engineering
University of Pennsylvania
Philadelphia, Pa. 19104

IEEE Transactions on Broadcast and Television Receivers (BTR) *(Aperiodical)*
The design and manufacture of broadcast and television receivers and components, and activities related thereto.
Chester W. Sall
RCA Laboratories
Princeton, N.J. 08540

IEEE Transactions on Broadcasting (BC)
(Quarterly starting with January)
Broadcast transmission systems engineering including the design and utilization of broadcast equipment.
Serge Bergen
10828 Fairchester Drive
Fairfax, Va. 22030

IEEE Transactions on Circuit Theory (CT)
(Quarterly starting with March)
Design and theory of operation of circuits for use in radio and electronic equipment.
Prof. Ben Leon
Electrical Engineering Dept.
Purdue University, Lafayette, Ind.

IEEE Transactions on Communication Technology (COM)
(Bimonthly starting with February)
Technology of telecommunications (theory and application) including telephony, telegraphy, and the transmission of other data, also point-to-point radio and television; treatment of systems and subsystems using one or a combination of means of transmission.
R. D. Slayton
Teletype Corp., 5555 Touhy Ave.
Skokie, Ill. 60076

IEEE Transactions on Computers (CPR)
(Monthly starting with January)
Theory, design, and practices relating to digital and analog computation and information processing.
Prof. Harry D. Huskey
Computer Center, University of California
Santa Cruz, Calif. 95060

IEEE Transactions on Education (E)
(Quarterly starting with March)
To foster improved relations between the electrical and electronics and affiliated industries and schools, colleges, and universities.
Prof. Sidney S. Shamis
Dept. of Electrical Engineering
New York University
University Heights, Bronx, N.Y. 10453

IEEE Transactions on Electrical Insulation (EI)
(Quarterly starting with February)
Electrical insulation common to the design and construction of components and equipment for use in electric and electronic circuits and distribution systems at all frequencies.
Charles L. Petze, Jr.
Delaware Research & Development Corp.
P.O. Box 110, Newcastle, Del. 19720

IEEE Transactions on Electromagnetic Compatibility (EMC) *(Quarterly starting with March)*
Origin, effect, control, and measurement of radio-frequency interference.
A. H. Sullivan, Jr.
7121 Wolftree Lane, Rockville, Md. 20852

IEEE Transactions on Electron Devices (ED) *(Monthly)*
Electron devices, including particularly electron tubes, solid-state devices, integrated electronic devices, and energy sources.
Dr. Glen Wade
Dept. of Electrical Engineering
University of California
Santa Barbara, Calif. 93106

IEEE Transactions on Engineering Management (EM)
(Quarterly starting with March)
Engineering management and administration as applied to technical, industrial, and educational activities in the field of electrical and electronics engineering.
Prof. A. H. Rubinstein
Dept. of Industrial Engineering
The Technological Institute
Northwestern University, Evanston, Ill. 60201

IEEE Transactions on Engineering Writing and Speech (EWS) *(Aperiodical)*
The study, development, improvement, and promotion of the techniques for preparing, organizing for use, processing, editing, collecting, conserving, and disseminating any form of information in the electrical and electronics fields.
L. M. Cole, Bell Telephone Labs., Inc.
Room 1E 238, Murray Hill, N.J. 07971

IEEE Transactions on Geoscience Electronics (GE)
(Quarterly starting with February)
Research, development, and techniques in instrumentation for geophysics and geochemistry, especially gravity and seismic measurements, magnetics, well logging, space exploration, meteorology, oceanography, and aerology.
Dr. A. A. J. Hoffman
Computer Center, Texas Christian University
Fort Worth, Tex. 76129

IEEE Transactions on Industrial Electronics and Control Instrumentation (IECI) *(Aperiodical)*
The application of electronics and electrical sciences to industrial processes.
R. B. Spooner
IMPAC Instrument Service Inc.
201 East Carson St., Pittsburgh, Pa. 15219

IEEE Transactions on Industry and General Applications (IGA)
(Bimonthly starting with January/February)
Application of electricity in industry, transportation, commerce, and the home.
D. S. Brereton
General Electric Co.
One River Road, Schenectady, N.Y. 12305

IEEE Transactions on Information Theory (IT)
(Quarterly starting with January)
Information theory and its application in radio circuitry and systems.
Carl W. Helstrom
Applied Electrophysics Dept.
University of California, Box 109
LaJolla, Calif. 92037

IEEE Transactions on Instrumentation and Measurement (IM) *(Quarterly starting with March)*
Measurements and instrumentation utilizing electrical and electronic techniques.
Prof. George B. Hoadley
Dept. of Electrical Engineering
North Carolina State College
Raleigh, N.C. 27607

IEEE Transactions on Magnetics (MAG)
(Quarterly starting with March)
Magnetic phenomena, materials, and devices as applied to electrical and electronics engineering.
E. W. Pugh
International Business Machines Corp.
Yorktown Heights, N.Y. 10398

IEEE Transactions on Man–Machine Systems (MMS)
(Quarterly starting with March)
Development and application of electrical and electronic knowledge germane to integrating human elements into man–machine systems.
Thomas B. Sheridan
Massachusetts Institute of Technology
Cambridge, Mass. 02139

IEEE Transactions on Microwave Theory and Techniques (MTT) *(Monthly)*
Microwave theory, microwave circuitry and techniques, microwave measurements, and the generation and amplification of microwaves.
S. Okwit
Airborne Instruments Laboratory
Walt Whitman Road, Melville, N.Y. 11746

IEEE Transactions on Nuclear Science (NS)
(Bimonthly starting with February)
Nuclear science and engineering instrumentation, plasma and high-energy physics, reactor controls, and radiation effects.
Richard F. Shea
14 Coronet Court, Schenectady, N.Y. 12309

IEEE Transactions on Parts, Materials and Packaging (PMP) *(Quarterly starting with March)*
Consideration of physical properties, performance characteristics and procedures, and techniques pertaining to component parts, materials, and circuit packaging and manufacturing.
Gustave Shapiro
Engineering Electronics Section
National Bureau of Standards
Washington, D.C. 20234

IEEE Transactions on Power Apparatus and Systems (PAS)
(Monthly)
Planning, research, development, design, application, construction, installation, and operation of apparatus, equipment, structures, materials, and systems for the safe, reliable, and economic generation, transmission, distribution, conversion, and control of electric energy for general industrial, commercial, public, and domestic consumption.
Prof. R. T. Smith
Dept. of Electrical Engineering
University of Oklahoma, Norman, Okla.

IEEE Journal of Quantum Electronics (QE) (Sponsored by ED and MTT)
(Monthly)
Physics of quantum electronic devices, such as masers and lasers, and associated techniques including modulation, detection, and nonlinear optic effects.
Dr. R. H. Kingston, M.I.T. Lincoln Lab.
Lexington, Mass. 02173

IEEE Transactions on Reliability (R)
(Quarterly starting with March)
Principles and practices used in reliability of electric and electronic equipment.
John A. Connor
TRW-Space Technology Laboratories
Bldg. R-4, Room 2198, One Space Park
Redondo Beach, Calif. 90278

IEEE Journal of Solid-State Circuits (SC)
(Quarterly starting with March)
Theory and design of solid-state circuits, and techniques, processes, and system considerations affecting their applications and performance.
Dr. James Meindl
Dept. of Electrical Engineering
Stanford University, Stanford, Calif.

IEEE Transactions on Sonics and Ultrasonics (SU)
(Quarterly starting with April)
Sonics, including ultrasonics and phonon technology.
Oskar E. Mattiat
Raytheon Submarine Signal Division
P.O. Box 360, Newport, R.I. 02840

IEEE Transactions on Systems Science and Cybernetics (SSC) *(Aperiodical)*
Science common to large collections of interacting functional units that together achieve a defined purpose. Areas include interdisciplinary subjects such as bionics, artificial intelligence, and self-organizing systems; and such aspects as modeling, optimization, reliability, and general theory.
John N. Warfield
Battelle Memorial Institute, Columbus, Ohio

IEEE Transactions on Vehicular Technology (VT)
(Aperiodical)
Vehicular technologies, which include communications and automotive electrical and electronics engineering. Communications include land, airborne, and maritime, as well as personal paging and citizen's communications services, when used as an adjunct to a vehicular system. Automotive includes all functional systems related to the vehicle and its operating medium, excluding systems associated with public transit.
Harry Nylund
Bell Telephone Laboratories, Inc.
Room 3F 630, Holmdel, N.J. 07733

Spectral lines

A new information service. The IEEE is participating with the IEE (Institution of Electrical Engineers) of London, England, in joint publication of *Electrical and Electronics Abstracts* (*EEA*), a widely used bibliographic reference or abstracts journal that has been published by the IEE under one name or another since 1898. Its other current name is *Science Abstracts, Series B*.*

In recent years *EEA* has grown significantly, and in 1968 it will provide reference to about 30 000 published items from the periodical and conference literature, making it the most comprehensive abstract service covering the IEEE's field of interest. A special arrangement has been made with the IEE so that an IEEE member may enter a personal subscription for *EEA* at the special rate of $21 for the year 1968, through the IEEE in New York. We hope that many members will take advantage of this offer.

Why should the IEEE join in publishing a relatively conventional abstracts journal in an era of computer-based information systems? The IEEE is very interested in the future development of such information systems but recognizes that a good abstracts journal provides the best base from which it is natural to develop a useful computer-based awareness and retrieval system. We intend to carry out that natural development, in cooperation with the IEE and other organizations active in the publication of technical and scientific information. To us, the key issue is the indexing method. The indexing for a published item is inherently the address part of a message; the text of the message is the abstract (or the document) that the user of the index is led to. An information service is thus a communication channel from the author to a potential reader with plenty of noise, interference, and randomness on the channel. Inadequacies in the indexing are like noise in the address part of the message transmission—the text may never reach the intended reader.

To follow the analogy a bit further, the meaningful content of the indexing (the address) should be the same, regardless of the channel used. The coding of the address may differ for diverse forms of bibliographic material—reference book, journal, card file, punched card deck, or computer tape—but the semantic content of the address should not, for it is in each case an identifier of that same set of individuals who should read the text of the message.

When you attempt to push the analogy further, it fights back. Indexing has substantial intellectual content, in a way not typical of ordinary communications addressing. That intellectual content—the succinct identification of an area of knowledge treated by the text—may have been put there by the indexer, by the author, or through computer manipulation of the text of the item or abstract. In any event, the quality of the index depends on the intellectual content.

Unfortunately, intellect and computer time are both expensive and in short supply. Therefore, it is highly desirable that all of the indexing required for a given published item, for the different possible communication channels, be done as part of one coherent process. That is our goal.

One big factor in making the single-input-event approach feasible now is that substantial progress has been made in computer-controlled composition methods. The IEEE will increasingly be using computer-controlled systems for composing its own indexes. A computer master-file record on magnetic tape will be prepared for each bibliographic item. The master file is our permanent index record of that literature and can be used to prepare not only the year-end indexes for each IEEE journal, but also a combined IEEE index, cumulative indexes (eventually), and magnetic tape products for sale to operators of computer-based retrieval systems. When the indexing is done at the time of publication, as it will be in 1968, the same indexing will also serve for *EEA* or other abstracts journals.

The staffs of the IEE in London and the IEEE in New York were found to be planning along almost identical paths. Since the two societies are essentially identical in the scope of their subject matter interests and share a common (or almost common) language, the logic of proceeding jointly toward those common goals became inescapable.

In 1968, *Electrical and Electronics Abstracts* will be analyzed, worked on, expanded, revised, and partially computerized, with staff on both sides of the Atlantic participating. If we work at it effectively, and have the benefit of enough advice and support from both members and nonmembers, both societies should have by 1969 a spectrum of bibliographic products, from journal indexes to computer services, that will serve a diverse set of users better than they have ever been served before.

We hope that in offering *Electrical and Electronics Abstracts* to the IEEE members we are starting in the development of a series of new information services for the membership. We feel that the IEEE has the responsibility to progress rapidly in that direction.

F. Karl Willenbrock

**Series A* is *Physics Abstracts; Series C* is *Control Abstracts.*

Authors

Negative conductance in semiconductors (page 47)

Herbert Kroemer (SM), head of the New Phenomena Section at the Fairchild Semiconductor R & D Laboratories, Palo Alto, Calif., received the Ph.D. degree in physics from the University of Göttingen, Germany, in 1952. He was subsequently associated with the Telecommunication Research Laboratory of the German Postal Service, RCA Laboratories at Princeton, N.J., the German Philips Research Laboratories, and Varian Associates at Palo Alto, Calif.

Dr. Kroemer has been involved with various aspects of solid-state physics technology and devices. Two recurrent themes of his work have been the search for device mechanisms operating at increasingly high frequencies and, as a consequence, an interest in the bulk effects of semiconductors. Studies of the Gunn effect, which began shortly after its discovery, currently represent a significant portion of his research effort.

Computers and computing—past, present, and future (page 57)

J. Nievergelt (M) is an assistant professor in the Department of Computer Science, University of Illinois. He received the Diploma in Mathematics from the Swiss Federal Institute of Technology, Zurich, in 1962, and the Ph.D. degree in mathematics from the University of Illinois in 1965. His main research interests are automata and formal-language theories and their application to problems of computer programming. He is currently involved in the design and implementation of a symbol manipulation language called EOL. His side interests include numerical analysis, game theory, and artificial intelligence.

Dr. Nievergelt is a member of the Association for Computing Machinery, the Society for Industrial and Applied Mathematics, Sigma Xi, and Phi Kappa Phi.

The high-brightness LED (page 62)

D. K. Hillman (*left*) is a project leader in Monsanto Company's newly formed New Enterprise Division, in St. Louis, Mo., where he is working on visible injection electroluminescent $GaAs_{1-x}P_x$ devices. He has been with Monsanto's Central Research Department since 1964. During this period he has worked on the development of incoherent visible and infrared emitters. His previous experience was with Tektronix, Inc., where he worked on GaAs diffused switching diodes. He received the B.S. degree in physics from the University of Nebraska in 1961.

G. E. Smith joined the Instrumentation Research Department of Fairchild Semiconductor in August 1967, following a three-year period with Monsanto, during which time he worked on integrated circuits, semiconductor lasers and display devices, and light-emitting diodes. He received the B.S. degree in physics and mathematics in 1958 from the University of New Zealand, where he later pursued a year of graduate study in nuclear physics and then joined the university's lecturing staff. In 1961 he joined the instrument engineering staff at Tektronix, Inc., Beaverton, Oreg.

A power supply for pulsed power welding (page 67)

E. H. Daggett (M) is a project manager on the staff of the Engineering and Equipment Development Department of AIRCO Welding Products Division of Air Reduction Company, Inc., Union, N.J.

He received the bachelor of science degree from the University of New Hampshire in 1948 and subsequently joined the Monsanto Chemical Company and later, Bell Telephone Laboratories, Inc. His prior work with the Air Reduction Company was at its Central Research Laboratories, where he developed electric welding equipment. He holds a U.S. patent on a power supply.

Mr. Daggett is an associate member of the American Welding Society and a registered professional engineer in the State of New Jersey.

Galvanomagnetic devices (page 75)

H. Weiss was born in 1920 in Nuremberg, Germany. He studied physics at the University of Erlangen following World War II. After finishing these studies he served as an assistant at the university, where he completed his thesis, on the photoconductivity of zinc oxide, in 1951. In 1952 he joined the Research Laboratories of the Siemens Schuckertwerke in Erlangen, where he worked with a group led by H. Welker, who discovered the semiconducting III-V compounds. His efforts were concentrated chiefly in the field of transport properties, especially the galvanomagnetic effects, of semiconductors—both at the research and application levels. In 1965 he received the physics prize of the Academy of Science in Göttingen for these investigations, which led to the discovery of the two-phase InSb–NiSb. At the time this article was written Dr. Weiss was manager of the Siemens AG Research Laboratories in Erlangen. He has since transferred to the Semiconductor Department of Siemens AG in Munich.

Education for innovation (page 83)

Daniel V. De Simone (M) received the B.S. degree in electrical engineering from the University of Illinois and the doctorate in law from New York University. Prior to entering government service in 1962 he was associated for six years with Bell Telephone Laboratories, Inc.

Dr. De Simone is currently director of the Office of Invention and Innovation at the National Bureau of Standards and in this capacity is responsible for programs to encourage the development and transfer of technology. He is also executive director of the National Inventors Council, which advises the Secretary of Commerce and other agency heads on issues concerning inventors and inventions. He performs still another role as the U.S. Commerce Department's program manager for technology, reporting to the Assistant Secretary for Science and Technology. In this capacity, he is responsible for programming, planning, and budgeting for the technologically oriented agencies of the U.S. Department of Commerce: the Patent Office, the Office of State Technical Services, and the NBS Institute for Applied Technology.

Digital control of shaft speed and position (page 90)

R. A. Millar (M) is a senior engineer in the Research and Development Section of Huntec Limited, Toronto, Canada. He is actively engaged in the design and development of electronic instruments, incorporating analog and digital techniques, for use in geophysical surveying and exploration.

He received the Higher National Certificate in Electrical Engineering from Acton Technical College in London, England, in 1957. In England he worked for Panellit Ltd. as a systems engineer on the design of a large-scale data-reduction scheme for a power station. Following his emigration to Canada he was employed as project engineer by Canadian Applied Research Ltd. and later by Ferranti-Packard Ltd., where he worked on an information retrieval system using microfilm techniques and digital logic. He was also involved in the design of a modular magnetic information display system.

Message from the President

S. W. Herwald *President IEEE*

When one tries to consider 160 000 individuals—most of them individualists—grouped into a voluntary organization, busily writing, organizing, and meeting on a part-time basis, it's easy to lose sight of the basic purposes of the IEEE. As a corollary to that, I would add that it is also easy to lose sight of the significance of the fact that IEEE membership is voluntary.

"Why am I a member?" "Why should I continue to be a member?" These are questions that every active IEEE member, from the President to the newest Student Member, must ask himself periodically. And, certainly, if he is an active participant in IEEE activities and functions, the individual will find the dominant reasons for his continuing association shifting from time to time. As each of us gets older, as our interests and perspectives change, we discover that there is a continuing but changing value for us in the Institute. My own interests have shifted at least five times over the past 20 years.

Most recently, my main concern has been to enhance communications in TAB, our Technical Activities Board. But this concern has underlined for me, in conjunction with the careful self-scrutiny made by the IEEE Executive Committee during most of 1966, what I believe is a fundamental and unchanging mission of our organization.

Basically, we are an organization of people interested in electrical and electronics engineering and its various interfaces with all worldwide activities. Thus, our underlying purpose in getting together is to enhance communication about our field. Our goal is to advance the technology that we all, in one way or another, depend upon. This enhanced communication is probably the single most important "product" of our Institute. Through IEEE, it is easier for a member to reach those whom he needs to reach in order to carry on the dialogues that are so essential for his professional work.

From this kind of viewpoint, it is easier to understand the whole army of activities that IEEE undertakes. We try to advance on many fronts: with the public, with the rest of the technical community, and, most important, among ourselves in our profession.

In the forefront of communication enhancement, of course, are our publications: IEEE SPECTRUM, the PROCEEDINGS OF THE IEEE, the TRANSACTIONS, and the JOURNALS.

Of great importance, too, are the Institute-sponsored technical meetings. Exclusive of our big New York International Convention, these meetings are a million-dollar-a-year proposition. In 1966, we sponsored 97 meetings.

But publications and meetings are just two of the more obvious foreground manifestations of how we enhance communications through IEEE. Behind these are the organizational elements—groups of volunteers and permanent staff—that keep the information exchange moving, to make sure that our tremendous activity answers the needs of as many members as possible.

I obviously cannot detail to you how all the different parts of this Institute—from the Board of Directors to the "grass roots" organizations of Sections and Groups—all interact to insure that our policies and activities match our members' real needs. Instead, I shall look at just one of our activities that bears on how we will enhance our communications function in the future.

This activity has to do with the information revolution, and our approach to it through our Information Services. By incorporating the best of what is now known about information retrieval, we have started a "kernel" system that can grow and evolve in the ever-shifting real-world scene. By relying on IEEE authors and editors, we should get the best possible indexing of our own literature, which constitutes a good proportion of that published in the world. This indexing, going into a master file on computer tapes, will serve as a kernel multiuse file. At first, it will be used for things like computer-controlled printout of indexed material (not yet a retrieval system). In 1968, this file will be extended to include also all IEEE-sponsored conferences. In due course, this master file will form the core of an information retrieval system as is appropriate to IEEE's unique needs. We are also joining with the Institution of Electrical Engineers of London in the publication of a comprehensive worldwide abstract of *all* professional literature pertaining to electrical and electronics engineering. This international venture offers both economic advantages and a better coverage of the entire field. These are but two aspects of what I mean by our enhanced communications.

The real backbone of all our efforts in IEEE, however, is voluntary servitude by those members who sometimes donate up to 50 percent of their time to work on society activities. And thank Heaven they do! No one could ever afford to pay for all that activity.

Fortunately, by and large, members' interests in the society are deep enough, and broad enough, that we usually find volunteers for nearly any job that needs to be done. In fact, one of the inherent values of the Institute is that we have room for individuals with almost any interest—from the person who wants to tell the world about his high Q values in single inductive transistors, to the generalist who's just monitoring the field.

Yet, it is my feeling that there are many individuals in IEEE who have not sufficiently appreciated that it is a voluntary organization and that, as such, communications within it go along a two-way street. To get more out of the enhanced communications that are made available to him, the individual must put more in, through his own deeper participation, through his membership in Groups, and through his personal letters, which receive much more attention than individual members probably imagine. In sum, what I am saying is that if you feel you are not getting enough out of your membership in IEEE, it may well be that you are not using the resources of the Institute that really are available. It is up to you to seize the value of enhanced communications that our Institute provides.

Negative conductance in semiconductors

Although progress in semiconductor research has proceeded at an astonishing pace during the last two decades, two recent discoveries are on the verge of revolutionizing the fields of microwaves and solid-state physics

Herbert Kroemer *Fairchild Semiconductor*

Until recently, investigators have been frustrated in their attempts at applying microwave and millimeter-wave frequencies to semiconductor devices. During the last few years, the discovery of avalanche transit-time and Gunn effects in bulk semiconductors has been met with overwhelming enthusiasm. The successful fabrication of models presently utilizing these negative-conductance phenomena has given these high-frequency devices an optimistic outlook for the future.

During the 19 years since the discovery of the transistor, the application of higher and higher operating frequencies has been one of the most persistent and frustrating objectives of semiconductor device research. Unfortunately, progress toward this elusive goal has been exceedingly slow, no matter how relentless the undertaking. The last few years, however, have introduced two discoveries that completely alter the complexity of this problem. It is possible for today's investigator to predict accurately semiconductor devices for the future that will not only operate over the entire conventional microwave frequency range, but the millimeter-wave frequency range as well.[1] The two discoveries referred to are embodied in the following events: (1) the realization of avalanche transit-time devices in silicon[2] (along the lines first proposed by Read[3] in the 50s), and (2) the discovery[4–6] and utilization[1] of the Gunn effect in gallium arsenide. It is the intention of this article to describe these two phenomena and their negative-conductance properties and applications in semiconductors.

Although they differ from each other in many respects, avalanche transit-time and the Gunn effect do have several aspects in common. For one, neither of them is related to the mechanisms of transistors. For another, both effects depend on the properties of "hot" electrons; that is, on the properties of electrons whose energy is large compared with kT, perhaps of the order of several tenths of an electronvolt or more. Finally, both phenomena involve the use of hot electrons in a two-terminal negative conductance of one form or another—the reason for the title of this article. The individual nature of negative conductance for these two cases is entirely separate, however. In avalanche transit-time devices, the negative conductance is caused by a phase shift (exceeding 90°, and ideally near 180°) between current and voltage. In devices that are related to the Gunn effect, the negative conductance is caused by a local negative conductivity within GaAs in strong electric fields; that is, at every point inside a GaAs crystal of such a device, the local current density decreases whenever the local electric field increases beyond a certain threshold.

The approach of any article attempting to deal with negative conductance in semiconductors should be twofold. First, it should describe and perhaps explain the physical origins of the two known effects causing negative conductance. And second, it should describe the manner in which these negative conductance effects can lead to the experimentally observed device behavior. This second aspect is particularly important in devices that are related to the Gunn effect, since, as we shall observe, a local negative conductance in these devices is capable of an almost incredible variety of external appearances. Consequently, a greater proportion of this article places emphasis on the Gunn-effect type rather than on the avalanche transit-time type of negative conductance. Furthermore, we shall concentrate on the fairly well-established basic concepts as they are presently understood, rather than on the large amount of experimental data and theoretical details still in a state of rapid transition—restricting any comments concerning technology to a minimum.

This emphasis on basic concepts implies that the material will not be presented in its chronologic sequence of discovery. The Gunn effect, in fact, notoriously exemplifies the case where the sequence of events leading to a discovery did not at all follow the sequence in which one might want to order and present such material conceptually, once the discoveries had taken place. Those who are not familiar with the discovery chronology of the Gunn effect should find this conceptual presentation of the facts easier to understand. Those who are already familiar with the experimental details and the order in which they were uncovered may at least gain a new point of view—hopefully, discovering that there is more order in this phenomenon than they might have suspected.

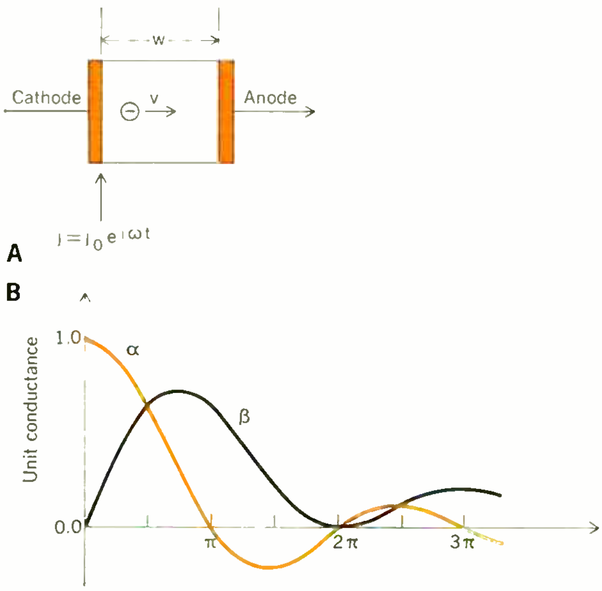
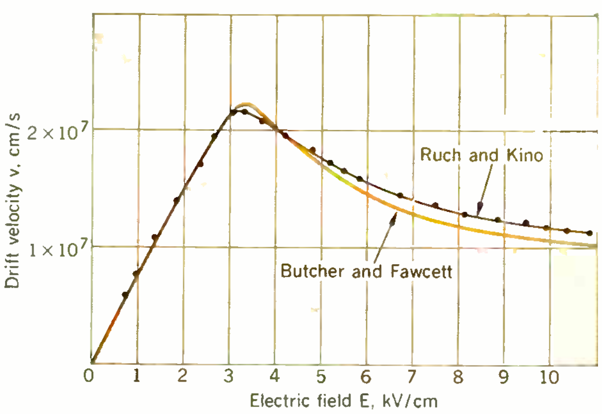


FIGURE 1. Negative conductance resulting from the combined phase shift of avalanche buildup and transit delay in silicon. A— p-n junction. B—Phase-shift characteristic.

FIGURE 2. Experimental (Ruch and Kino) and theoretical (Butcher and Fawcett) velocity-field characteristics of gallium arsenide.



Phase-shift negative conductivity in silicon

Avalanche transit time. As previously stated, the negative conductance of avalanche transit-time devices is attributed to a phase shift between the current and voltage of p-n or p-i-n junctions that are biased into the avalanche breakdown range. This phase shift consists of two components. One of these is a phase delay caused by the finite transit time of electrons through the space-charge layer of the junction. The other is a phase delay caused by the avalanche multiplication process itself, resulting from the condition that the rate at which electron-hole pairs are generated inside the avalanching region of a reverse-biased p-n junction is proportional to the density of electron-hole pairs that are already present in that region. For fields sufficiently far into the avalanche breakdown range, this rate of generation will exceed the rate at which electron-hole pairs can leave the avalanche region; and, as a result, both density and current will exponentially grow with time. What ultimately limits the current is its own space charge, which weakens the field inside the avalanching region of the p-n junction to exactly the value necessary to sustain the avalanche without further growth.

Obviously, this space-charge buildup takes time. If one should now superimpose a small alternating voltage of sufficiently high frequency over such a sustained avalanche, the space-charge readjustment would not be able to follow that voltage, and one would obtain a high-frequency ac component of the avalanche current that would not be space-charge limited. This current will continue to build up during every cycle of the alternating voltage, even after that voltage has gone through its maximum; ideally, so long as the voltage is positive. However, this means that the alternating current will go through its maximum at that instant at which the alternating voltage declines and goes through zero. In other words, the avalanche current exhibits a 90° inductive phase delay relative to the driving alternating voltage. Any additional phase delay can then lead to a negative conductance.

This (ideally) 90° avalanche phase delay combines with the phase delay caused by the finite transit time of electrons through the space-charge region of the p-n junction in the way described by Fig. 1. If one assumes that the periodic ac density $j_0e^{i\omega t}$ is only generated within a very thin layer at the cathode end of the space-charge region in this p-n junction,* and that the electrons travel with a uniform velocity v through this space-charge region, then one must sum over all the current contributions inside the p-n junction (generated at different times) in order to compute the total current that flows through the external leads. Calculations show that this current is given by

$$J(t) = (\alpha - i\beta)j_0e^{i\omega t}$$

where $\alpha = \dfrac{\sin \omega\tau}{\omega\tau}$ $\beta = \dfrac{1 - \cos \omega\tau}{\omega\tau}$

* The assumption of only a very thin avalanche region all the way on the cathode side tremendously simplifies the analysis, without sacrificing any of the essential concepts. If the avalanche region is located at an intermediate position, and/or if it occupies a significant fraction of the space-charge layer thickness, both the mathematical analysis and—to some extent—the frequency behavior of the device, become more complicated, and do not introduce any newer physical concepts. The reader interested in these problems is directed to Ref. 1.

Here, $\omega = 2\pi f$; and τ is the transit time through the space-charge region. The two quantities α and β are shown in the curves depicted in Fig. 1. Both curves decay in an oscillatory fashion with increasing frequency, and although α oscillates between positive and negative values, β is always positive or zero.

If one now assumes that the original local current density that is generated at the cathode interface is a purely inductive current created by a strong avalanche multiplication, or

$$j_0 = \frac{E_0}{i\omega\lambda}$$

where λ is some proportionality factor that might be called an inductivity, then it is found that the overall conductance G of this device (the real part of the admittance) is proportional to $-\beta/\omega\lambda$. The idealized conductance, therefore, is either negative or zero for all frequencies, with the largest negative value occurring for a frequency approximately equal to one half of the inverse transit time, or a transit angle of 180°. Its value is zero for the inverse transit-time frequency, and for all integer multiples of this frequency. It is this negative conductance brought about by the β term displayed in Fig. 1 that represents the underlying mechanism of avalanche transit-time devices.

In our idealized example, this negative conductance exists essentially for all frequencies. In practice, however, one is restricted to operate at a frequency close to the reciprocal transit time. Three reasons account for this. First, negative conductance attains its maximum at this point. Second, this is where the α term, which represents either an inductance or a capacitance, depending on its sign, goes through zero. Finally, our simplifying assumption that there exists a purely inductive current density j_0 is never entirely correct. This is particularly true at low avalanche multiplications or at low frequencies where space-charge effects become dominant; in these cases, the avalanche phase delay is always somewhat less than 90°. In this case, the current density j_0 is not purely inductive as it flows through the cathode interface, but contains a conductive contribution, without a phase shift, and this conductive portion of the internal current density contributes to the overall conductance in proportion to the α term of Fig. 1.

This α term is, of course, always positive for low frequencies and, for sufficiently low frequencies, even a weak α term is much larger than the β term. It is, therefore, not possible to obtain an overall negative conductance with transit angles less than 10° to 30°, and negative conductance has usually already disappeared for much larger transit angles.

For transit angles larger than 180° and approaching 270°, the negative conductance of the device rapidly decreases and the capacitance increases. Therefore, for all practical purposes, one must consider that avalanche transit-time devices work well only in a frequency range around a 180° transit angle, although this range may be rather wide—typically, as wide as 2:1.

Device aspects. Having sufficiently described the physical mechanism of these devices, an evaluation of their merits, in comparison with such other devices as those based on the Gunn effect, must consider three primary aspects. One is available technology, the second is noise behavior, and the third is the relation of frequency to transit time, and therefore to the thickness of the device. Of these three, the existence of a sophisticated technology is the strongest asset of impact avalanche transit-time devices—an enormous advantage indeed. Since these devices can be fabricated from silicon, using essentially the same materials and device technology that are used for transistors and integrated circuits, the problems encountered in developing new materials or in formulating a new technology are virtually nonexistent. I believe this advantage to be so important that, in all applications where these devices can provide the required performance, their usage will be given priority.

Drawbacks, however, are encountered in the other two device considerations. Unfortunately, the avalanche multiplication process is a relatively noisy one[7,8]—similar to the noise in a photomultiplier—and it does not, at the present time, appear likely that these devices will ever incorporate low-noise characteristics. For those applications where minimum noise is a necessity, and there are many, one must look for some other class of devices, such as those utilizing the Gunn effect, which offer substantially less noise.

The second drawback, involving the relationship of frequency to device thickness, is a more indirect one. The implication is that, for a given cross-sectional area of a device, the capacitance increases in proprotion to the frequency, and the impedance decreases in inverse proportion to the square of the frequency. Since, in practice, one always has to work at some reasonable impedance level (whatever that level may be), an increase in frequency must also be followed by a reduction in area of a proportion equal to the inverse square of the frequency. This, of course, leads to a power falloff of 6 dB per frequency octave. In essence, this is only an ultimate limitation, since the situation is presently fairly academic and does not become an important consideration until one reaches millimeter-wave frequencies (unless one is interested in very high pulsed powers). Furthermore, because the high state of development for existing materials and technology is presently available in producing these devices, the realization of these frequencies is possible right now rather than, say, in five years. Indeed, Bowman and Burrus at Bell Telephone Laboratories, Inc., have achieved oscillations at frequencies ranging to 340 GHz, with power levels at 300 GHz still in the milliwatt range.[9] It is quite likely, however, because these investigations were based on an already highly developed technology, that the power levels obtained are already fairly close to the ultimate limits.

Local negative conductivity in gallium arsenide

The Gunn effect. Since the current density in a semiconductor is proportional to both the density of electrons and their drift velocity, a decrease in current density—with an increasing electric field—can be brought about by a decrease in either of these quantities. Both types of decrease have actually been observed. For example, a decrease of electron density, in the presence of an increasing field, can be brought about by field-enhanced trapping. This occurs in gold-doped germanium at cryogenic temperatures, and in high-resistivity GaAs at room temperatures. However, these field-enhanced trapping effects are, in general, very slow, and the device potential of this form of negative conductivity is, at best, very limited. Certainly, the microwave frequency range is beyond this potential.

The second possibility for reducing current density in an increasing field, that of decreasing the drift velocity, takes place in GaAs above a 3-kV/cm field strength. It is this mechanism which underlies the Gunn effect, although that was not recognized at the time the effect was discovered. This decrease in drift velocity, in turn, is a result of the conduction-band structure of GaAs. That the GaAs band structure could lead to a decrease in drift velocity with increasing field, i.e., a negative mobility, had already been recognized by Ridley and Watkins[10] and by Hilsum[11] before the discovery of the Gunn effect. Particularly, Hilsum's paper was quite specific and quantitative about this possibility. However, these predictions were promptly ignored, and it was not until after the discovery of the Gunn effect that they were taken seriously and invoked as a means of explaining the Gunn effect.[12]

Even then, direct measurement of the dependence of drift velocity on the electric field, and direct evidence for the existence of the negative differential mobility were not available. Measurements of this kind have finally been published during the last year by several groups working independently on this problem.[13-17] Figure 2 gives an example of perhaps the most beautiful and certainly the most direct result—that of Ruch and Kino[16] at Stanford University. Their experimental curve speaks for itself, and we merely wish to add that other evidence obtained from Gunn-effect models[6,18] has shown that the drift velocity maintains a value approximately equal to Ruch and Kino's minimum value in fields up to the order precipitating avalanche breakdown (about 200–300 kV/cm). Moreover, Fig. 2 shows a theoretical curve by Butcher and Fawcett[19] that represents a continuation and extension of the work Hilsum performed in 1961. The close agreement between both curves indicates the accuracy level theoretical understanding of this effect has attained.

As previously noted, negative-mobility behavior is a result of the conduction-band structure of GaAs (Fig. 3). GaAs is a direct-gap semiconductor with a conduction-band minimum occurring at $k = 0$, the center of the Brillouin zone. It is this fact that accounts for the low-minority carrier lifetimes that make GaAs a poor transistor material, as well as for the high-radiation recombination probability that makes it an excellent laser material. This property is also an essential ingredient of the Gunn effect. It is typical of direct-gap semiconductors that the conduction-band effective mass is relatively low, and GaAs is no exception to this rule. Its effective mass is of the order of 7 percent of the free-electron mass m_0.

In addition to this central valley, GaAs possesses what we shall call satellite valleys; that is, additional conduction-band minima with higher effective masses at higher energies—energies that are large compared with kT. This, too, is a very common feature of the type of semiconductor under discussion. In the case of GaAs, the satellite valley energy is about 0.36 eV. These satellite valleys are located along the ⟨100⟩ directions of k space (probably on the surface of the Brillouin zone), and the combined density-of-states mass for all satellite valleys is of the same order as the free-electron mass itself.[20]

At room temperature, and in the absence of any strong electric fields, all electrons will occupy the central valley, since the energy separation between it and the satellite valleys is large compared with kT. Because of the very low effective mass in the central valley, electrons have the high mobility that is characteristic of these semiconductors; but, for the same reason, electrons are also very easily heated by an external electric field. When this heating does take place, an increasing fraction of electrons reaches energies approaching the magnitude of the

FIGURE 3. Conduction-band structure of GaAs.

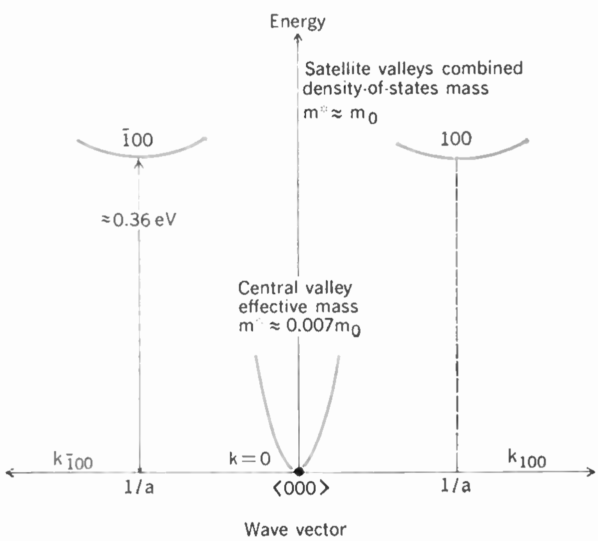
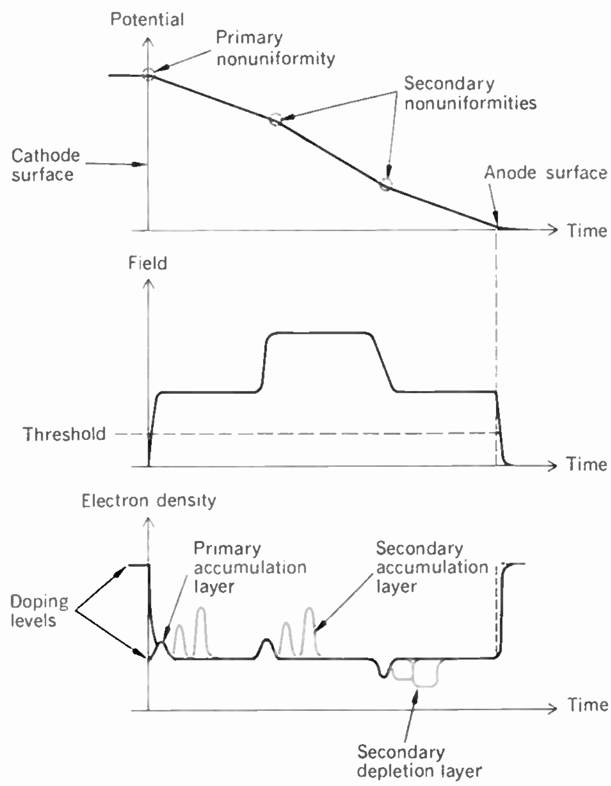


FIGURE 4. Traveling accumulation and depletion layer instabilities in a medium of negative conductivity.



satellite-valley energies, thereby becoming scattered into these valleys. The large effective mass of the satellite valleys, and the possibility of intervalley scattering between them, cause the mobility of these high-energy electrons to be much lower than the mobility of the central-valley electrons. As a result, the overall average electron mobility and conductivity of such a crystal will decrease as soon as the electric field reaches a value large enough to sustain electron transfer into the satellite valleys. This, in itself, does not imply the existence of a negative differential mobility or conductivity, but once the electron transfer has started, the fraction of high-energy electrons steeply increases with the increasing electric field. Even though the velocity of those electrons that stay behind in the central valley keeps increasing in proportion to the electric field, the rate of electron loss from the high-velocity central valley into the low-velocity satellite valleys is so rapid that, above a certain threshold field, the current, and, therefore, the average drift velocity, actually drops with further increase of the electric field, leading to the negative differential mobility that is observed. In GaAs, this effect takes place somewhere between 3.0 and 3.5 kV/cm, and the transfer appears to be essentially complete above 20 kV/cm.

Similar behavior could be expected for any other semiconductor with a similar band structure; namely, a low-mass central valley combined with a set of high-mass satellite valleys. Prerequisites also include an energy separation between high- and low-mass valleys that is large compared with kT but smaller than the energy gap of the semiconductor. The latter is necessary so that avalanche breakdown does not set in before the transfer of electrons into the satellite valleys. There are, indeed, several semiconductors that fulfill this requirement and do exhibit the Gunn effect. Of the III–V compounds, these are InP,[4] and InAs under hydrostatic pressure.[21] This hydrostatic pressure must be large enough to raise the InAs energy gap above its satellite valley energy, which for low pressures is greater than the energy gap, thus leading to avalanche breakdown rather than the Gunn effect. Of the II–VI compounds, both CdTe[22] and ZnSe[23] have exhibited the Gunn effect, and this list will probably expand with future developments.

Devices without internal space charge—the LSA mode of the Gunn effect. To continue our evaluation of devices that rely upon this negative mobility, it is obvious that the conceptually simplest possible model would consist of a uniformly doped semiconductor with a pair of parallel ohmic contacts, and completely devoid of any internal space charges. In such a case, the internal electric field would be uniform and simply proportional to the applied voltage. The current, in turn, would be proportional to the drift velocity at this field level. The entire device would then have a current–voltage characteristic of the same relative shape as the velocity–field characteristic (Fig. 2). That is, it would have a voltage-controlled negative conductance. Coupled to an external resonant circuit, this device could then excite oscillations at a frequency determined by the resonant frequency of the circuit (combined with the capacitance of the device). but independent of the transit time of the electrons. This concept of an extremely simple mode of oscillation for a negative-conductivity crystal has actually been materialized by Copeland[24,25]; but it has been given a rather obscure name—the LSA mode of the Gunn effect, or simply the LSA diode. The abbreviation LSA originally stood for "large signal amplification"; more recently, for "limited space-charge accumulation." However, even the latter interpretation of these three letters obscures the fact that, conceptually, this is the simplest mode of oscillation for a negative-mobility crystal. For all practical purposes, it should be considered *the* fundamental mode of oscillation.

The term "limited space-charge accumulation" merely reflects the historical truth that, in earlier experiments, the Gunn effect had always been associated with a very strong formation of space-charge layers.[1-4] Our imperfect understanding of the effect had originally led us to believe that space-charge-free oscillations were not possible, until Copeland of Bell Telephone Laboratories, Inc., actually discovered such oscillations during computer simulations of the Gunn effect, establishing the criteria for their experimental realization. We shall discuss these criteria briefly in a subsequent section. As it turned out, Shuskus and Shaw of United Aircraft had apparently observed this mode of oscillation before,[26,27] but its nature was, at that time, not recognized.

Within a few months after Copeland's pioneering work, Copeland himself[28,29] and many others[30] utilized the LSA mode in drastically extending both the power and the frequency range of Gunn-effect devices. The highest frequencies achieved thus far with this mode are of the order of 150 GHz.* This does not yet approach the highest frequencies achieved with avalanche transit-time oscillators.[9] That is really not surprising, in view of the fact that both the materials and the device technology of gallium arsenide are much less developed than that of silicon. Nevertheless, for those frequencies at which LSA oscillators have operated, they have substantially outperformed avalanche transit-time devices with respect to both power and noise.

Since this area is undergoing extremely rapid deveopment, it would be rather pointless to quote any specific—and unpublished—achievements at this time, for they would almost certainly be obsolete by the time this article appeared in print.[46] I would like to state my belief that the chain of events, from the discovery of the Gunn effect to the discovery of the LSA mode, represents one of the greatest breakthroughs in semiconductor physics, probably the biggest since the discovery of the transistor itself. It is assuredly one that is likely to revolutionize the microwave field, particularly in the millimeter-wave region, as much as the gas laser is revolutionizing the field of optics.

Space-charge instabilities

The critical nl product. Let us now turn our attention to those modes of the Gunn effect that do involve space-charge effects. Historically, they were discovered first, since they are, experimentally, considerably simpler to obtain than the LSA mode. Conceptually, however, they are definitely more complicated, a condition that has contributed to the slow initial understanding of the mechanics underlying the Gunn effect.

Space-charge complications associated with the Gunn effect arise in a medium of negative differential conductivity because any nonuniformity of the electric field, hence any space charge, tends to build up exponen-

* J. A. Copeland, personal communication.

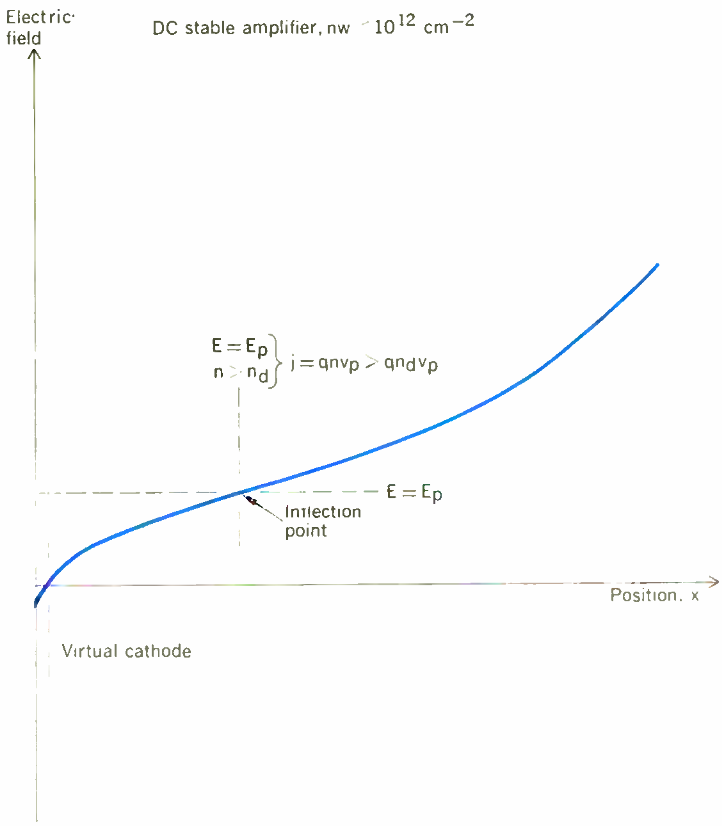
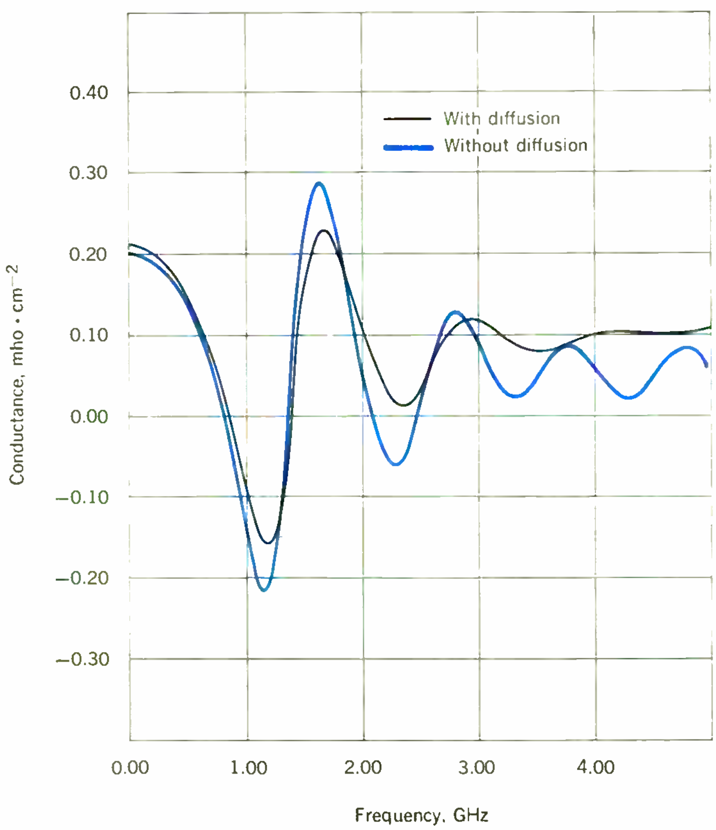


FIGURE 5. Field distribution in a subcritically doped negative-mobility crystal.

FIGURE 6. Theoretical conductance versus frequency of a 100-μm-thick crystal free from ionized impurities at a 4.8-kV/cm bias field.



tially with time. This is demonstrated in Fig. 4. The top of this illustration shows an idealized potential distribution, such as one might find immediately after the application of a bias potential. Included are three different types of field nonuniformities. The first of these is simply caused by a field transition at the cathode interface, ranging from low fields inside the cathode to high fields inside the remainder of the crystal. This "primary" field nonuniformity is, of course, inevitable. The other two field nonuniformities indicated in Fig. 4 represent the two types that may arise from either imperfect doping or merely statistical fluctuations, e.g., noise. These will be referred to as "secondary" field nonuniformities. Since, at least in the one-dimensional case, each field nonuniformity is associated with a space charge, each of these nonuniformities must correspond to either an accumulation or a depletion of electrons (called "accumulation" or "depletion" layers). The primary field nonuniformity at the cathode is related to a primary accumulation layer; the two secondary field nonuniformities are related to an accumulation layer and a depletion layer, respectively. If one now assumes that the field in all three sections of the crystal is within the range of negative mobility, then it is apparent that both the secondary accumulation and depletion layers must increase in strength with time. This follows from reasoning that the middle region, which has the highest electric field, actually has the lowest drift velocity. As a result, electrons tend to pile up at the left edge of this high-field region, where an electron accumulation layer already exists; and the electron density continues to decrease at the right edge, where a depletion layer exists. This, in turn, further increases the field strength of the middle high-field region and further decreases the field strength of the outer low-field regions—thereby also increasing the rate of space-charge layer formation.

We are confronted with a runaway process. Given enough time, it would continue until the field strength in the low-field regions had dropped so far below the threshold field that the drift velocity in these regions would once again reach the same value as the drift velocity in the high-field region. In other words, the velocity along a low-field branch may reach the same value as the minimum velocity along the high-field branch.[31] It remains to be demonstrated that this limiting space-charge buildup is only developed for sufficiently high doping levels.

Since all of this space-charge buildup takes place in a moving electron stream, both the accumulation and depletion layers travel along with approximately the same velocity as that stream. This is not only true for the secondary accumulation and depletion layers, but also for part of the primary accumulation layer, which detaches itself from the cathode and moves as an accumulation layer through the crystal. In fact, during the initial stages, this is usually the strongest of the various space-charge layers; and, in weakly doped crystals, it may remain the dominant space-charge layer throughout the entire period of travel.

Ultimately, all of these space-charge layers will disappear into the anode. Naturally, the primary accumulation layer will disappear last, followed by a rebuilding of the field inside the crystal to values above the threshold field. At that time, a new set of space-charge layers, consisting of one primary and possibly several secondaries, will form. It is this periodic nucleation and dissipation of space-charge layers that give rise to the current oscillations of the Gunn effect.

A complete theory of this growth process, particularly in the presence of statistical fluctuations in the crystal, is difficult to expound, and requires the use of numerical techniques.[32,33] However, so long as the deviations from a uniform field are still weak—during the early stages of the space-charge buildup—the growth of these space-charge layers is given by

$$Q(x, t) = Q(x - vt, 0) \exp\left(\frac{t}{\tau_n}\right)$$

where
$$\tau_n = \frac{\epsilon\epsilon_0}{qn\mu_n}$$

is the absolute value of the negative dielectric relaxation time of a crystal with negative mobility μ_n, with ϵ the dielectric constant, q the electronic charge, and n the electron density. If this law remained valid throughout the entire growth of the space-charge layer, then the maximum growth during one transit time would be given by

$$G_n = \exp\left(\frac{l}{v\tau_n}\right) = \exp\left(\frac{qnl\mu_n}{\epsilon\epsilon_0 v}\right)$$

where l is the thickness of the crystal. Now, obviously, in order for space-charge instabilities to occur, this total growth factor must be large in comparison with unit value. This means that the product of doping (electron density) and thickness (sample length) must satisfy the inequality

$$nl > \frac{\epsilon\epsilon_0}{q}\frac{v}{\mu_n} \approx 10^{12}\ \mathrm{cm}^{-2}$$

When this requirement is not satisfied, the formation of strong space-charge instabilities should not be expected.[12] In fact, oscillations probably will not occur at all, any current generated through the crystal remaining stable.

Small-signal amplification for subcritical doping ($nl \ll 10^{12}$ cm^{-2}). At first, one might expect that a subcritically doped crystal would have a static negative conductance similar to that of the tunnel diode. This, however, is not the case. In 1954, Shockley[34] had already mathematically shown that a decrease of drift velocity in an increasing electric field will not, in general, lead to a decrease in static current as the static voltage increases. The underlying explanation for this, in the case of such weak doping, is that the primary accumulation layer extends essentially throughout the entire crystal. Hence, as the voltage increases, the amount of charge stored in this layer increases in such a way that the overall current through the device continues increasing, in spite of the decrease in drift velocity of the individual electrons.

This situation is described in more detail in the dc stable amplifier of Fig. 5, where electric field versus position is shown for such a crystal. This electric field, somewhere within the crystal, must increase from the low values at the cathode contact to the high values corresponding to the applied bias. In the process, it must also pass through a plane where it equals the threshold field E_p of the velocity–field characteristic. Within this plane, of course, the velocity is equal to the peak velocity, v_p, and the current is given by

$$j = qnv_p$$

Since this plane is in the region of an increasing field, it must also be in a region of excessive electron density. Therefore, the following inequality must hold

$$j > qn_d v_p$$

where n_d is the doping density ($n_d < n$). Furthermore, this current must be constant throughout the entire thickness of the device. However, since the velocity is lower everywhere other than in the threshold plane, one must conclude that the electron density must be higher everywhere than in this plane. Since a space charge already existed in this plane, it follows that there must be an even larger

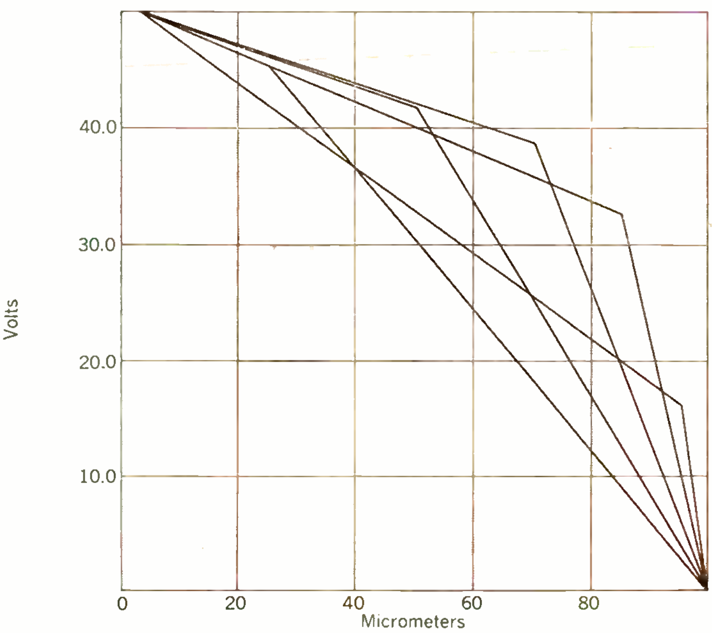
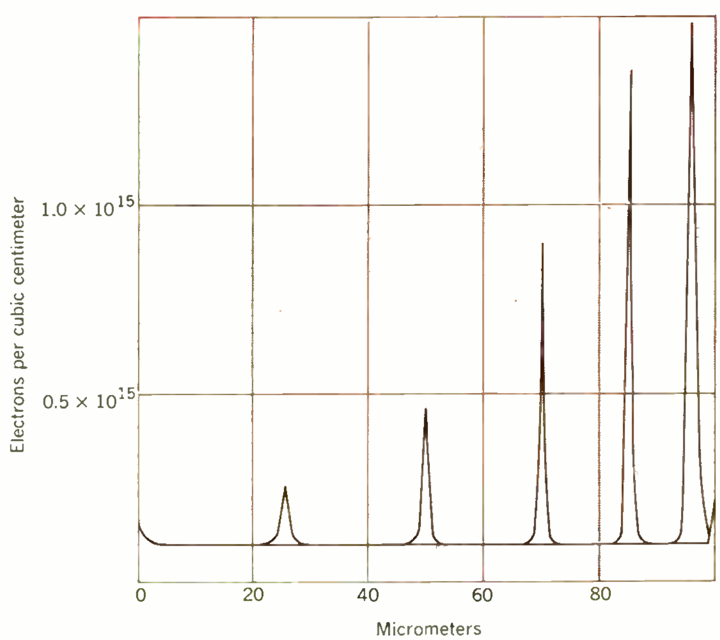


FIGURE 7. Dynamics of the pure accumulation mode, representing potential distribution for a doping level $n_d = 10^{14}$ cm^{-3} in intervals of 0.2 ns after voltage turn-on.

FIGURE 8. Electron density corresponding to Fig. 7.



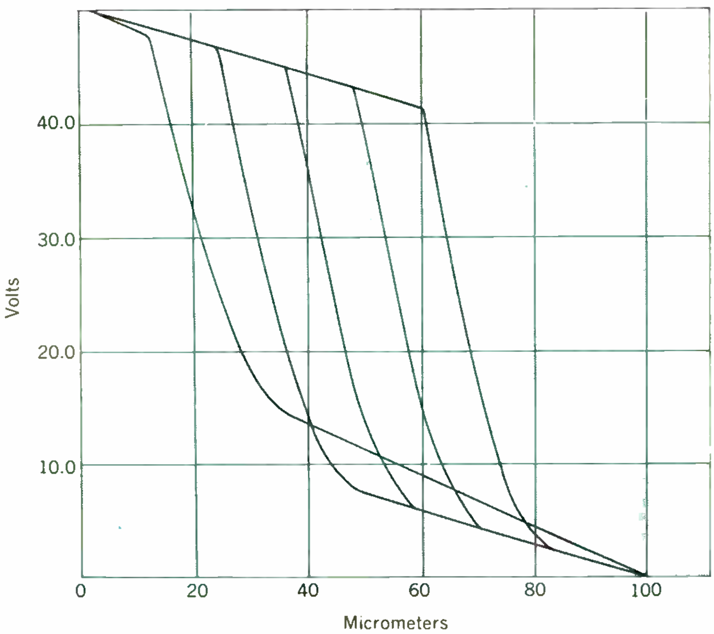
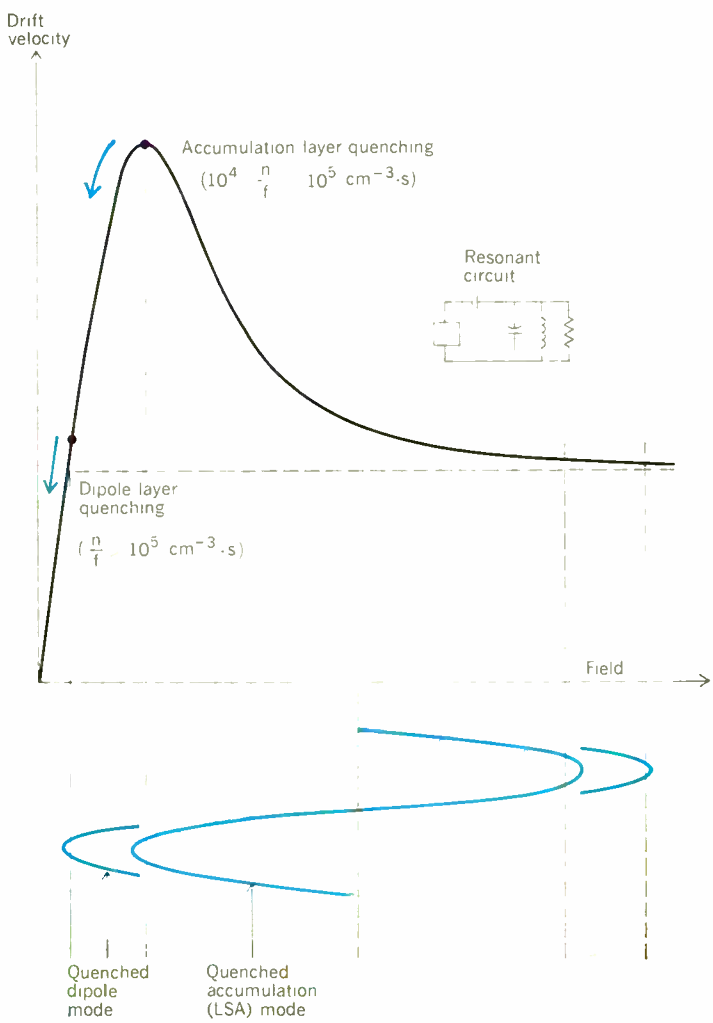


FIGURE 9. Dynamics of the mature dipole mode, representing potential distribution for the same doping level and time intervals as Fig. 7.

FIGURE 10. Quenching of accumulation and dipole modes in a resonant circuit that periodically drives the bias field below threshold.



space charge throughout the entire crystal.

With increasing bias voltage, the entire field curve will get shifted upwards. As a result, the plane in which the electric field passes through the threshold value will shift to the left—closer to the cathode. While this happens, the curve also gets steeper, indicating, of course, that the excess amount of electron density above the doping density will increase. Because the velocity in the threshold plane is, by definition, constant and equal to the peak velocity, the current density must also increase with an increase in bias. This is, of course, intrinsically what is contained in Shockley's 1954 theorem.

It was discovered by Thim *et al.*[35] of Bell Telephone Laboratories, Inc., that such a subcritically doped crystal will exhibit a negative differential conductance at high frequencies, specifically at frequencies close to the transit-time frequency. Both Thim and others[36–41] have extensively used this property to construct experimental microwave amplifiers from bulk GaAs. Our own personal investigations have included extensive research on the theory of these amplifiers,[42] particularly within the limits of zero doping and in the absence of trapping effects. Figure 6 represents one of the results of these theoretical calculations for the specific case of a 100-micrometer-thick crystal under a bias field of about 4.8 kV/cm, calculated both with and without the inclusion of electron diffusion effects. It is discerned from this illustration that there is, indeed, a pronounced negative conductance around the transit-time frequency, even for zero doping. Experimentally, such a negative conductance has never been observed within the limit of very high resistivity. This almost certainly results because such GaAs always contains high densities of traps, and the theory is not applicable to crystals with trapping effects. It should be noted that numerous statements have appeared in the professional literature to the effect that, even in the absence of traps, negative conductance within the limit of zero doping should be nonexistent. However, these statements can be attributed to improper linearizations in the theory, and are certainly incorrect.[42]

The pure accumulation mode ($nl \approx 10^{12}$ cm^{-2}).[32,33] As one increases the doping level to around 10^{12} electrons per square centimeter, space-charge instabilities—and that means oscillations—do, indeed, set in. At that doping level, however, since the dielectric relaxation time is still sufficiently long and the resulting space-charge layer growth sufficiently weak, the space-charge dynamics are entirely dominated by the propagation and growth of the primary accumulation layer. The secondary accumulation and depletion layers play a very subordinate role in this effect.

Figure 7 exhibits the calculated dynamics of the internal potential distribution for such a crystal under the influence of a 5-kV/cm bias. The increasingly sharp bends in potential, caused by the growing accumulation layer, are easily recognized. Apparently, the field near the cathode first drops to low values, and then increases to higher values—the origin of the current oscillations. Figure 8 denotes the electron density of this crystal, and shows particularly well how the primary accumulation layer grows, and, incidentally, how its movement slows down as it approaches the anode.

The mature dipole (Gunn) mode ($nl \gg 10^{12}$ cm^{-2}). Further increases in the doping level to values larger than 10^{12} electrons per square centimeter sufficiently shorten the dielectric relaxation time so that even relatively weak internal-field inhomogeneities or space-charge fluctuations can build up into fully developed space-charge layers. As it turns out, however, the simultaneous existence of more than one accumulation and more than one depletion layer is unstable. Furthermore, out of the multiplicity of accumulation and depletion layers that may initially form, only one dipole pair will remain.[32] This invariably leads to the type of propagating dipole domain that is embodied in Fig. 9.[32,33] Normally, the doping level which this illustration represents would not be sufficient to produce this mode; however, the depletion layer was created by initializing the computations with a weak electron depletion near the cathode.

As some readers might recall, such dipole layers, rather than the pure accumulation layers previously described, have actually been observed in potential-probing experiments.[6,18,43] There are perfectly comprehensible reasons that account for this preference. In order to be able to perform such probing experiments, the crystal has to be sufficiently thick. Therefore, for the available doping levels, the nl product almost inevitably becomes larger than 10^{12}; which, of course, is the condition precipitating formation of mature dipole domains.

These potential-probing experiments may have resulted in the widespread misapprehension that this mature dipole-domain mode is the one in which actual microwave oscillators operate. However, this is only true to a very limited extent, the dipole mode frequently being undesirable for this application. It is, indeed, true that this mode has been of tremendous use in those scientific experiments that have clarified the basic mechanism of the Gunn effect, but these investigations were, for the most part, performed at relatively low frequencies. In realistic microwave devices, this mode has, in fact, several drawbacks. For one, the high electric fields that can build up inside the dipole domains can lead to avalanche multiplication,[18] which is noisy and may lead to an electrical breakdown of the entire device. For another, the wave shape associated with the fully matured dipole mode is not particularly suitable for microwave oscillators. Finally, in the interest of low heat dissipation, it is generally desirable to work at doping levels as low as possible, and certainly lower than those that are required for the formation of mature dipole domains. As a result, most actual non-LSA microwave oscillators employ devices that are much closer to the pure accumulation mode than they are to the mature dipole mode.

The space-charge layer propagation in the mature dipole mode is considerably slower and much more uniform than the pure accumulation model. This uniformity indicates that during a large portion of each oscillating cycle the current density is essentially constant.

Quenched space-charge instabilities in resonant circuits

The quenched dipole mode. Thus far, we have assumed that voltage across the device is constant with time. A question arises concerning the fate of space-charge instabilities within a crystal that is inserted into an oscillating, high-Q resonant circuit, particularly if this circuit causes the voltage across the crystal to dip periodically below the threshold voltage. When this occurs, an exciting phenomenon takes place, which is clearly illustrated in Fig. 10.

Let us first consider the case of the mature dipole domain.[43–45] As indicated in Fig. 9, most of the voltage across such a crystal is dropped across the high-field domain itself. Hence, as the circuit reduces the overall bias voltage, this voltage reduction decreases the thickness of the high-field domain. Ultimately, at some particular voltage, the accumulation and depletion layers will simply neutralize each other, and the dipole domain will vanish. Upon closer inspection, this quenching begins slightly above the point at which the bias field drops to that value for which the velocity on the positive mobility branch of the velocity-field characteristic is equal to the velocity in the velocity valley along the high-field (negative mobility) branch. Once the field drops below its quenching value, the entire dipole domain disappears; and when the voltage across the device recovers, a new domain becomes nucleated immediately after the bias field regains the threshold.

In such a circuit, the oscillations will take place at the resonant frequency of the circuit itself, rather than at the transit-time frequency. If this resonant frequency is substantially larger than the transit-time frequency, this implies that the domains become quenched long before they reach the anode. In that event, the remainder of the crystal merely acts as a parasitic series resistance—except during that short portion of each cycle when a new domain is being nucleated. As a result, this quenched dipole mode, although tunable, is generally a mode of low efficiency.

The quenched accumulation (LSA) mode. If one only had an accumulation layer to quench, the resulting situation would be altogether different. In this particular case, the entire crystal section between the accumulation layer and the anode would be a series resistance of negative value. This, in turn, would not dissipate, but would actually deliver oscillation energy to the circuit (except for that short portion of each cycle during which the quenching takes place). Within the limit of sufficiently high frequencies, the penetration of the accumulation layer into the crystal will be so short that essentially all the energy that is delivered to the circuit will be generated by this negative series resistance. The role of the accumulation layer itself can then be neglected.[25]

This is Copeland's LSA mode once again, and our description has now shown how it is achieved; namely, by periodic quenching of the primary accumulation layer that detaches itself from the cathode during every cycle. It can further be shown that, in order to quench a single accumulation layer, it is not necessary to reduce the field to that point at which a dipole domain would become quenched. Instead, it is sufficient to reduce the field only somewhat below the threshold field of the velocity–field characteristic.[25] This is because the field difference across a primary accumulation layer is much weaker than the field differences that build up in a mature dipole domain—making the primary accumulation layer easier to quench. Finally, if the circuit frequency is adequately higher than the transit-time frequency, the nl product need not necessarily be kept close to 10^{12} electrons per square centimeter in order to prevent the formation of dipole domains. One can (and, in fact, should) increase n beyond this value as the circuit frequency goes up. Copeland has shown[24,25] that the proper operating condition for the LSA mode requires n/f, the ratio of electron density to frequency, to be between 10^4 and 10^5.

A conclusion

After outlining the present status of negative-conductance effects in semiconductors, I would like to conclude with a few remarks concerning possible future applications of these devices. It has often been stated that the importance of bulk phenomena lies in their potential ability to replace low-power microwave tubes. I firmly disagree! I believe that they will not replace microwave tubes any faster than transistors have replaced receiving tubes—nor is replacing microwave tubes a particularly challenging objective. The true essence of these bulk-effect devices lies in their overwhelming potential for creating newer developments in the field of microwaves that ordinarily could not have been derived from tubes alone.

In the same manner that the transistor and integrated circuit created such new fields as large-scale digital computers, in *that* sense will these devices create newer microwave applications. What form this progress will take is anyone's guess. That they will arrive is a certainty.

Revised text of a paper published originally in *Festkörperprobleme*, vol. 7, published by Vieweg-Verlag, Braunschweig, Germany, 1967; and presented at the German Physical Society–IEEE Region 8 Conference on Semiconductor Device Research, Bad Nauheim, Germany, April 19, 1967.

REFERENCES

1. For a survey of the state of this development as of the fall of 1965, see the "Special Issue on Semiconductor Bulk-Effect and Transit-Time Devices," *IEEE Trans. Electron Devices*, vol. ED-13, Jan. 1966. (See also Ref. 46.)

2. Johnston, R. L., DeLoach, B. C., and Cohen, B. G., "A silicon diode microwave oscillator," *Bell System Tech. J.*, vol. 44, pp. 369–372, Feb. 1965.

3. Read, W. T., "A proposed high-frequency negative-resistance diode," *Bell System Tech. J.*, vol. 37, pp. 401–446, Mar. 1958.

4. Gunn, J. B., "Microwave oscillations of current in III-V semiconductors," *Solid-State Commun.*, vol. 1, pp. 88–91, Sept. 1963.

5. Gunn, J. B., "Instabilities of current in III-V semiconductors," *IBM J. Res. Develop.*, vol. 8, pp. 141–159, Apr. 1964.

6. Gunn, J. B., "Instabilities of current and of potential distribution in GaAs and InP," *Symp. Plasma Effects Solids*, Dunod, Paris, 1964.

7. Hines, M. E., "Noise theory for the Read type avalanche diode," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 158–163, Jan. 1966.

8. Josenhans, J., "Noise spectra of Read diode and Gunn oscillators," *Proc. IEEE*, vol. 54, pp. 1478–1479, Oct. 1966.

9. Bowman, L. S., and Burrus, C. A., Jr., "Pulse-driven silicon p-n junction avalanche oscillators for the 0.9- to 20-mm band," *IEEE Trans. Electron Devices*, vol. ED-14, August 1967.

10. Ridley, B. K., and Watkins, T. B., "The possibility of negative resistance effects in semiconductors," *Proc. Phys. Soc. (London)*, vol. 78, pp. 293–304, Aug. 1961.

11. Hilsum, C., "Transferred electron amplifiers and oscillators," *Proc. IRE*, vol. 50, pp. 185–189, Feb. 1962.

12. Kroemer, H., "Theory of the Gunn effect," *Proc. IEEE*, vol. 52, p. 1736, Dec. 1964.

13. Gunn, J. B., and Elliott, B. J., "Measurement of the negative differential mobility of electrons in GaAs," *Phys. Letters*, vol. 22, pp. 369–371, Sept. 1966.

14. Chang, D. M., and Moll, J. L., "Direct observation of the drift velocity as a function of the electric field in gallium arsenide," *Appl. Phys. Letters*, vol. 9, pp. 283–285, Oct. 1966.

15. Thim, H. W., "Potential distribution and field dependence of electron velocity in bulk GaAs measured with a point contact probe," *Electron. Letters*, vol. 2, pp. 403–405, Nov. 1966.

16. Ruch, J. G., and Kino, G. S., "Measurements of the velocity-field characteristic of gallium arsenide," *Appl. Phys. Letters*, vol. 10, pp. 40–42, Jan. 1967.

17. Acket, G. A., "Determination of the negative differential mobility of n-type gallium arsenide using 8 mm microwaves," *Phys. Letters*, vol. 24A, pp. 200–202, Feb. 1967.

18. Heeks, J. S., "Some properties of the moving high-field domain in Gunn effect devices," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 68–79, Jan. 1966.

19. Butcher, P. N., and Fawcett, W., "Calculation of the velocity-field characteristic for gallium arsenide," *Phys. Letters*, vol. 21, pp. 489–490, June 1966.

20. Ehrenreich, H., "Band structure and electron transport of GaAs," *Phys. Rev.*, vol. 120, pp. 1951–1963, Dec. 1960.

21. Allen, J. W., Shyam, M., and Pearson, G. L., "Gunn oscillations in indium arsenide," *Appl. Phys. Letters*, vol. 9, pp. 39–41, July 1966.

22. Foyt, A. G., and McWhorter, A. L., "The Gunn effect in polar semiconductors," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 79–87, Jan. 1966.

23. Ludwig, G. W., Halsted, R. E., and Aven, M. S., "Current saturation and instability in CdTe and ZnSe," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 671–672, Aug./Sept. 1966.

24. Copeland, J. A., "A new mode of operation for bulk negative-resistance oscillators," *Proc. IEEE*, vol. 54, pp. 1479–1480, Oct. 1966.

25. Copeland, J. A., "LSA oscillator diode theory," *J. Appl. Phys.*, vol. 38, pp. 3096–3101, July 1967.

26. Shuskus, A. J., and Shaw, M. P., "Current instabilities in gallium arsenide," *Proc. IEEE*, vol. 53, pp. 1804–1805, Nov. 1965.

27. Shaw, M. P., and Shuskus, A. J., "Current instability above the Gunn threshold," *Proc. IEEE*, vol. 54, pp. 1580–1581, Nov. 1966.

28. Copeland, J. A., "CW operation of LSA oscillator diodes—44 to 88 GHz," *Bell Syst. Tech. J.*, vol. 46, pp. 284–287, Jan. 1967.

29. Copeland, J. A., and Spiwak, R. R., "LSA operation of bulk *n*-GaAs diodes," presented at the 1967 Internat'l Solid-State Circuit Conf., Philadelphia, Pa.

30. For example, Kennedy K., "Negative conductance in bulk gallium arsenide of high frequencies," Thesis, Cornell University, Sept. 1966.

31. Ridley, B. K., "Specific negative resistance in solids," *Proc. Phys. Soc. (London)*, vol. 82, pp. 954–966, Dec. 1963.

32. Kroemer, H., "Nonlinear space-charge domain dynamics in a semiconductor with negative differential mobility," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 27–40, Jan. 1966.

33. McCumber, D. E., and Chynoweth, A. G., "Theory of negative-conductance amplification and of Gunn instabilities in 'two valley' semiconductors," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 4–21, Jan. 1966.

34. Shockley, W., "Negative resistance arising from transit time in semiconductor diodes," *Bell System Tech. J.*, vol. 33, pp. 799–826, July 1954.

35. Thim, H. W., Barber, M. R., Hakki, B. W., Knight, S., and Uenohara, M., "Microwave amplification in a dc-biased bulk semiconductor," *Appl. Phys. Letters*, vol. 7, pp. 167–168, Sept. 1965.

36. Hakki, B. W., and Knight, S., "Microwave phenomena in bulk GaAs," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 94–105, Jan. 1966.

37. Thim, H. W., and Barber, M. R., "Microwave amplification in a GaAs bulk semiconductor," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 110–114, Jan. 1966.

38. Foyt, A. G., and Quist, T. M., "Bulk GaAs microwave amplifiers," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 199–200, Jan. 1966.

39. Hakki, B. W., and Beccone, J. P., "Microwave negative conductance of bulk GaAs," *Proc. IEEE*, vol. 54, pp. 916–917, June 1966.

40. McWhorter, A. L., and Foyt, A. G.," Bulk GaAs negative conductance amplifiers," *Appl. Phys. Letters*, vol. 9, pp. 300–302, Oct. 1966.

41. Thim, H. W., "Temperature effects in bulk GaAs amplifiers," *IEEE Trans. Electron Devices*, vol. ED-14, pp. 59–62, Feb. 1967.

42. Kroemer, H., "Detailed theory of the negative conductance of bulk negative mobility amplifiers, in the limit of zero ion density," to be published. (This paper also contains extensive references to other theoretical works on subcritically doped devices.)

43. Gunn, J. B., "Properties of a free, steadily traveling electrical domain in GaAs," *IBM J. Res. Develop.*, vol. 10, pp. 300–309, July 1966.

44. Carroll, J. E., "Oscillations covering 4 Gc/s to 31 Gc/s from a single Gunn Diode," *Electron Letters*, vol. 2, p. 141, Apr. 1966.

45. Carroll, J. E., "Resonant-circuit operation of Gunn diodes: a self-pumped parametric oscillator," *Electron. Letters*, vol. 2, pp. 215–216, June 1966.

46. For a state-of-the-art survey as of the spring of 1967, see "Second Special Issue on Semiconductor Bulk-Effect and Transit-Time Devices," *IEEE Trans. Electron Devices*, vol. ED-14, Sept. 1967.

Computers and computing—

Some of the concepts basic to computer design were anticipated more than 100 years ago, but real progress in this field had to wait for the advent of electromechanical and, later, electronic devices

J. Nievergelt *University of Illinois*

PAST

present

Future

Computer science has had a fascinating history, dating back to the mid-19th century. Its most dramatic progress, of course, has been over the past two decades or so. Hardware development has been phenomenal during this latter period. We have now reached the stage where the major problem of large-scale-computer manufacturers is in the area of software development. It would appear that the "speed frontier" in computer technology has been crossed, but now we are faced with the much more difficult problem of a "complexity barrier."

Computer science, the body of knowledge that has sprung up in the last 20 years in connection with and as a consequence of the proliferation of electronic computers, shares with many young and rapidly developing sciences a peculiar property: it is not possible to assess the present state of the art without touching upon the past and looking into the future as far as one can see.

It is necessary to look ahead because the conceivable future is sometimes just a decade away from the production line, and because the computer user's tomorrow is the researcher's today. And it is necessary to look into the past because we have not yet had time to assess it and draw final conclusions from past experience. Yesterday's ideas, out of fashion today, may well come back to prominence tomorrow.

The computer world's ideas of the past, present, and future therefore merge into one inseparable pool. I could not attempt to describe the state of the art without mentioning some of the history and making some predictions, wrong or right, about the years ahead.

Computer organization and hardware

Any elementary explanation of what a computer is and how it works is likely to include the type of diagram shown in Fig. 1.

Since control, memory, arithmetic, and input/output are certainly features that any computer must have, the pattern shown in Fig. 1 can be stretched to fit the organization of any conceivable computer. But the implicit message the picture conveys in assigning a separate box to each of these four functions—namely, that it is possible to separate and localize them—is increasingly at variance with recent trends toward complex systems of semiautonomous machines, many of which may be large computers in their own right.

Figure 1, which essentially dates back to the famous 1947 report, "Preliminary Discussion of the Logical Design of an Electronic Computing Instrument" by Berks, Goldstine, and von Neumann, became the blue-

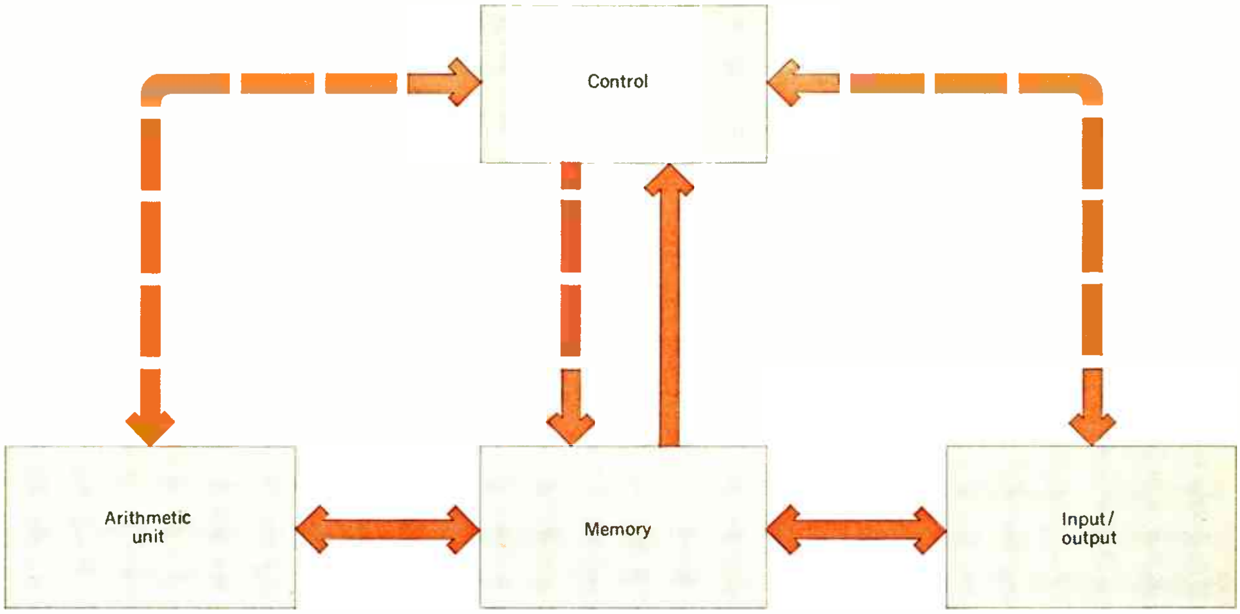


FIGURE 1. Basic computer functions.

print for a large number of computers said to be of the von Neumann type. This concept is now so well known that it suffices to point out the most essential feature that distinguishes a von Neumann-type computer from machines developed earlier or concurrently (Harvard's Mark I, the Bell Telephone Laboratories' relay computers, the University of Pennsylvania's ENIAC): the idea of storing the program that directs the machine's computation in the same memory that stores data, in order to allow the computer to modify the program as it is being executed.

Except for this stored-program concept, which allows automatic program modification,* the same design had been anticipated 100 years earlier by Charles Babbage in England. This was the age when the seafaring nations made great efforts to publish tables of logarithms, trigonometric functions, and astronomical data, and to this end sometimes employed on one project 100 people for the mere purpose of performing routine additions, which required no insight into the computational process at all. After partially completing his "difference engine," an entirely mechanical special-purpose device that could be initialized to evaluate automatically any 6th-degree polynomial at regular intervals, Babbage went on to design his "analytical engine." He was undoubtedly inspired by the Jacquard loom, a card-controlled automatic loom capable of weaving very elaborate designs. As Lady Lovelace, an able contemporary expositor of Babbage's work, pointed out, "the analytical engine weaves algebraic patterns just as the Jacquard loom weaves flowers and leaves."

Babbage and several of his contemporaries proved by their writings a complete understanding of the basic principles on which general-purpose computers operate. Their failure to influence history—to start the computer revolution 100 years earlier than it actually did—must be ascribed to the fact that the only implementation of their ideas possible at that time was mechanical, and a mechanical device of the complexity of a general-purpose computer seems to be all but impossible to build and reliably maintain.

*This feature, once considered crucial, has become less important now that computers are equipped with index registers.

It is an insight of fundamental importance that the organization of a computer has to match the technology available. Babbage unwillingly provided the first example of this principle by the unsuccessful outcome of his experiment. Mechanical devices will do for desk calculators and cash registers, but not for a general-purpose computer. The realization of his idea had to wait for the appearance of electromechanical devices. Several relay computers were successfully completed in the late 1930s and early 1940s. The designers of these machines were unaware that they were rediscovering Babbage's concept. And these relay computers were at first logically not as advanced as the analytical engine had been. I mention these facts from the prehistory of computing to point out the relevance of "old" ideas in a field that has not yet reached maturity.

From the point of view of economic impact, the history of computers might be said to begin in 1946 when the University of Pennsylvania's ENIAC initiated the era of electronic computers. The significance of this event can be appreciated by comparing electronics with the other technologies with which computer pioneers had experimented.

First, consider speed of operation. The basic time unit of a mechanical device, say Babbage's analytical engine, is of the order of one second. You just cannot accelerate, rotate by a precisely determined angle, and stop a cogwheel much faster than that. Relays decreased this basic time unit a thousandfold; a spring-loaded switch can be opened or closed in a millisecond, but not much faster than that, because it bounces back and forth a few times before it comes to rest. With vacuum tubes another factor of 1000 can be gained, since there are no movable parts at all. Today, semiconductor electronics gives us components that change state in a nanosecond—the time it takes light to travel about a foot (30 cm).

Second, consider the complexity of computers. Babbage failed in his attempt to construct a machine of several thousand moving parts. The ENIAC in 1946 combined 18 000 vacuum tubes, a rather formidable feat to this day. The IBM STRETCH, however, in 1961 incorporated 150 000 transistors. The ILLIAC IV, now being designed at the University of Illinois, will embody in its integrated circuits the equivalent of several million transistors.

Reliability, physical size, and power consumption are just some of the other reasons computer development today is vitally dependent upon electronics, and upon further advances in electronics on a similar scale as those we have witnessed in the last 20 years. As I shall now follow some of the highlights of the development of computers up to today's large-scale systems, and then extrapolate to tomorrow's regional or global information-processing and communications system, you should keep in mind some of the figures that were mentioned. Everything we have today, and much of what we expect in the future, depends on that factor of 1000 of improvement in hardware with which computer technology has been so generously provided.

The development that changed the simple organization of von Neumann's computer (Fig. 1) into today's system of processors of a large-scale computing center occurred in small steps. The first innovation was caused by the great disparity in speed between the memory, control, and arithmetic unit, which operate at electronic speeds, and the input/output (I/O) unit, which usually involves moving parts. The solution to this was straightforward, but had far-reaching implications. When data have to be taken out, the control transfers this information to a buffer, signals the I/O unit to take care of it, then proceeds with the main calculation. When the I/O unit is finished, it interrupts the main processor with a message that it is ready for further work. We thus have two rather independent machines, the central processing unit (CPU) and an I/O processor.

The number of different I/O units, or peripherals, soon mushroomed. To the paper-tape reader and punch were added card readers and punches, line printers, magnetic tapes, plotters, visual displays, and, for applications such as process control, real-time inputs and outputs that interface through analog/digital converters. And since the main memory is never large enough, backup storage devices such as drums and disks were added.

Coordinating the action of some or all of these units became a major task, well worth the services of one or two smaller computers, the satellite processors. Thus a basic disparity in speed of operation and the desire to use a costly computer in the most efficient way led, in an almost compulsory manner, to a system configuration of the type shown in Fig. 2.

Figure 2 would not have been complete without remote consoles. Time sharing, or the use of one computer by many people (apparently) simultaneously, arose from a nostalgia for the old days, when it was customary to ponder over a problem at the console and to keep the computer idle during most of this time. No doubt this is the ideal *modus operandi* for the user, but in the name of efficiency it had to give way to batch processing (that mode of operation in which the computer can be kept busy and the user kept waiting all the time). The basic idea of time sharing is again simple. Time is sliced into short intervals, anywhere from a fraction of a second to minutes, and the slices are given according to some priority scheme in turn to the various users who sit at consoles and request time. An additional advantage is that these consoles, which are at present usually teletypewriter units, can be far from the central computer, linked by a telephone line, so that a user has access to a large computer directly from his office or home.

From this steady expansion in the complexity of computer systems in the past there emerges clearly a vision for the future: that of a network of computers, data banks, and terminals, all linked by communication lines, which spans a wide area such as the North American continent. You, the average citizen, lease an information-processing and communications terminal from this network just as you rent a phone now. But apart from having a phone, you will have direct access to the nation's libraries. Your children may receive their education at home by programmed instruction. You may attend a conference while remaining at home by turning on a closed-circuit-television link so that you can hear and see each of your business partners. As J. R. Pierce of Bell Laboratories aptly puts it, you will not commute to work any longer, you will communicate to work. Computation (or information processing) and communication (or information transmission) are two aspects of the same activity, and soon there may be no way to tell them apart.

The greatest problem in implementing a global information-processing and communication network will undoubtedly be the software to make it work. We will see later that providing the software for a computer system of the kind shown in Fig. 2 currently taxes our capability. First, I must mention a trend toward the opposite of the picture I have drawn so far—the small computer, costing less than $10 000, of which several types are now on the market. The small computer has definite advantages over a terminal hooked up to a large computer. It does not cause the delays (sometimes years) that seem to be unavoidable in the delivery of hardware and software of large systems; it is reliable and easy to maintain; and last but not least, a small computer does not cost much more than some remote consoles. The limitations of small computers are obvious: no access to large files of data, little programming support, and the impossibility of performing an occasional "large" calculation.

The availability of families of compatible computers, introduced by the IBM 360 series, has added considerably to the attractiveness of the small computer. It is now possible to change from one computer to the next bigger one in the line with a minimum of reprogramming. Hence, small computers will increase in importance in the near future for providing the easiest way for a small user to enter the computer field. It is to be expected, however, that in the long run small stand-alone computers will be replaced by remote consoles with some processing ability of their own, with a large computer system in the background available when it is needed.

Software and applications

Software—the sets of programs required to be supplied with a computer in order to satisfy the demands on versatility, efficiency, and ease of use that have become customary—has emerged in the past 20 years as the

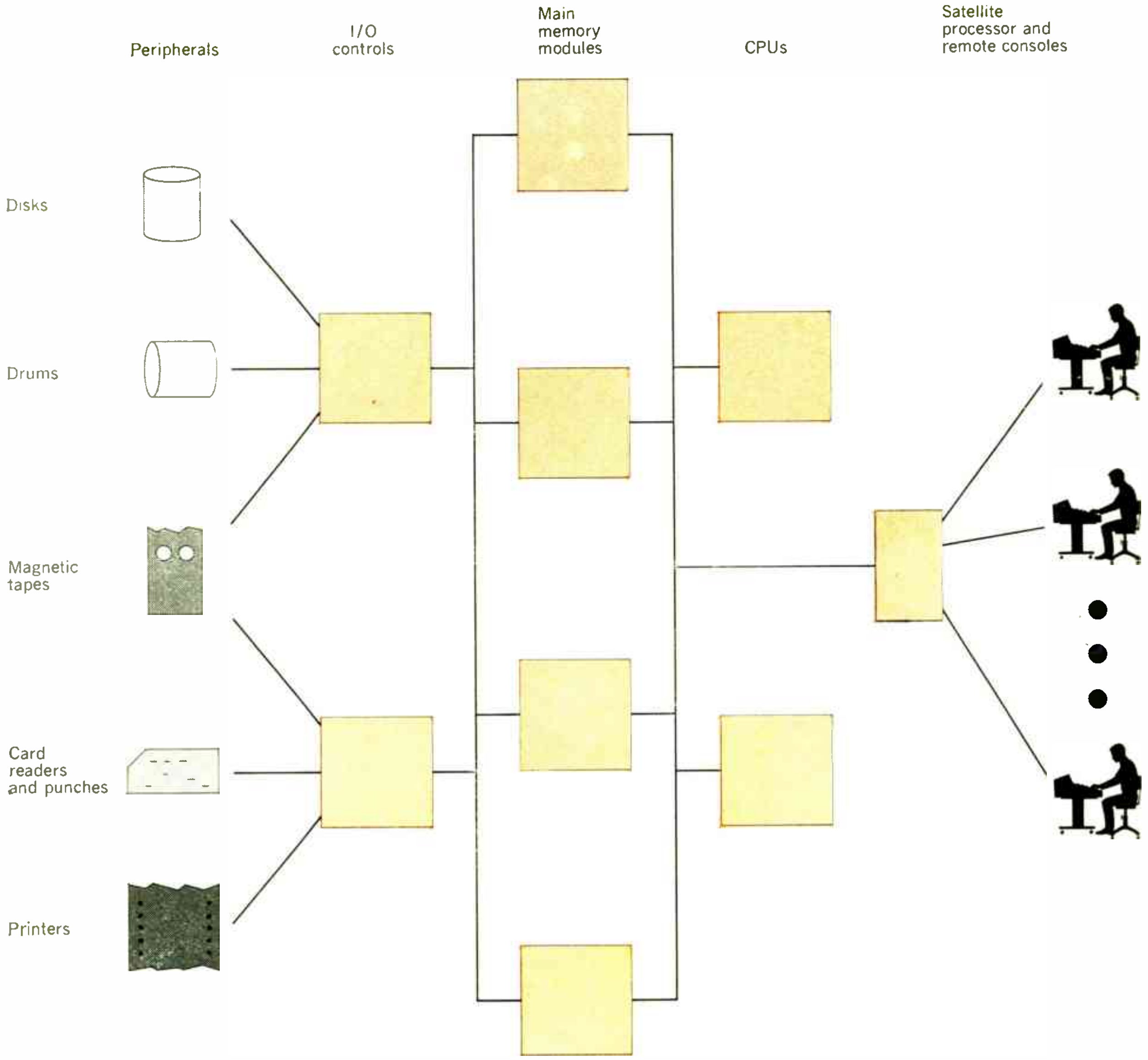


FIGURE 2. Typical computer system configuration.

major problem of the manufacturers of large-scale computer systems. The resources put into the development of software for one of the new families of computers may well exceed those put into the development of the machines.

To understand the reason for this "software explosion," a phenomenon that I doubt anyone would have predicted 15 years ago, one has to consider the patterns of computer development, from the von Neumann-type computer of the late 1940s to today's computer systems configurations, as was done in the preceding section. The analogous developments in the range of computer applications and the demands of the user must also be examined.

Babbage embarked on his "analytical engine" project in order to automate the computation (and the printing) of tables of numbers. The ENIAC computer was designed and used to compute ballistic trajectories. Von Neumann had in mind mainly the solution of partial differential equations of hydrodynamics. A program for this kind of numerical computation is not very complicated: 1000 instructions are sufficient for many numerical computations that run for hours. Moreover, it was not expected that many programs would be written. A program's lifetime was intended to be long, and hence any effort in programming was supposed to be a long-range investment. Under these conditions it made sense to write a program in binary form with absolute addresses, with every bit of the program explicitly specified by the programmer. Nothing but the user's program, the computer, and the flipping of a few switches are involved. This situation could well be called a Garden of Eden configuration, because it is unlikely ever to occur again.

The initial steps toward software, which led to assemblers first and compilers later, were in response to users' demands for convenience. It is easier to remember, for example, that MPY instead of 011010 denotes multiplication; thus, mnemonics enter the picture. And it is preferable to know what X and TOTAL mean than to keep track of what should be stored in locations 11010010 and 11010011; thus, symbolic addresses are introduced. Mnemonics and symbolic addresses ushered in the age of

I. General-purpose digital computers installed in the United States

Year:	1955	1958	1961	1964	1967
Dollar value, millions:	75	460	1100	2000	4200
Number:	244	2550	7550	18 200	36 000

Figures courtesy of International Data Corporation.

assemblers. Assembly language, augmented by macro-facilities and program libraries, became the programmer's major tool in the 1950s and proved highly successful. It provided complete control over the computer, down to the individual bit, yet removed most of the drudgery involved in "octal absolute" coding.

Some experienced programmers still prefer assembly-language coding to some of the later methods. However, in order to provide a means for occasional users, who do not want to learn the intricacies and details of assembly languages, to write their own programs, "autocodes" came into existence. These codes enable the users to write such statements as $X = 3.14*Y + X$, which may be equivalent to many machine instructions. Subsequent generalizations of autocodes are called *compiler languages* (because elaborate programs called compilers are needed to translate statements similar to the foregoing statement into machine language), *higher-level languages* (because they allow the programmer to omit the details and to concentrate on the essentials of his program), or *problem-* or *procedure-oriented languages* (because they are designed to suit one type of problem particularly well). Initially, problems usually were basically related to either numerical computation or business-data processing. This fact is evident from the names of the two most widely used representatives of this type of language: Fortran (derived from "formula translation") and Cobol (derived from "common business-oriented language").

With the spread of assembly- and compiler-language programming, closed-shop batch processing became the accepted mode of operating a computer. The user no longer had direct access to the machine; instead he submitted a deck of cards and received his results after a turnaround time ranging from hours to days.

In this situation, which had become typical by the early 1960s, software was already an important, if not yet critical, ingredient of a large-scale computer system. In particular, it differed in one significant aspect from today's software: when a user's program ran, *it* was in control of the machine. In the next phase of computer development, however, the omnipresent operating system would never quite relinquish control. It would monitor and time the user's program, then kick it off in order to call another one.

Operating systems—the programs that regulate the flow of other system programs, user programs, and data through the computer while allocating the computer's resources to the various activities that proceed at any one time—became a necessity for several reasons: the use of many I/O units that operate over a wide range of speeds; a hierarchy of storage, ranging from the fast-access main memory to slow, large backup storage devices; the expansion of the computer to several concurrently operating processors; and, finally, the proliferation of programming languages. A computation center no longer had its own programming language but had to have several compilers, interpreters, and subroutine packages so that the user could choose the language to suit his problem: Fortran, Algol, or Mad for numerical computation; Cobol for business applications; Lisp or IPL-V for list processing; Comit or Snobol for string manipulation; GPSS or Simscript for simulation. Along with these exists a host of languages often used only locally or for special purposes.

The task of designing, implementing, and debugging the software for a line of computers is a monumental task, not only because of the sheer magnitude of the undertaking but also because of the logical difficulty of the effort needed. For the latter reason, the brute-force approach of putting a thousand system programmers on the job has not been very successful. In fact, it has invited the joke: "If it takes one woman nine months to produce a child, how long does it take nine women?" But, on the other hand, alternate approaches have worked no better, and thus delays in software delivery have become standard.

We are pushing against the limits of complexity that we know how to handle. Never before have programming projects of such a degree of complexity been seriously attempted as those that occur today in software development and in a few specific applications.

Computing has many frontiers; the one of speed was emphasized earlier. But many programs that tax the speed of our fastest computers can be programmed with a moderate effort. They are of the same kind of problems as those for which computers were originally designed: logically simple but computationally hard. However, we also run into problems that are computationally light but logically so complex that we do not know how to design, document, and debug them. This kind of problem is relatively new. But since we also face these difficulties when we are trying to expand the use of computers to other problems not of the traditional scientific computation type (short code, long run)—problems that are potentially among the most challenging applications of computers—this complexity barrier may well be the most important limitation that the computer field will be subject to in the immediate future.

Where are we now?

According to all measurable parameters, we are still on the way up. Table I gives the estimated annual value of general-purpose digital computers shipped by American manufacturers and the estimated number of general-purpose digital computers installed in the United States.

In the face of such imponderables as the basic understanding of computers and their applications and, perhaps even more important, their implications to our society, it is almost impossible to predict the future. Much more thought, experience, and time will be required before computer science becomes a stable, dependable constituent of our civilization. As usual, however, people will not wait for this to happen before exploiting the computer's potential as far as they can, and hence we can expect a continued rapid expansion of the field for some time. But it should be firmly kept in mind that we may not always know what we are doing.

Revised text of a talk presented at a conference of the American Gas Association, St. Louis, Mo., May 3, 1967.

The high-brightness

The advent of the light-emitting diode has given us a device that offers a significant increase in performance over previous devices, thus also significantly increasing compatibility with microcircuits

When the visible-light-emitting diode became commercially available in 1963, it brought to the market a device that promised advantages in the areas of efficient and reliable operation and of cost. This discussion of the light-emitting device considers some of the trade-offs with regard to the LED as well as the kind of performance that can be expected in actual use at the present state of the art. Particular emphasis is placed on the significant increase in performance over previous similar devices and the resultant increased compatibility with microcircuits.

Visible-light-emitting diodes have been understood conceptually since about 1956, and they reached the marketplace in 1963. With these devices came the promise of high efficiency, low-voltage operation, reliability, and low cost, and hence compatibility with microcircuits. Earlier devices were high in cost, low in efficiency, had unknown reliability, and "perhaps" were compatible with high-power microcircuits.

Diode materials

Visible-light-emitting diodes (LEDs) are commonly made from gallium phosphide (GaP), silicon carbide, and the alloy gallium arsenide phosphide (GaAsP); our work has been with GaAsP. When GaP is alloyed with GaAs during the crystal-growing process, the band gap of the resulting semiconductor increases with the mole percent of GaP. Gallium arsenide is a direct semiconductor, one with a high probability of direct-band radiative minority carrier recombination. This recombination process is very efficient; internal efficiencies near unity have been reported for GaAs. Unfortunately, GaP is an indirect semiconductor, with a band structure not unlike that of germanium and silicon. Direct-band transitions are an improbable second-order phenomenon in these materials. However, GaAsP is a direct-band semiconductor for compositions of 45 percent GaP and less and efficient light-emitting devices can be easily produced with the direct-band materials. Above 45 percent GaP the efficiency of photon generation falls off rapidly. For 48 percent GaAsP the band gap is approximately 1.975—and the emitted photon energy is nearly equal to the band gap, according to Einstein's equation:

$$E = \frac{hc}{\lambda}$$

$$\lambda = \frac{12\,400\ \text{Å}}{E_g}$$

where E_g is the band gap in electronvolts. Thus, 48 percent GaAsP will produce light-emitting diodes with peak wavelengths of 6300 Å. The useful peak wavelength capability range of GaAsP is from nearly 9000 Å, for GaAs, to 6300 Å.

The alloy GaAsP is commonly produced by the epitaxial deposition of the alloy material on a GaAs substrate. The growth of large single crystals by the Bridgman or Czochralski techniques has been limited by the high partial pressures of phosphorus, up to 45 atmospheres, needed to maintain equilibrium. In this material p-n junctions can be useful as radiation sources. The minority carriers needed for radiative recombination are generated by the injection of electrons into the p region when the device is forward-biased.

Junctions are readily formed by taking advantage of the qualities of zinc, an acceptor that diffuses rapidly at modest temperatures. Junction regions can be defined by etching mesas—leaving separate p-type islands—or by diffusing discrete areas through a silicon oxide or silicon nitride diffusion mask. Separate areas on a flat field are useful for monolithic arrays of devices, allowing relative ease of dielectric isolation for metalized contacts.

Optical measurement

As the LEDs emit visible light with forward bias, we shall discuss their radiation characteristic as well as a method for optimization of the emission to a measure-

LED

D. K. Hillman *Monsanto Company*
G. E. Smith *Fairchild Semiconductor*

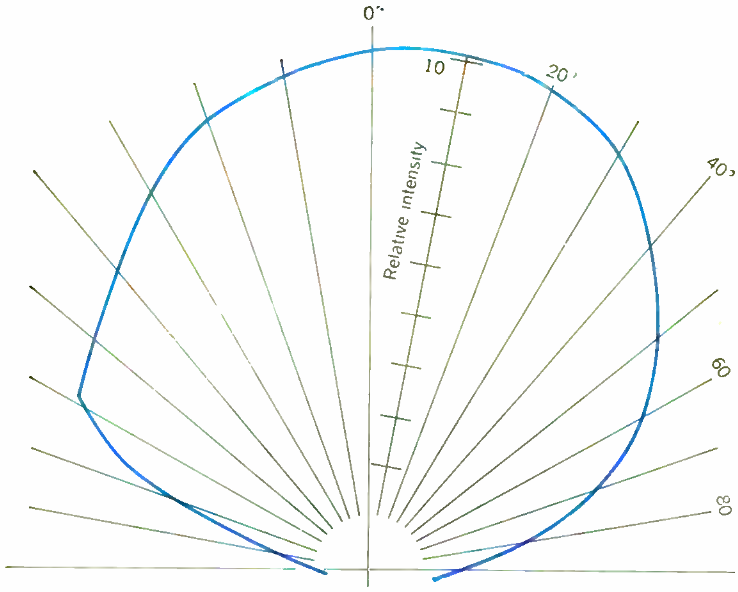
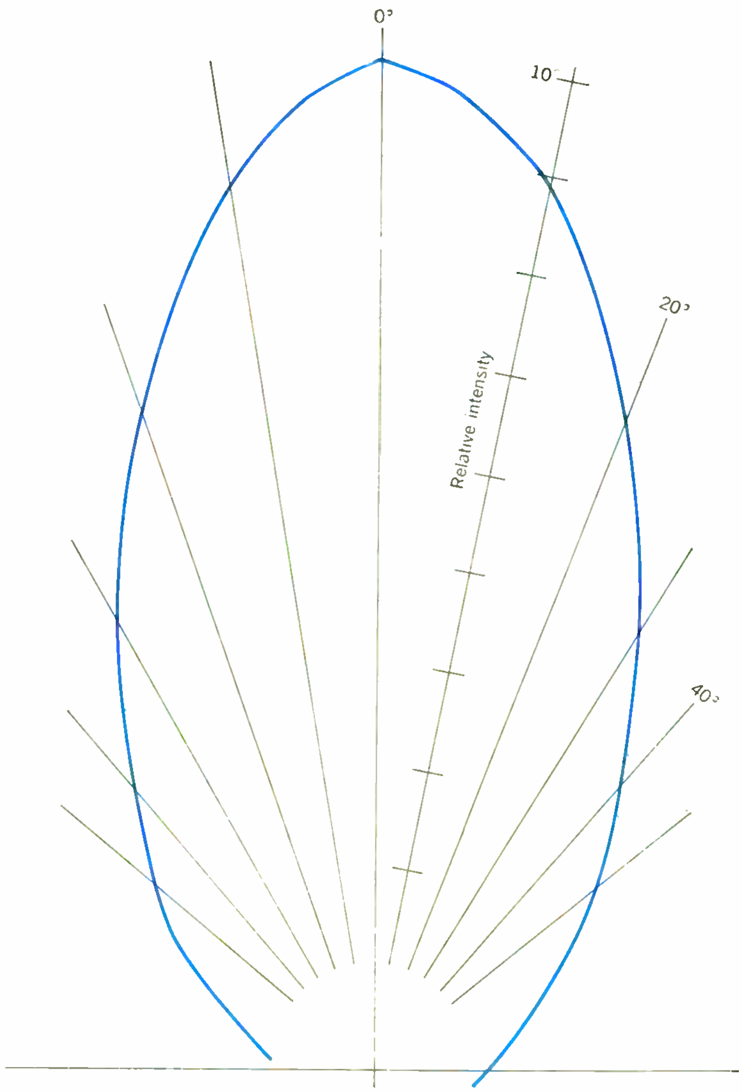


FIGURE 1. Angular distribution of GaAsP diode without lens.

FIGURE 2. Angular distribution of GaAsP diode with epoxy lens.



ment related to the human eye.

The spectral distribution is represented by a sharp peak. The radiation pattern is skewed slightly by optical band-edge absorption on the lower-wavelength side. The peak varies in wavelength depending on the semiconductor composition. The amplitude is affected by the diode conversion efficiency as well as by gains in input power and by the type of measurement amplifier used. At half-intensity the spectral width is typically 300 Å, although we have seen devices with half-widths as narrow as 150 Å and as wide as 500 Å.

The efficiency and radiated power can be measured with a solar cell. Internal quantum efficiency describes the ability of the device to convert electron current to photon current and external quantum efficiency is defined as the photon current leaving the device as compared with the electron current. The external efficiency is far below the internal efficiency, and is limited by absorption, reflection, and refraction losses.

Neither efficiency nor total radiated power alone will indicate how much visual stimulation a light source will produce, because the sensitivity of the human eye is strongly dependent on wavelength. Standard luminosity curves show that peak sensitivity occurs at 5550 Å and is down four orders of magnitude at 7500 Å and 3900 Å. A sharply peaked light emitter at 6500 Å falls on a steeply rising portion of the luminosity curve. Thus, changes of a few hundred angstroms in peak wavelength can grossly affect visual stimulation. You may say that this is no problem; simply choose a composition that peaks near 5500 Å and stick with it. However, as fast as the luminosity curve rises, the efficiency obtained falls at an even faster rate.

The lower the wavelength, the higher the visual stimulation, but also the lower is the total radiated power. Therefore, a wide range of wavelengths can have similar brightness.

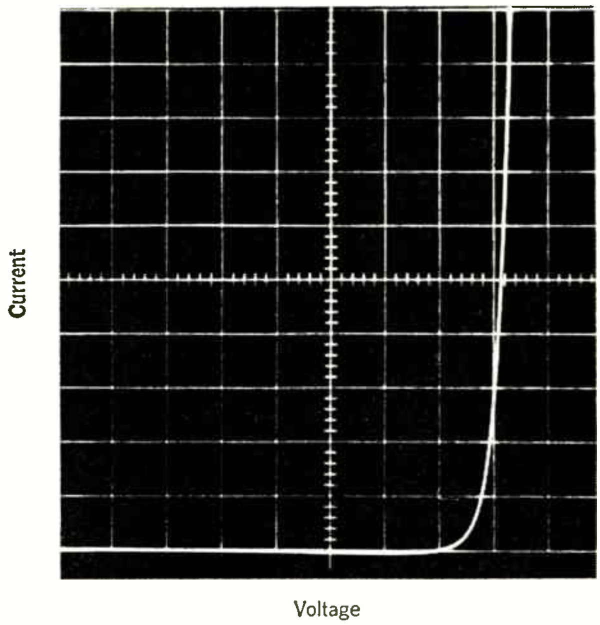
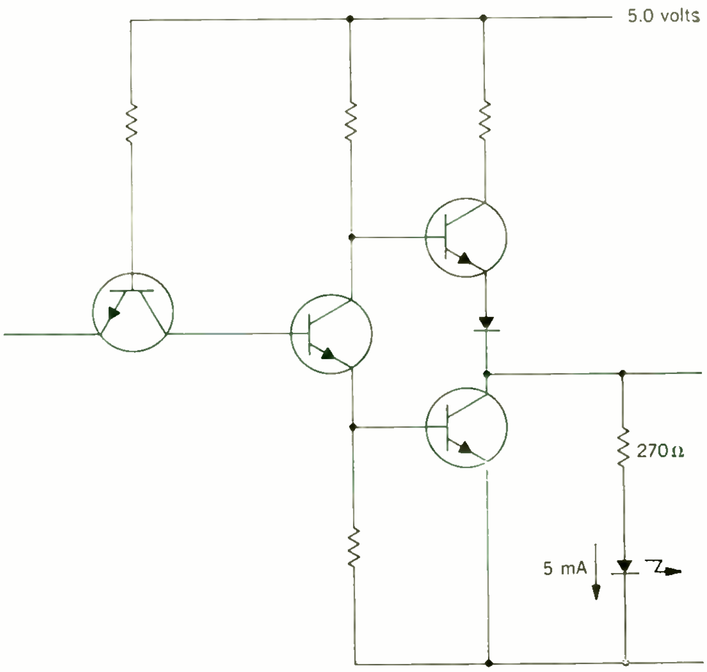


FIGURE 3. Current-voltage characteristic of GaAsP diode. (Scale: 1 mA/div and 0.2 V/div)

FIGURE 4. Typical T²L gate driving GaAsP diode.



The photometric unit

Common usage for the photometric unit of visual stimulation for emitters such as phosphors and electroluminescent diodes is the brightness of the surface, expressible in footlamberts (fL). This photometric unit for light emitters presents serious problems, not the least of which is that the IEEE Standards Committee has discouraged its use. The physiological brightness characteristic is a function of the intensity of the source, but it is also a function of the observer and the ambient lighting conditions. These problems are eliminated for purposes of characterization of the light source by the use of a physical measurement quantity, luminance, which is proportional to the luminous intensity per unit projected area. This quantity can be calculated from radiometric measurements. In particular, if the total radiated power is known from a solar cell or thermopile measurement, then the luminous efficiency can be computed by evaluating the following integral:

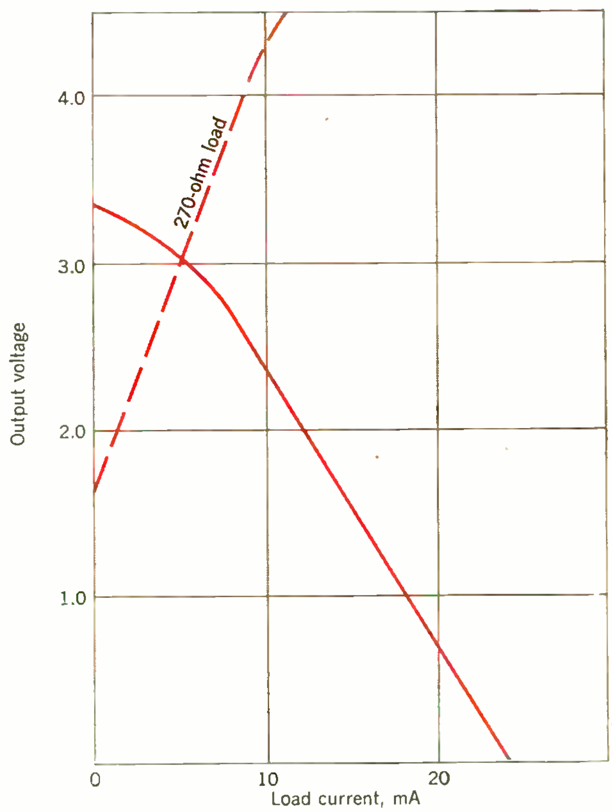
$$K = \frac{\left[\dfrac{680 \text{ lumens}}{\text{watt}}\right] \int_0^\infty \bar{y} P_\lambda \, d\lambda}{\int_0^\infty P_\lambda \, d\lambda}$$

where y is the photopic luminosity curve and P_λ is the radiated flux per unit wavelength λ. The luminous flux F is then

$$F = K \int_0^\infty P_\lambda \, d\lambda$$

The intensity is flux per unit solid angle for vanishingly small solid angles. Thus,

FIGURE 5. The "1" logic level; $V_{CC} = 5$ volts.



$$I_\theta = \lim_{\omega \to 0} \frac{F_\theta}{\omega}$$

where I_θ is the intensity at angle θ from the normal.

Brightness is the intensity per apparent area:

$$B_\theta = \lim_{A_\theta \to 0} \frac{I_\theta}{A_\theta}$$

where A_θ is the area at an angle θ. For plane surfaces $A_\theta = A \cos \theta$. This is not true for the semiconductor emitter because of aberrations by the epoxy lens.

As can be imagined, the calculation from radiated power measurements is a tedious one, needing corrections to primary standards for a monochrometer and solar cell. An easier method, which avoids going through the calculations, is to employ one of the commercially available photometric measuring instruments.

The active part of such an instrument is a photomultiplier with a suitable correcting filter that approximates the standard observer luminosity function. A telescope in the instrument allows the selection of an 0.02-cm-diameter spot on the semiconductor emitter that is projected on the photomultiplier through an aperture. The combination of aperture and accurate focus defines both the solid angle subtended by the detector and an area of the emitting surface; thus the photomultiplier output is proportional to the brightness. The instrument is calibrated by a secondary brightness source and a series of red filters, supplied by the National Bureau of Standards, which are necessary to correct for inaccuracies in the sensitivity as compared with the luminosity function. When it is properly calibrated, the brightness meter reads directly in footlamberts. With it we can specify the performance of a particular device and, probably more important, we can compare different groups of diodes.

It should be pointed out that the level of radiation of emitting diodes is a function of current density. In any discussion of this function, we must know the brightness per unit of current density. For example, a 0.4-mm-diameter mesa, emitting light to a surface brightness of 50 fL at 50-mA forward current, has a brightness figure of merit of about 1 fL/A/cm². The figure of merit must be adjusted to the apparent area of the chip if a magnifying package is used.

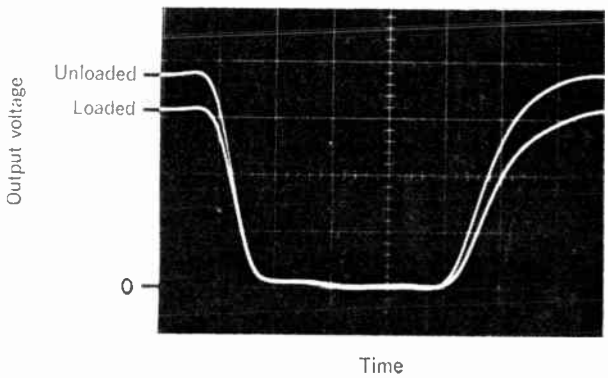


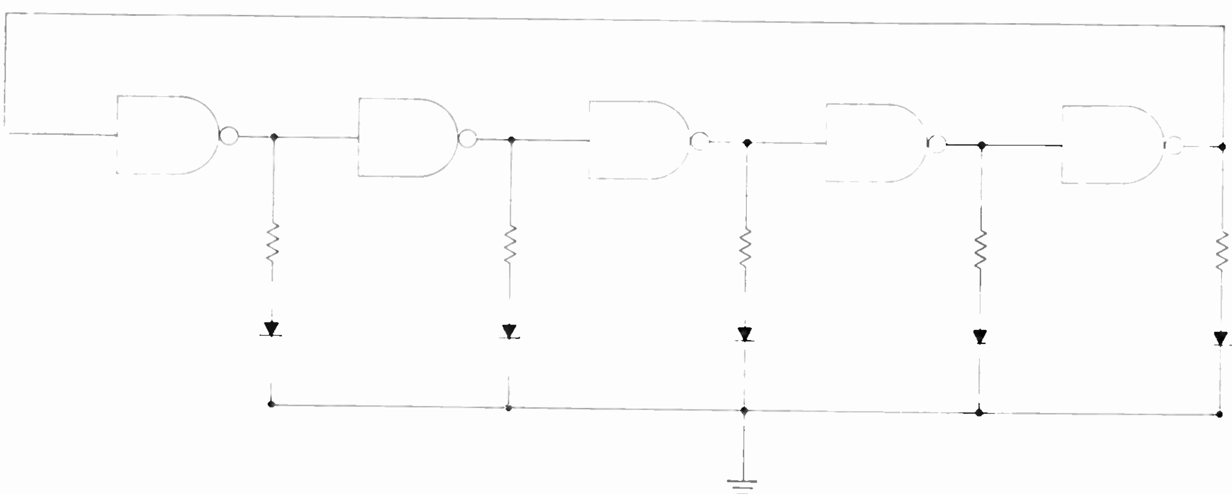
FIGURE 6. T²L circuit loading by GaAsP diode and 270-ohm resistor. (Scale: 1 V/div and 10 ns/div)

The improvement in performance

We have made some progress in improving the quality and quantity of light from LEDs, as evidenced by the spectra typifying an improved device. The changes in performance include virtual elimination of a broad secondary peak. This peak has been associated with a deep impurity band, which has been reduced and finally eliminated by extreme care in the material-growth and device-fabrication procedures. The secondary peak makes evaluation of devices by comparison techniques difficult because that radiation is not visible, even though it does represent radiated energy and improves the measured efficiency of the device. The secondary peak represents a competing radiative process; by eliminating it, the energy represented by the primary peak is increased.

Also, the intensity of radiation in the peak is considerably higher in the improved device. We can compare the two devices, by use of a brightness meter. We see that the first diode exhibited a brightness of 140 fL at a 50-mA forward current, whereas the improved device showed 550 fL under the same conditions. This case represents an

FIGURE 7. Propagation delay measurement with T²L gates loaded with GaAsP diodes.



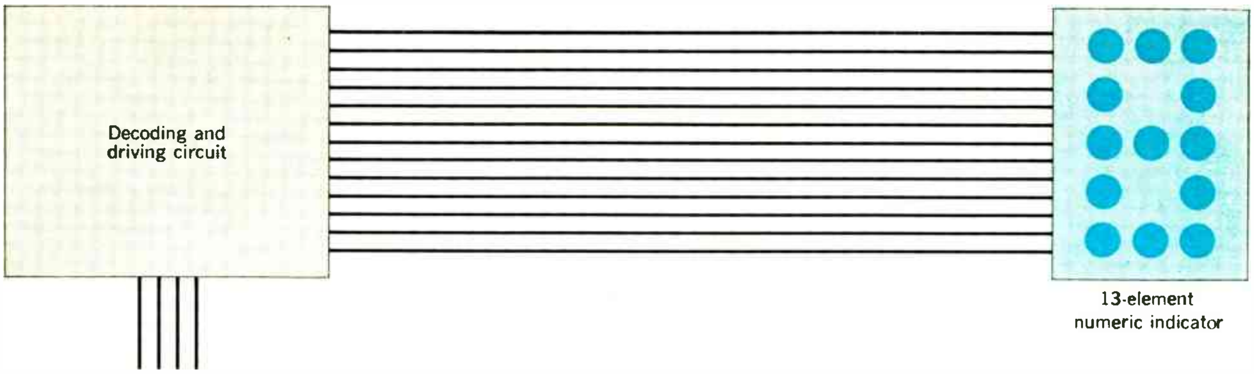


FIGURE 8. Alphanumeric character readout.

improvement of >290 percent; improvements in brightness of 230 percent are typical. The important point to remember is not that one can obtain 1000 fL at 150 mA but rather that good visibility is possible at, say, 30 fL at 5 mA. As efficiency increases, power consumption goes down for a constant brightness, which is what is going to interest the microcircuit user.

Figure 1 presents the angular distribution from a diode without any encapsulant. This distribution closely represents a circle, which is what is expected for a Lambertian, or perfectly diffuse, radiator. The index of refraction of GaAsP is very high, about 3.5, and hence internally generated radiation will be internally reflected for angles greater than 16 degrees. This internally reflected emission is lost by absorption. When an epoxy overcoat is placed on the surface, the critical angle into the epoxy is increased to about 25 degrees. If the epoxy surface is curved, we can take advantage of the increased emission from the chip.

Figure 2 shows the angular distribution of a device with an epoxy lens. The result of the curved epoxy surface can be seen in the increased directionality of the emission. Typical lenses produce a factor of two increase in intensity.

Figure 3 represents the electrical performance of the diodes. The devices can be characterized electrically by parameters similar to those of other semiconductors. Typically, at 25°C: V_F = 1.6 volts at 10 mA, $R_F < 5$ ohms at 10 mA, $dV_F/dT = 2.3$ mV/°C, $C = 120$ pF at $V = 0$ volts, and $P_{max} = 125$ mW. The reverse characteristics are uncontrolled and no recommendation can be given for operation in this region.

The oscilloscope trace of Fig. 3 represents the forward current–voltage characteristic of the GaAsP diode. The current sensitivity is 1 mA/div, which indicates the kind of voltages needed at the power levels we can expect from microcircuit use. At 1.6 volts the device draws 3 mA.

With the improved performance now obtainable, 20 to 30 fL can be achieved at a 5-mA forward current. For example, the diode represented by Fig. 3 read 38 fL at 5 mA. This is a suitable level for use with logic elements for diagnostic purposes. Several forms of integrated circuit logic gates can drive these diodes without severely affecting proper operation of the circuit. The various T^2L families are most suitable.

Shown in Fig. 4 is an example of a T^2L gate with the diode in series with a 270-ohm resistor in the output. When the gate is in the "1" level, the diode is limited to about 5 mA. Resistors in the range of 200–500 ohms can be used to obtain a suitable operating current without overloading the gate output.

Figure 5 presents the output voltage versus load current of the gate. The 270-ohm load line has been included; note that it is offset by 1.6 volts when the diode begins to conduct. Since the 270-ohm resistor is large compared with the dynamic forward resistance of the diode, it masks any effects of the diode. At the "1" logic level the output will set at about 3 volts with 5 mA in the diode.

Oscilloscope traces of the output of the gate when the input is pulsed are given in Figure 6. The upper trace is the unloaded output; the lower is the output when loaded with the diode and resistor. The resistor provides some decoupling from the diode capacitance and allows the gate to operate at a reasonable speed. Full fan-out capability of the gate is retained since worst-case input for these particular circuits is 2 volts.

Figure 7 presents a ring oscillator that is used to measure the propagation delay of the logic gates. The odd number of inverting gates forces the circuit to oscillate. Propagation delay is related to the period of oscillation. Typical values for this gate are supposed to be 13 ns. In our arrangement, without the diodes in the circuit, the propagation delay was typically 6.5 ns. When each gate was loaded with a diode and a 270-ohm resistor the delay increased to 7.0 ns. The power required in an application such as this (15 to 30 mW) is comparable to that needed for other low-power lamps.

Finally, Fig. 8 alludes to the possibilities of alphanumeric character readout. A 3 × 5 numeric array can be operated from 2 volts, consuming less than 50 mA or 100 mW, as compared with 500 mW for neon-type indicator tubes. The current levels are such that a monolithic integrated circuit decoder and driver can be built using current techniques.

Conclusion

Although we have seen a fourfold increase in the brightness of GaAsP diodes, we do not wish to imply that there will be no further improvement. Devices have been tested in the laboratory that, under the same conditions as described, have measured 1000 fL at less than 15 mA input—which is 60 times the brightness that we had to strain to achieve a short time ago.

This work was supported in part by Wright-Patterson Air Force Base, Contract AF 33(615)-3618.

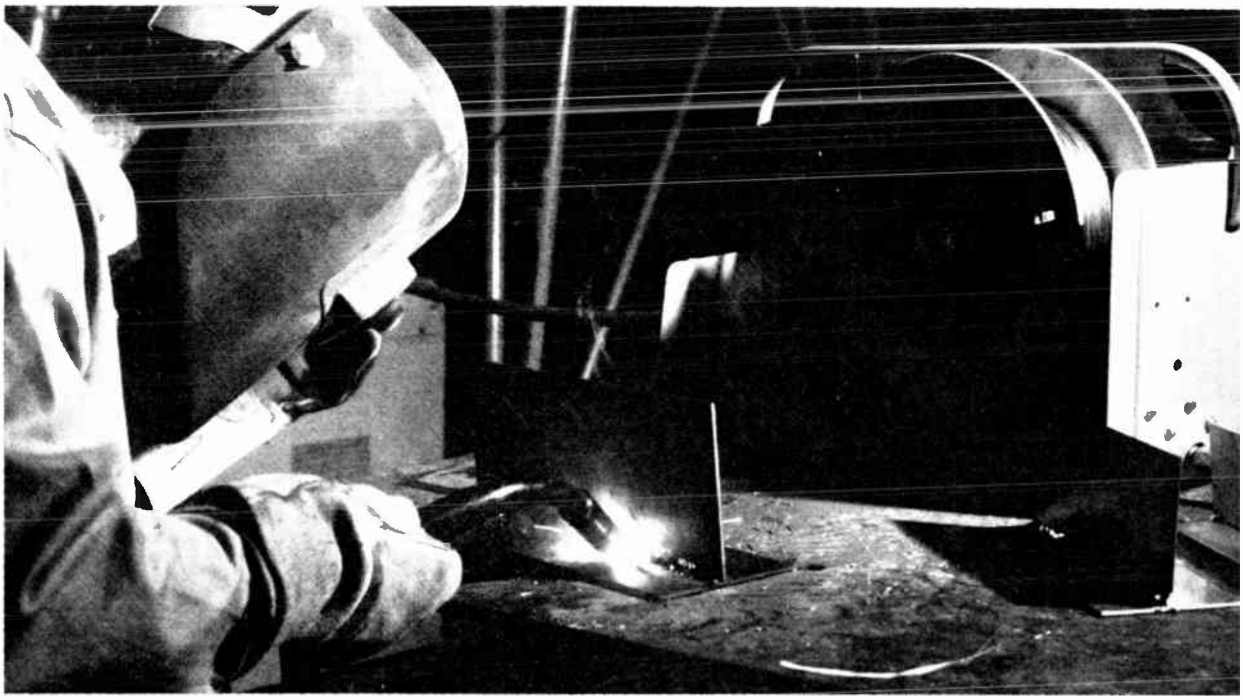


FIGURE 1. The gas-metal-arc process employs a wire consumable electrode that is continuously fed through the welding torch within a cone of shielding gas. Hand-held torches, as shown, can be used in either horizontal or out-of-position welding. Mig torches are also used in automatic welding machines.

A power supply for pulsed power welding

The dc power supply parameters are all-important in gas-metal-arc welding: too little or too much current can either ruin a weld or make operations difficult. Pulsed power welding seems to be an ideal compromise

E. H. Daggett *Air Reduction Company*

The use of pulsing current for gas-metal-arc welding has increased the versatility of the method because of the ease of varying the average current to suit the application. A power supply consists of two properly designed constant-potential power sources in parallel to provide two levels of current. The pulse current is supplied by a half-wave rectifier, and background current by a filtered, three-phase rectifier. The dynamic characteristics must be considered to get good arc stability during current commutation between the two supplies.

The advent of gas-metal-arc (Mig) welding (Fig. 1) has provided industry with one of its most useful tools for joining metal to metal. It is a radical departure from "stick" welding, which uses individual, coated electrodes.

The gas-metal-arc torch utilizes a consumable electrode in the form of a small-diameter wire, which is continuously fed through the torch by a feeder mechanism. Power is supplied to the wire electrode at the insulated torch. The arc between the wire electrode and the workpiece is shielded by a cone of inert gas, to prevent contamination of the weld by atmospheric oxygen or nitrogen. A cross section is shown in Fig. 2.

The power supply voltage and wire-feed speed are two of the most critical parameters in the Mig process. Normally a filtered dc supply is used; but this presents some problems. Too low a current will produce shallow, irregular welds; too high a current causes metal to flow so freely as to preclude all but horizontal work.

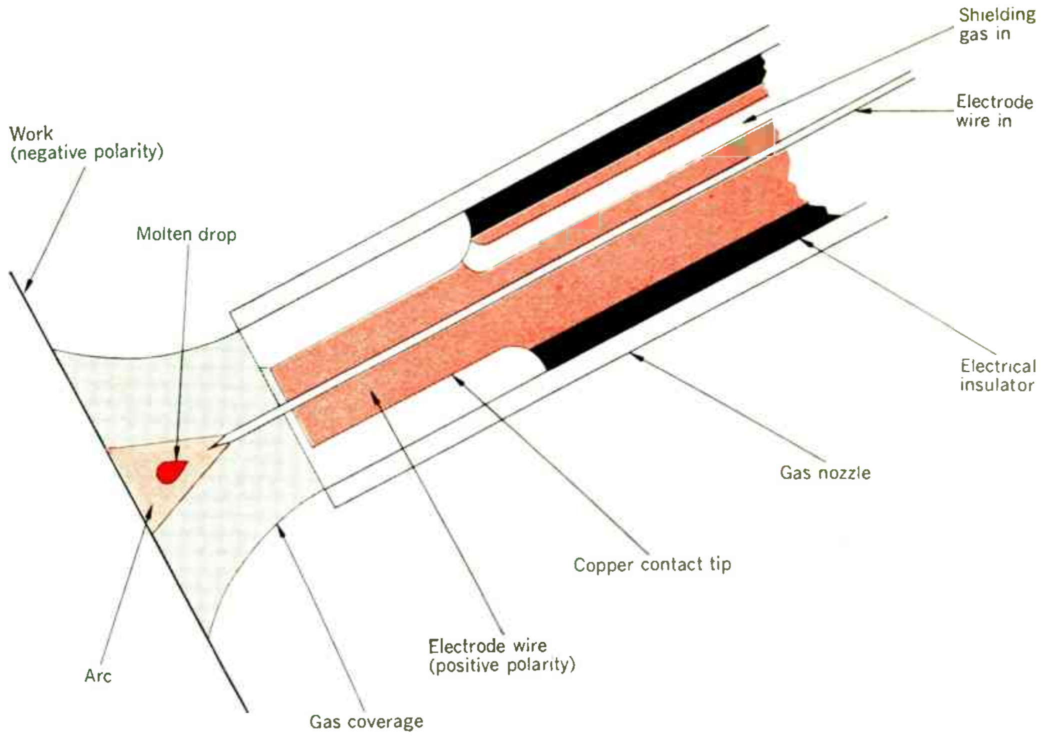


FIGURE 2. Cross section of welding gun.

Basic Mig process

The welding arc between the work and the continuously fed electrode wire provides most of the heat, and the rest is due to resistance heating of the "stick-out" portion of the electrode between the arc and the current contact tube. Currents, which normally range between 100 and 450 amperes, are proportional to the electrode wire-feed rate and diameter. Common wire diameters run from 0.8 mm to 2.3 mm, and are fed at rates from 2.5 to 11.5 meters per minute.

Several pure gases and mixtures may be employed to shield the arc and weld metal from atmospheric contamination, but the principal gas used is argon, with the addition of one or more of oxygen, carbon dioxide, or helium. The electrode is the positive arc terminal; and there are two modes of metal transfer from the electrode to the work (Fig. 3).

The transfer modes may be demonstrated by starting the arc at low current (low wire-feed speed) in inert gas, and noting the large globules of molten metal that form on the end of the electrode. These globules may grow to four or more times the electrode diameter before they are pulled off by gravity. As they grow, they wobble around and disturb the arc plasma, so that it moves around the work. The heated zone in the work is shallow. The weld metal deposit is irregular, and the drops may grow large enough to cause a short circuit, with attendant blasts of liquid metal.

With an increase in wire feed, the current increases, and at a specific amperage there is a sudden reduction in the size of the drops, and an increase in the number of drops, the plasma cone size, and the penetration of heat into the work due to the increased current density on the end of the electrode. This high-current-density condition, or spray-transfer process, has many welding applications; and the point at which the change in transfer mode occurs is called the transition current. For each combination of wire size, type, stick-out, and shield gas, there is a specific transition current. (For instance, for 1.6-mm-diameter aluminum electrodes the change occurs at 175 amperes in argon gas.)

Metal transfer by spray arc is desirable because the drops are small enough to pass through the plasma without disturbing its shape, and the electromagnetic force on the drops is greater than gravity, causing the drops to move along the axis of wire from the molten electrode tip to the work.[1]

The pulsed power process

The pulsed power welding process may be used with electrodes and shielding gases that normally operate only in the high-current-density or spray-transfer region. This implies an inert, or predominantly inert, gas, and an arc free of short circuits. The transfer of metal from the electrode to the work is restricted to intervals of high current interposed between intervals of lower current.

Normally, to obtain proper transfer conditions, power levels may be too great to allow out-of-position welding. In pulsed power welding, the power level is between the spray-transfer and dip-transfer regions, and fills the operational gap between them. It consists of a low background current, with superimposed, high-current pulses.

In the dip-transfer process, metal is transferred by contact between the molten electrode and the work, and the transfer takes place during a short circuit, with a closely controlled rate of current rise. The short-circuiting intervals are separated by arcing intervals, and occur randomly from 100 to 200 times per second. It is necessary to use carbon dioxide gas in this process, which may be objectionable when welding some metals. The average power level may be too low for heavy work and result in poor fusion.

The combination of two distinct current levels in pulsed welding results in lower average current, electrode melt-off rate, and heat input to the work. A study of the waveshape of the pulse with respect to time has shown that square, triangular, and half-sine waves all result in metal transfer during the pulse—and, of course, the half-sine wave is the easiest and least expensive to produce. The principle of the process is to pulse the arc long enough to transfer a given number of drops and then to reduce the current to a lower level to allow the weld to cool.

The pulse current is adjusted to nudge into the spray current range in order to force a fixed number of drops off, normally a single drop. For a given electrode size and type, the pulse current height is fixed; it is proportional to the peak voltage of the sine wave.

If the lower level were to persist for one-half second or more, the arc would revert to a large globular transfer, and would be undesirable. The few milliseconds' pause between half-wave rectifier pulses meets the requirements of pulse duration and repetition rate. The lower level is called "background" current. The two levels of current and their pulse repetition rates have been studied in many combinations,[2] and it has been found that 60 pulses per second works for a majority of welding conditions.[3] This is fortunate, because it is available from the power line. The lower 50-Hz power in some countries may also be used for many applications. Figure 4 shows the relation between pulsed current and transfer.

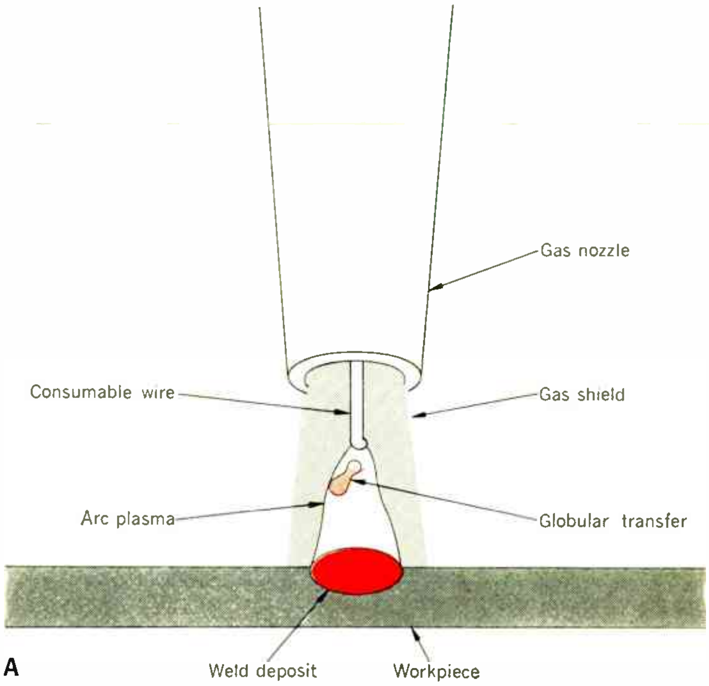
The power supply circuit

The background level and pulsed level of current are provided by two separate supplies that are connected in parallel. Each of these supplies is of the constant-potential type, with only about two or three volts droop per 100 amperes. Each supply is made variable by the use of sliding-tap transformers, and variations as small as one volt are available.

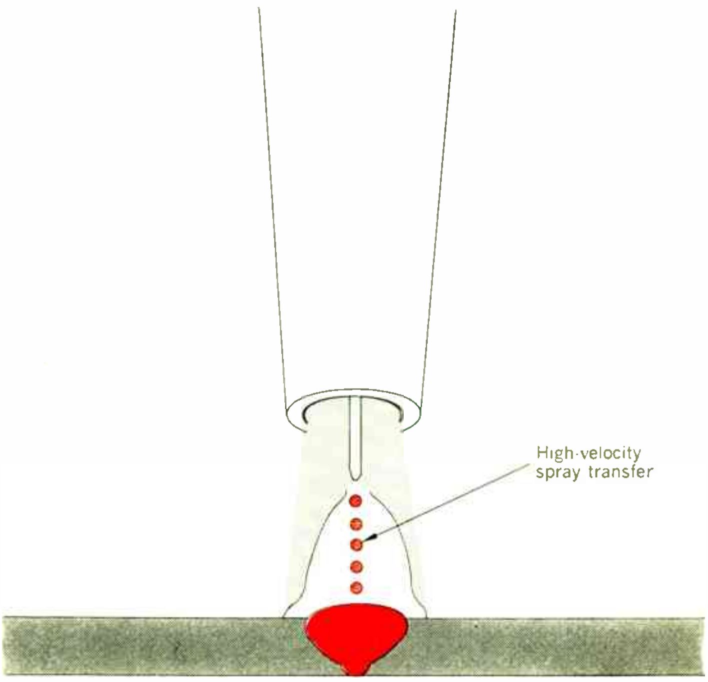
The source providing the background current is a three-phase, full-wave rectifier that utilizes silicon diodes. The output of this supply is filtered with a small smoothing choke that reduces ripple to a low level. A voltage variation from the supply of from 7 to 37 volts dc is possible. This circuit was chosen to provide a fairly pure direct current for the background interval. The diode rectifiers used in this supply are protected by small capacitors to suppress voltage surges in the circuit. The sliding tap on the transformer secondary is mechanically connected to a recording drum that is calibrated in volts.

The pulse voltage is generated by a single-phase, half-wave rectifier system fed from a transformer with a sliding tap. One-volt variations are possible with this transformer. The diode is also protected by a small capacitor for surge protection. Both transformer primaries have several voltage options to accommodate most common commercial supply voltages. The output of the pulse rectifier is tied in parallel with the output of the background rectifier, and their combined currents flow through a swinging choke to the welding electrode, which is positively poled. The workpiece is connected to the common negative terminals of both supplies. The output voltage available from the pulse supply is higher than that of the back-

FIGURE 3. Metal transfer as a function of current density. Low current density (A) produces a slow, irregular-shaped, globular metal transfer, with relatively low heat and erratic welds. High current density (B) produces a high-velocity spray transfer, clean regular weld, and high heat with corresponding free flow of molten metal.



B



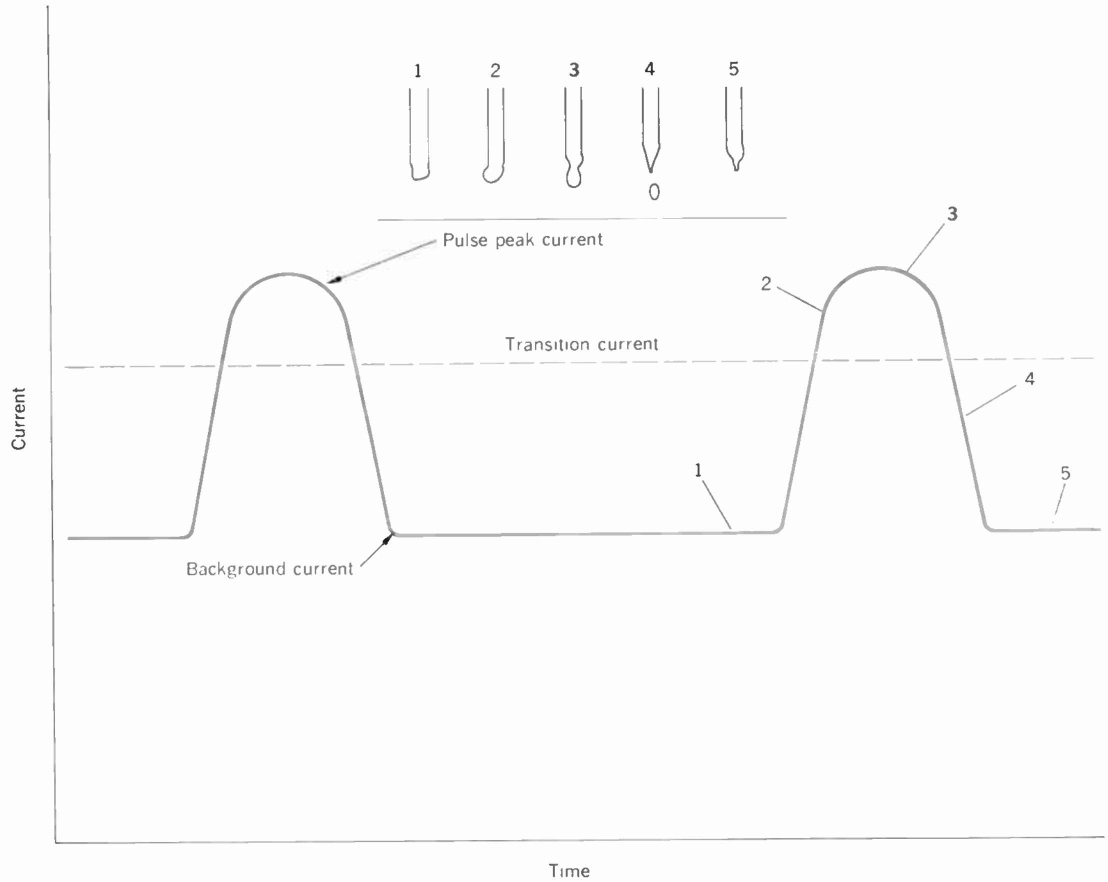


FIGURE 4. Current waveform and stages of metal transfer.

ground supply. Because of the parallel output connection and the blocking action of the diodes in each supply, the current flows from the supply that has the highest instantaneous voltage. Since the pulse voltage is available only during the positive cycles of the line voltage, pulses can occur only half of the time. The peak voltage available from the pulsing supply is from 30 to 60 volts, which will accommodate a range of electrode sizes and metals.

The swinging choke that carries the combined output currents is used to stabilize the arc during commutation. As large currents flow during the pulse, internal drops across the leakage reactance of the transformer cause momentary output voltages below that required by the arc at that instant. This causes the arc to go out, which results in unsatisfactory welding. Arc outages are prevented by discharge of the choke, which keeps the welding current flowing; but the inductance has to be minimal during the pulse to prevent large voltage drops across it. The two requirements of inductance are met by using a swinging choke that saturates at the proper current level.

Due to the nature of this power supply, there are unique problems of metering. Each electrode size and type needs a particular peak current to transfer the metal properly. This current is proportional to the peak voltage available and must be preset before the welding is started. The panel meter is an average-reading meter, but may be converted to a peak-reading meter by pressing a button that connects a diode and capacitor, so that peak value on the sine wave is read. The ammeter is conventional, and reads the voltage drop across the shunt carrying the total current. A popular clamp-on ammeter used with the pulsing waveform will show a different value than the average-reading meter because it reads the rms value of the waveform. All process data have been recorded using average current values so that they may be compared with other processes.

The electrical input rating of the power supply was determined by operating both taps at maximum. A three-phase wattmeter showed 12 kW consumed in this test when a resistive load and 30 volts output were used. Because the three-phase supply and the single-phase half-wave supply combine to give an unbalanced load to the line, two currents were 28 amperes each, and the remaining line read 16 amperes. The design of the transformer that supplies the pulse circuit must consider the effect of a net direct current in the core, because the half-wave rectifier connection will produce saturation. This is accomplished by using the correct lamination material; also, an air gap may be necessary in the core.

Auxiliary power for wire feeders, water circulators, and similar equipment for the welding package is provided by a 1-kVA transformer at 115 volts.

REFERENCES

1. U.S. Patent 3 071 680.
2. Boughton, P., and Lucey, J. A., "The use of pulsed current to control metal transfer in welding," *Brit. Welding J.*, vol. 12, pp. 159–166, Apr. 1965.
3. Agnew, S. A., and Daggett, E. H., "Introduction to pulsed-spray arc welding," *Welding J.*, vol. 45, pp. 284–289, Apr. 1966.

Galvanomagnetic devices

Classic textbook phenomena since the days of Faraday, Hall and magnetoresistance effects might have remained the sole province of students, inventors, and theorists had it not been for modern solid-state technology

H. Weiss *Siemens Aktiengesellschaft*

Galvanomagnetic properties in the form of Hall and magnetoresistance effects have been rescued from relative obscurity by the successful emergence of sophisticated semiconductor materials and techniques. Long a part of the standard curriculum, these phenomena have unleashed a plethora of useful electromagnetic devices by the adroit application of relatively simple principles. Although solid-state research has certainly been enhanced by the electron-mobility and band-structure analysis capability of galvanomagnetic effects, who would have thought that potentiometers, choppers, gyrators, and push buttons could find new substance with long-established phenomena?

Galvanomagnetic effects are solid-state phenomena that have been well known to investigators for over a century. Until the advent of recent breakthroughs in solid-state research, however, utilization of these effects was restricted to the classic textbook references (unless one considers the bismuth spiral for measuring strong magnetic fields). This is not to imply that there has existed a lack of ideas for technically applying these phenomena—a notion fully belied by the mass of scientific, technical, and patent literature that has accumulated through the years. Rather, history has repeatedly demonstrated that the interval from the discovery of a phenomenon to its eventual application within a practical device may be quite long, indeed. A phenomenon may be of outstanding scientific interest, but unless its characteristic values—e.g., resistance, current and voltage capacity, temperature dependence, size, cost—can be implemented by available techniques, its relative advantages cannot be readily marketed. In short, a maturity must be developed between effect and technique. Hence, the vital prerequisites for successful application of galvanomagnetic devices were contained in the technologies realized by the development of the transistor, contactless signal generation, and present-day high-quality magnetic materials.

The magnetoresistance effect was discovered by W. Thomson in 1856, and the Hall effect by E. H. Hall in 1879. These two galvanomagnetic phenomena have enabled solid-state researchers in recent years to measure the concentration and mobility of charge carriers, and to analyze band structures in solids. The first report concerning the use of the Hall effect (in germanium) in an instrument used to measure magnetic fields was published by G. L. Pearson in 1948.[1] Extensive technical use of the Hall generator began after 1952, when the first results concerning high electron mobility in indium antimonide (InSb) were published by Welker *et al.*[2-4] More recently, the discovery of the magnetoresistance effect in more tenable forms has precipitated a growing number of technical applications.[5,6]

The mode of operation of galvanomagnetic devices is relatively easier to understand than that of transistors or rectifiers, since one is essentially dealing with a homogeneous semiconductor. With heterogeneous devices, lifetime, surface, and diffusion effects play a decisive role in overall operation; whereas these effects can be totally disregarded in our present study.

The Hall effect

According to the equation for the Hall voltage U_H,

$$U_H = \frac{R_H}{d} i_1 B \qquad (1)$$

where

R_H = Hall coefficient, $m^3/A \cdot s$
i_1 = control current, A
d = thickness of semiconductor, m
B = magnetic induction, $T = V \cdot s/m^2$

with 42 distribution points (Fig. 4). The drum, with a diameter of 1 meter, has 16 magnetic bands. At the perimeter of the storage device, there are located 44 adjustable perpendicular rods with Hall generators attached. The speed is 0.0043 m/s at the perimeter; and the appropriate combination of Hall generators reacts at 44 positions of the drum. Such an assembly for controlling a circular conveyor seems eminently suitable for automatic assembly lines, such as those employed in automobile factories.

The magnetoresistance effect

For many types of applications, the internal resistance of a Hall generator is much too small, and the consequent high control current and low Hall voltage necessitate additional amplifiers. In these instances, however, the magnetoresistance effect can be employed successfully. In contrast to the Hall generator, this semiconductor is divided into long, narrow strips such that the resistance per unit area may be varied over three orders of magnitude. The quantity $\Delta R/\Delta B$, which corresponds to U_H/i_1B for Hall generators, can be adapted to several kΩ/dT without technological difficulties; that is, several orders of magnitude more than is shown in Table II. Thus, high signal voltages can be obtained with small currents. The exploitation of magnetoresistance in common electronic circuits, especially when transistors are involved, requires resistances of the order of 100 ohms. An additional advantage of the magnetoresistance device makes use of the fact that only two leads are necessary; therefore, the circuitry is similar to that of a standard fixed resistance, or a variable resistance with sliding contacts.

It should be noted that the magnetoresistance effect in homogeneous InSb is much too small for any worthwhile application. The resistance increases by only 55 percent in a magnetic field of 1.0 T.[11] In order to obtain a device with a high magnetoresistive effect, it is necessary to short-circuit the Hall voltage by suitably shaping the semiconductor and electrodes. The equipotential lines, with and wihout a magnetic field, must display the same position; and the current lines must be rotated through the Hall angle ϑ.

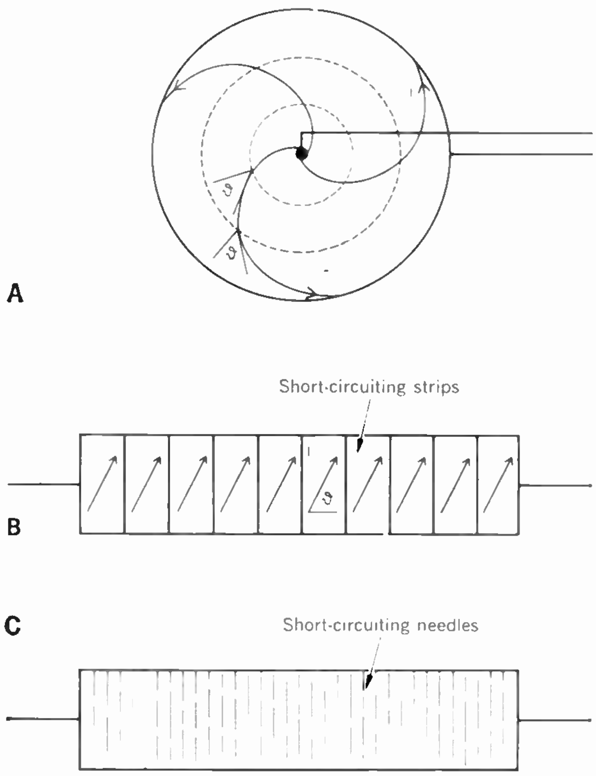


FIGURE 5. Structures with a high magnetoresistance effect in InSb. A—Disk. B—Raster plate. C—InSb-NiSb eutectic.

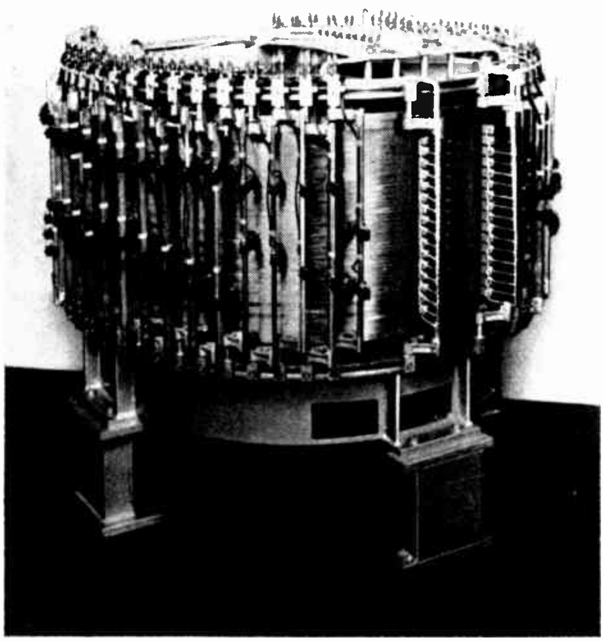
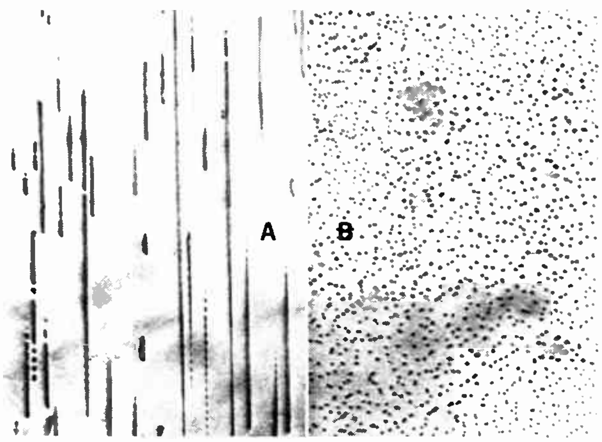


FIGURE 4. Drum storage with magnetic tapes and vertical rods for the mounting of Hall generators.

FIGURE 6. Polished surfaces of InSb-NiSb eutectic (200 ×). A—Surface parallel to NiSb needles. B—Surface perpendicular to NiSb needles.



The simplest form that satisfies these conditions is a disk with a hole in its center [Fig. 5(A)]. The inner, as well as the outer, ring electrodes are made of a good conductor, such as silver, and are soldered directly to the semiconductor. Consequently, it has been found that, with slightly n-doped InSb, the resistance can be increased by a factor of 38 in a field of 1.0 T.

Examination of very thin samples (10 μm thick) have yielded zero-field resistances of up to 2 ohms. Therefore, a different form—the so-called raster plate—must be chosen for the field-dependent resistance. It consists, in principle, of a long, narrow semiconductor plate fastened to an insulating substrate. This plate, as described in Fig. 5(B), is covered with short-circuiting metal strips, which lie transverse to the longitudinal direction.[12]

The result is not a complete suppression of the Hall voltage, as it is in the case of the disk, but an extremely large reduction, such that the lines of current flow, corresponding to the arrows in Fig. 5(B), are rotated by the magnetic field by nearly the Hall angle ϑ. In this type of construction, it is possible to vary the resistance within wide limits by changes in the length, width, and thickness of the plates, as well as by changes in their number. As a result, a resistance ratio of 15 in a field of 1.0 T can be obtained with n-doped InSb.

Apart from the use of suitable forms of homogeneous indium antimonide for producing a high magnetoresistance effect, an alternate method has been devised for obtaining the same result. Instead of applying short-circuiting strips to the specimen, a crystal of InSb can be fabricated that incorporates a parallel arrangement of short-circuiting NiSb needles[13] within its interior [Fig. 5(C)]. Figure 6 illustrates a specimen of indium antimonide–nickel antimonide eutectic at a magnification of 200. The specimen actually consists of InSb with 1.8 percent by weight of NiSb. Figure 6(A) shows a sectional view of the crystal which is parallel to the short-circuiting needle-shaped inclusions; Fig. 6(B) shows a view perpendicular to that plane. The needles have a mean length of 50 μm, and diameter of less than 1 μm. The specific conductivity of the nickel antimonide amounts to approximately 7×10^6 mho/m. It is, therefore, larger than the conductivity of intrinsic or slightly n-doped indium antimonide by at least two orders of magnitude. Moreover, if a semiconductor plate is cut from the two-phase material, with the position of the needles as represented in Fig. 5(C), and the plate is inserted in a magnetic field that is perpendicular to the plane of the diagram, the result is the analog of the raster plate described in Fig. 5(B).

Since nickel antimonide crystallizes into the nickel arsenide lattice, it exhibits a tendency to form long, thin needles within InSb, as shown in the micrograph (Fig. 6) of the eutectic. Over a large region, the needles display an arbitrary orientation. An investigation of the solidification process, however, has disclosed that the needles grow in a direction that is approximately perpendicular to the solid–liquid phase boundary. Figure 6(A) duplicates a micrograph of a lapped and polished section that was cut parallel to the crystal surface area—crystal growth started at the bottom and continued upward. Figure 6(B) shows a section that is normal to the growth direction. Electron-optical and X-ray investigations have shown that these needles are single crystals whose long axes coincide with the c direction of the NiSb lattice. Since the solidus and liquidus curves coincide at the eutectic point, neither the process of normal solidification nor repeated zone melting has any significant effect upon the uniform distribution of the needles. Hence, it is quite feasible to grow blocks 30 cm in length and 2 cm in width that have nearly rectangular cross sections.

The impurity content of InSb may be determined from measurements of the specific conductivity and of the Hall coefficient made at or below liquid-air temperature. Such measurements indicate that the addition of NiSb

FIGURE 7. Relative resistance R_B/R_0 of field plates with different n-doping as a function of magnetic induction B.

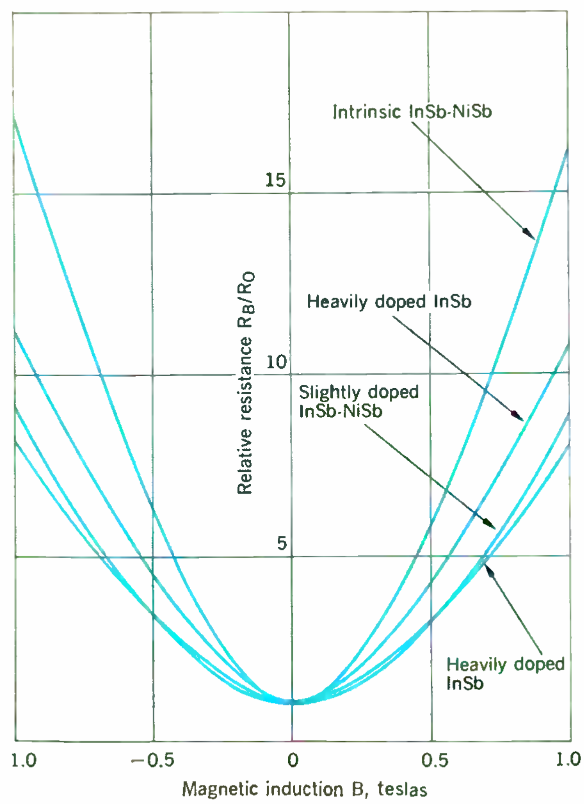
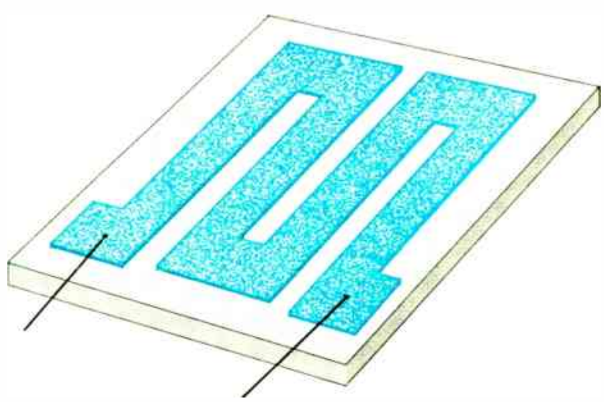


FIGURE 8. Structure of a field plate. Semiconductor is mounted on an insulating substrate.



to InSb in the eutectic composition at 1.8 percent by weight introduces less than 10^{15} electrons per cm^3. The semiconductor matrix of InSb that contains the needles remains, for all practical purposes, unchanged. This represents a decided advantage to the investigator, since it is possible to dope indium antimonide independently, without considering the presence of the second phase, thus controlling the temperature dependence of the electrical properties, their magnitudes, and their reproducibility. To obtain n-doped material, tellurium is added to the pure indium antimonide melt in the same manner nickel antimonide was added. With this technique, temperature dependences as small as 10^{-3} per °K are possible. By using the magnetoresistance device as a base-emitter resistance, one can adjust the temperature coefficient to that of a germanium or silicon transistor by suitable doping.

Figure 7 demonstrates that both methods—surface metallic strips and built-in needles—provide quite a good short circuit and, therefore, a large increase in resistance. The relative resistance R_H/R_0 of field plates is given as a function of the magnetic induction. Curves are presented that correspond to intrinsic and slightly doped InSb-NiSb, and to heavily doped InSb. Note that, with intrinsic InSb, the resistance increases by a factor of 17 in a field of 1.0 T.

The field plate, the magnetoresistance device schematically displayed in Fig. 8, is fabricated in the following manner. An InSb-NiSb plate, approximately 16.5 mm by 18.5 mm in size, is cemented to a ceramic plate of similar dimensions, and is ground to a required thickness of 20 μm. By means of the photographic mask detailed in Fig. 9(A), a light-sensitive lacquer, and a suitable etchant, the shapes of 44 individual resistors are etched in relief [Fig. 9(B)]. Thereafter, the substrate is cut or broken into the individual devices, and two leads are soldered onto each element [Fig. 9(C)]. By this technique, one is able to attain zero-field resistances of up to 200 Ω/mm^2.

What are the problems connected with the production of the two-phase eutectic InSb-NiSb in large quantities at low cost? In this area, the work of A. Müller and M. Wilhelm has been invaluable and deserves mention.

Investigations have uncovered a major irregularity that occurs during solidification of the eutectic (Fig. 10). It has been found that the needles arrange themselves in striped arrays with empty regions between them.[14] A similar picture involving doping striations may also be obtained for Te-doped InSb. These variations of doping or of the distribution coefficient of Te in InSb are normally caused by a periodic variation in the speed of crystallization during the solidification process. The reason is due to a temperature oscillation in the semiconductor melt at a temperature gradient.[15] To avoid the striations, the temperature gradient near the solid-liquid interface must be less than 2 degrees per cm.

Small temperature gradients are not the sole cause of this parallel alignment of needles. By normal freezing of the InSb-NiSb melt, one may obtain a polycrystalline rod with grains that are elongated primarily in the growth direction (Fig. 11). Specifically, the orientation of 100 needles relative to the pulling direction was measured for several positions along a rod. Figure 12 plots the distributions obtained. It was discovered that the c axes of neighboring needles exhibited angles of up to 30 degrees,

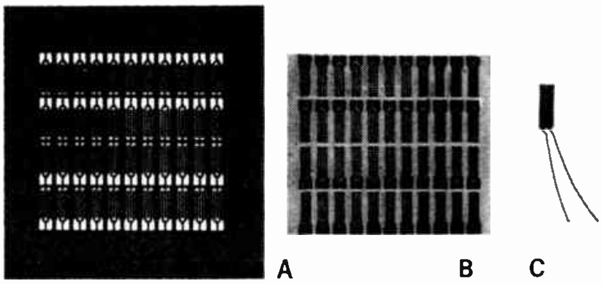


FIGURE 9. Fabrication of field plates (1.5 mm × 4 mm). A—Photographic mask for 44 devices. B—44 etched resistors. C—Individual field plate.

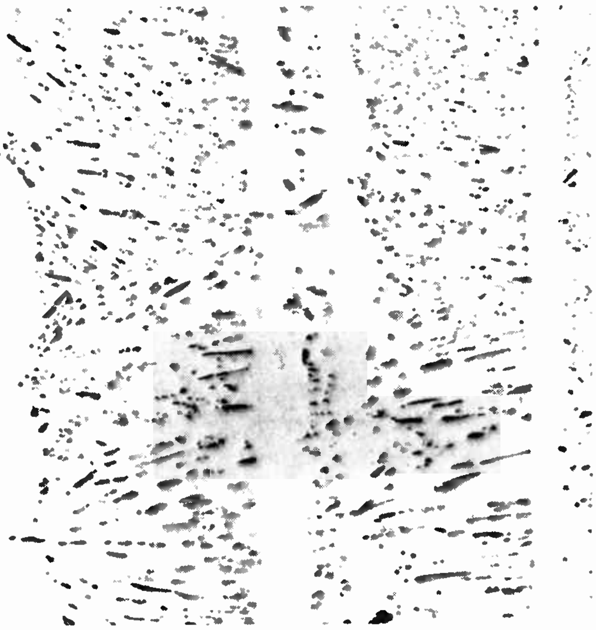
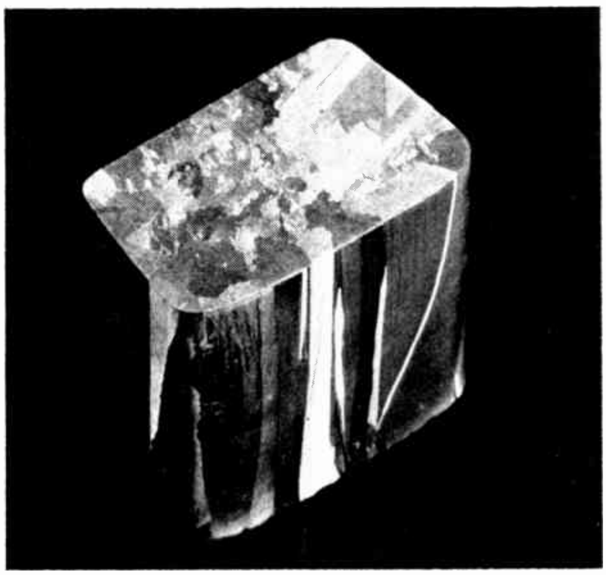


FIGURE 10. Striations in a two-phase InSb-NiSb eutectic. The growth direction is from left to right.

FIGURE 11. Polycrystalline rod of InSb-NiSb. This cross-sectional area is 2.5 cm × 2 cm.



with the deviation from the growth direction amounting to as much as 20 degrees. An interesting observation embodies the fact that the various maxima of Fig. 12 do not point to the same direction. There is no direct correlation between the direction of the needles and the orientation of the crystallites.

However, decisive progress has been made by the discovery of InSb-NiSb with a fiber matrix.[16] This is a polycrystalline rod in which the monocrystalline grains are elongated fibers that are parallel to each other and to the growth direction of the crystal (Fig. 13). The individual fibers have their axes oriented in the ⟨110⟩-direction, and are rotated at random. Figure 14 reproduces the micrograph of a fiber crystal. The NiSb needles are readily seen within the parallel grains of different brightness. In such a crystal the axes of the needles point to almost the same direction—80 percent do not deviate from the growth direction by more than 5 degrees. The fibers and the needles seem to mutually stabilize their parallel growth with a high yield. Pure InSb without NiSb cannot crystallize in this form.

With the aid of the newly developed two-phase materials, several new devices and components have been developed for contactless control techniques. For this particular purpose, it was necessary to design suitable magnetic systems.

The most obvious application is contained in a variable resistance with no sliding contact. Figure 15(A) demonstrates the principle. The right of this illustration com-

FIGURE 12. Relative distributions of needle orientations for different positions along a rod of InSb-NiSb of the type shown in Fig. 11. The growth direction is zero degrees.

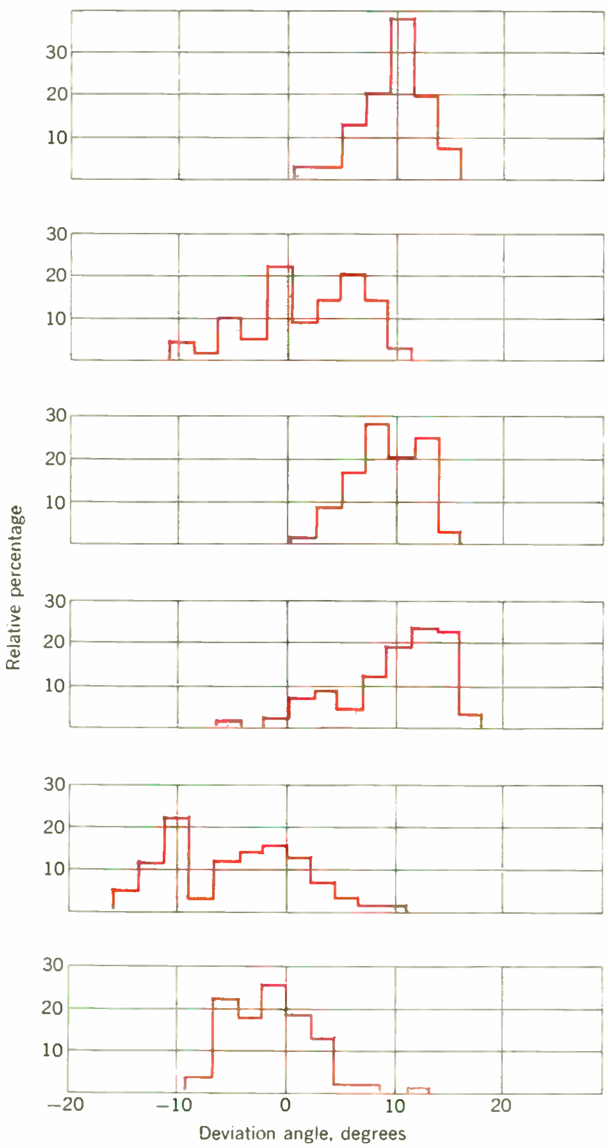
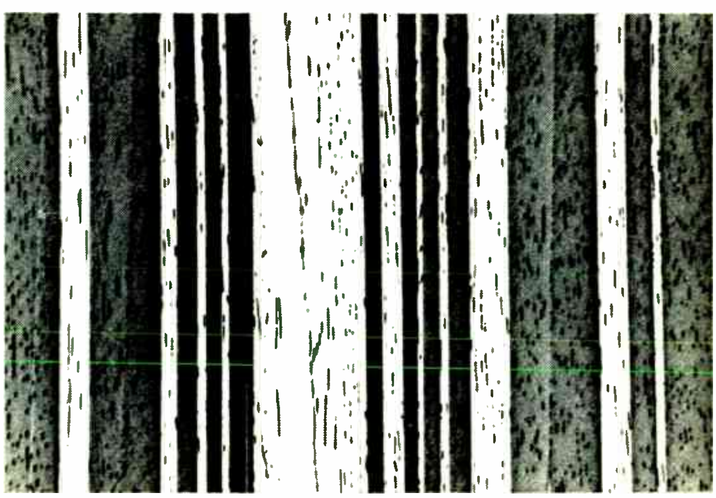


FIGURE 13. Fiber crystal of InSb-NiSb with the single thin elongated crystals oriented in the ⟨110⟩-direction.

FIGURE 14. Polished surface of a fiber crystal corresponding to Fig. 13. The surface is parallel to the InSb fibers and to the intruded black NiSb needles.



prises a hairpin-shaped semiconductor with two leads attached. The colored rectangle on the left represents the magnetic flux of a permanent magnet acting upon the field plate; its effective surface area is determined by the shape of the magnet pole caps. If the magnetic flux is displaced from left to right, then the resistance increases as a larger portion of the semiconductor is contained within the flux. The limit of this increase is reached when the flux arrives at the terminals of the field plate, i.e., $b = l$. If the semiconductor is homogeneous, and if the two legs of the field plate are constant in width and thickness, then, for a displacement Δb, an additional segment of the field-dependent resistance (proportional to Δb) is introduced into the magnetic field. Furthermore, if the magnetic induction of the driving flux is homogeneous, such that the portions already covered do not change their resistance, then the resistance increases linearly with b. Figure 15(B) denotes the relative dependence of the resistance upon the displacement b, according to the length l of the field plate, for $R_B/R_0 = 9$.

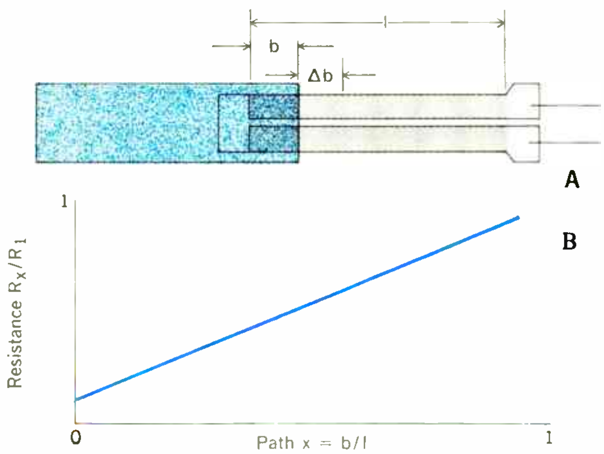
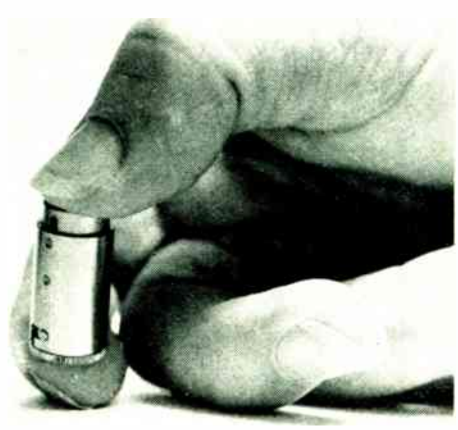


FIGURE 15. Contactless variable resistance device. A—Fundamental configuration of the field plate and movable magnetic flux. B—Relative resistance R_x/R_1 as a function of the relative path $x = b/l$.

FIGURE 16. Field-plate push button.



A practical application of this device can be seen in the construction of a noncontacting indicator for monitoring the nominal current acting on the brake of a locomotive used by the German Federal Railway.[17] For this particular usage, a field plate with a small temperature dependence is controlled by a permanent magnet, which can rotate about its own axis. The controlled resistance varies linearly with the angle of rotation over a range of ratio 1 to 7.

The internal structure of a field-plate potentiometer offers another type of application. The magnetic circuit consists of two parallelly oriented permanent magnets, a common air gap, and a field plate placed within the air gap below the upper pole cap. A highly permeable, shaft-mounted screw completes the circuit. Hence, the magnetic flux envelops the semiconductor as the screw-operated shaft is turned—a linearity error comprising a fraction of a percent over an angle of 270 degrees being obtainable. By an appropriate configuration of the screw on the shaft, any dependence of resistance upon angle is feasible.

A field-plate push button, proportionally illustrated in Fig. 16, features a small permanent magnet that is attached directly to the button. If it is pressed into the cylinder, it traverses a field plate whose resistance increases over a ratio of 1 to 10 with a displacement of less than 1 mm. This device is usually applied in cases where reliable nonbouncing switches are required. Other applications of magnetoresistance include a highly sensitive field-measuring device with great temperature stability,[6,18] a contactless chopper,[6,19] and the brushless motor,[6] which has been used in tape recorders since 1966.

Galvanomagnetic effects have been so successfully adapted to practical devices that there are virtually unlimited prospects for the development of new and interesting applications. These innovating concepts have certainly already enriched the spectrum of the semiconductor market.

REFERENCES

1. Pearson, G. L., "A magnetic field strength meter employing the Hall effect in germanium," *Rev. Sci. Instr.*, vol. 19, pp. 263–265, Apr. 1948.
2. Welker, H., *Z. Naturforsch.*, vol. 7a, p. 744, 1952.
3. Kuhrt, F., *Siemens-Z.*, vol. 28, p. 376, 1954.
4. Hartel, W., *Siemens-Z.*, vol. 28, p. 370, 1954.
5. Weiss, H., *ETZ*, vol. 17, p. 289, 1965.
6. Weiss, H., *Solid-State Electron.*, vol. 9, p. 443, 1966.
7. Kuhrt, F., *Solid-State Electron.*, vol. 9, p. 567, 1966.
8. Guenther, K. G., *Z. Naturforsch.*, vol. 8a, p. 248, 1953; vol. 13a, p. 1081, 1958.
9. Wieder, H. H., *Solid-State Electron.*, vol. 9, p. 373, 1966.
10. Kuhrt, F., Stark, G., and Wolf, F., *Elektron. Rundschau*, vol. 13, p. 407, 1959.
11. Weiss, H., "Galvanomagnetic properties of InSb," *J. Appl. Phys.*, vol. 32, pp. 2064–2068, Oct. 1961.
12. Weiss, H., and Welker, H., *ETZ*, vol. 76, p. 513, 1955.
13. Weiss, H., and Wilhelm, M., *Z. Physik.*, vol. 176, p. 399, 1963.
14. Müller, A., and Wilhelm, M., *Z. Naturforsch.*, vol. 20a, p. 1190, 1965.
15. Müller, A., and Wilhelm, M., *Z. Naturforsch.*, vol. 19a, p. 254, 1964.
16. Müller, A., and Wilhelm, M., *Z. Naturforsch.*, vol. 22a, p. 264, 1967.
17. Lössel, W., *Siemens-Z.*, vol. 39, p. 600, 1965.
18. Von Borcke, U., Martens, H., and Weiss, H., *Solid-State Electron.*, vol. 8, pp. 365–373, Apr. 1965.
19. Hieronymus, H., and Weiss, H., *Solid-State Electron.*, vol. 6, pp. 463–476, Sept.–Oct., 1963.

Education for innovation

Practical, creative engineers are desperately needed to solve the complex problems of modern society; yet there is a proliferation of walking formula indexes issuing from our colleges. Perhaps we are educating innovation into oblivion

Daniel V. De Simone *U.S. Department of Commerce*

Engineering education should encourage students to strive for a mastery of fundamentals and the cultivation of excellence—but this is not enough. Engineering education must be kept alive and relevant; and, to encourage creativity, it must stimulate the imaginations of students. The typical educational yardstick of a student's performance, however, is the accuracy with which he can repeat, by rote, information obtained from a lecture or text. Original or unconventional approaches to problems are discouraged, and their proponents often penalized, thus discouraging and depressing the student. Must inventiveness be sacrificed to education?

An engineering school that merely imparts information is an expensive waste. It was made an anachronism by Johann Gutenberg's invention of the printing press in the 15th century.[1] Education needs to be kept as alive as the intensive civilization in which we live, and as relevant as the concerns that linger after reading a newspaper.

Engineering education should encourage students to strive for the mastery of fundamentals, the discovery of the relatedness of things, and the cultivation of excellence. But it should also be a creative experience, stimulating the imagination of students and helping them to prepare themselves for the contests and the challenges of an imperfect world. It should encourage them to believe they can do the "impossible" from time to time, even if it means doing violence to precedent. It should ensure, through them, the continued advance and renewal of our society—without sacrificing human values.

These are the aims of creative engineering education; that is, education for innovation. Taken together, they constitute a formidable set of goals.

Noble illusions

Engineering is a profession, an art of action and synthesis, and not simply a body of knowledge. Its highest calling is to invent and innovate. Shaping the professional attitudes and skills of design needed in the practice of this art is therefore difficult in an engineering educational system that, with few exceptions, stresses the acquisition of knowledge and skills of analysis to the virtual exclusion of all else. In the absence of deliberate counter measures, it is an educational system that stifles creativity.

Consider the typical measure of a student's performance; it is often enough to stamp out whatever inventive and innovative qualities he may have possessed when he entered school. By and large, his performance is measured by the fidelity with which he feeds back the information he has absorbed from lectures and texts. Perfect feedback occurs when he gives the right answers to problems that have been assembled for him. These problems are solved through analysis based on unassailable principles of science and engineering.

Creative individuals are oppressed by this regime; and the real world of invention and innovation is foreign to it. Although in school one must never fail, an inventor fails all the time, and is elated in those rare instances when he succeeds. "An inventor," Charles Kettering once remarked, "is simply a fellow who doesn't take his education too seriously." Indeed, Marshall McLuhan maintains, with much persuasiveness, that going to school interrupts the education of students. The outside world, McLuhan observes, is far richer in information than is the schoolroom.

Seldom are students encouraged to see and experience the relevance of their studies to the "outside" world. Yet it will soon be theirs to change, for better or worse.

The extension of man's capabilities

Variations of the terms "engineering," "invention," "innovation," and "creativity" have already been used, and they will be encountered many more times. It would be useful to try to compose some reasonably serviceable ground rules for these expressions, or at least to note the various ways in which they are employed by different authors. But the inquiry need not be abandoned just because our definitions are imprecise.

Creativity has become a fashionable term, not because of its legitimate uses, but because of its wanton employment by sloganeers. It is exciting to be creative. Thus comes "creative living" if one uses "ABC Mouthwash" or moves into the "Cubicle Towers." The result is that when serious commentators speak of creativity, the lack of response may be due to an overdose of such advertisements.

the world or the world to itself.[5] We are concerned here with technological innovation; but we are primarily interested in innovations in education for innovation.

The consequences of change

Technological invention and innovation are the business of engineering. The consequences of technological change are therefore properly a concern of engineering; certainly, they should be studied in the engineering schools.

Innovation has been essential to the process by which the United States has grown and renewed itself; there is a very significant relationship between it and economic growth. Although estimates of the contribution of technological progress to increases in the Gross National Product (GNP) are imprecise, all economists seem to agree that the contribution is substantial.[6–8] For example, if we compare the change in the labor input with the change in GNP over the period 1947–65, we note a marked difference between these two factors. The average annual hours of work remained practically constant, whereas the GNP nearly doubled. Without presuming to

Creativity threatens change and disruption, which is why creative ideas are less well received than are creative appellations. To be creative has been characterized as having the quality or power of producing "effective surprise."[2] The dictionary defines it as the power or quality of causing "to come into being, as something unique that would not naturally evolve or that is not made by ordinary processes." Donald MacKinnon says that creativity involves "a response that is novel or at least statistically infrequent, . . . [and] to some extent . . . adaptive to reality"[3] and Myron Coler observes that creativity concerns "something which is new, rather unexpected, and nontrivial."[4] We might sum these observations on the meaning of creativity and say that creativity involves effective surprise—something novel, unexpected, and nontrivial.

Engineering is epitomized in the processes of invention and innovation. As we use these terms, invention is the process of conception. Innovation, which may or may not include invention, is the complex process of introducing ideas into use or practice, and includes the vital element of entrepreneurship. Thus, society benefits from innovation, not from invention alone, and often there is a long lapse between the two. Heron, for example, discovered the principle of the steam turbine engine two millennia before its introduction.

Through invention and innovation, ideas are conceived and introduced into the economy as new products and processes; or into an organization to change its way of doing things; or into a society to change its ways of thinking or to provide for its social needs and to adapt itself to say how much of this increase in GNP was attributable to technological innovation as opposed to other factors such as education, we can say confidently that it played a major role.

The most dramatic consequence of technological change is evidenced in man himself.[9] It has been noted that 20th century man has precisely the same brain and body as his ancestors of some 20 000 years ago. The significance of this is that technological innovation, by extending man's capabilities, has obviated the need for his evolution and has, instead, changed the world about him. Under the seas and in space, it is *technology* that sustains him, not his outmoded physical attributes.

Human values

Not all of the consequences of technological change are cause for rejoicing. Many of them, such as environmental pollution, are deleterious and avoidable, and call for social and political innovation. But others, such as temporary dislocations of employees, firms, and industries, are concomitants of change, which is inherently disruptive—although these, too, require innovative approaches, so that they can be anticipated and ameliorated. Still other objections to technology are spiritual and philosophical.[10]

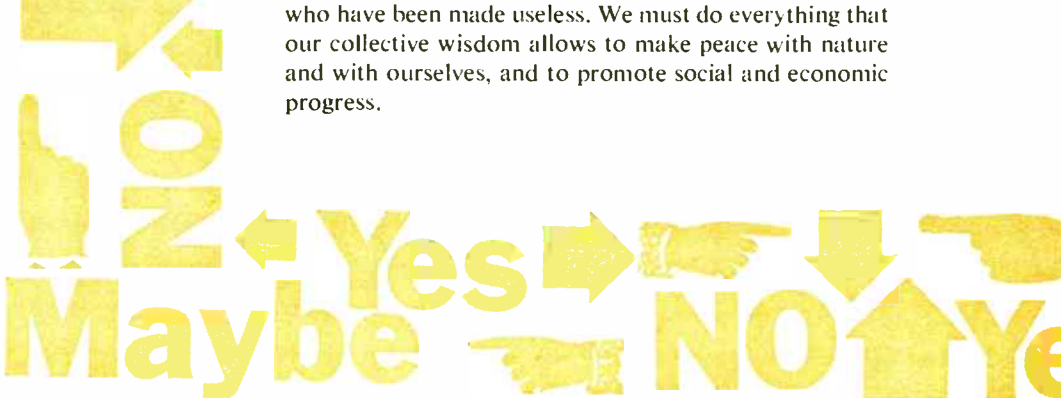
Mahatma Gandhi resented the subjugation of India by the West and wanted to sever the ties that had been established through India's adoption of Western technology. He advocated an abandonment of this technology and a return to the unblemished past. But much as the people of India revered him, they were unwilling to follow his lead on this score. They would not endorse a policy of technological regression.[11]

The lesson for engineering students is that it is unlikely that any society will ever deliberately arrest its technological advance. Therefore, it is the obligation of engineering education to be concerned at least as much with the quality of technological change as with its quantity.

It is an error, of course, to hold that the disagreeable aspects of society are due to science and technology. The fault lies in the manipulators of change, in an attitude of mind that sacrifices human values for other objectives. Engineering students should know this, for they will be the future instruments of technological change.[12]

Not by science alone

Society has properly extolled the achievements and potential of science, which aims at the extension of man's knowledge. In the process, however, the role of engineering has been underplayed and relatively disregarded in, of all fields, engineering education. And this is truly unfortunate, for engineering is the extension of man's capabilities—no less noble an object than the extension of his



knowledge. We need first-rate scientists; we also need first-rate engineers.

Creative design courses are available at only a handful of U.S. engineering schools. Moreover, very few Ph.D. theses in engineering have anything at at all to do with creative engineering design. The great majority of these theses involve abstract, science-oriented dissertations. Again, to cite this fact is not to detract from the importance of science. Rather, it is to voice legitimate concern over the philosophy and direction of engineering education. Although it is quite acceptable for Ph.D. candidates in the sciences to profess no interest in the application of new knowledge to the solution of practical problems, it is disconcerting when most Ph.D. candidates in engineering display similar attitudes in their graduate efforts. If engineers, like scientists, are educated in an environment that voices disdain for the practical application of science, then to whom will society turn to solve its problems?

The needs of society

The first and most important aspect of an engineer's education, J. Herbert Hollomon has observed, is to become aware of the character and needs of the society in which he will live. The requirements of engineering cannot be determined intelligently unless the needs of society are first perceived. What kind of a society do we have in the United States? Consider, first, the increasing concentration of human beings in complex metropolitan environments, which, by the turn of this century, will more than double in population. In the process our society is becoming more interdependent and, in the not-too-distant future, will likely become one faced with social dilemmas of leisure and affluence. In large part due to technological prowess, it has become the preponderant world power, with all the profound consequences and obligations that stem from this ascendancy. Gradually we are passing out of the state of being an industrial civilization and becoming a society whose social, economic, and political life is shaped not so much by the industrial mechanism as by the technologies of communication and information processing. We are at the center of an electrically contracted "global village."[13]

We are a society that cannot much longer condone the outrages we commit against ourselves. We have but a little way to go, as history is measured, before the wanton destruction of our environment reaches disastrous proportions. Most of our rivers are cesspools. Flying high over many of our large cities, one sees them straddled by an overhanging sewer of noxious fumes and contaminants. And, as just one illustration of what we do to the land, the ravagement of strip mining is not truly believable until one happens upon the destitute remains and imagines himself on the moon's surface. We also can no longer ignore the tens of thousands of fellow human beings who are being slaughtered and maimed on our highways every year. The scales of opportunity must be balanced for those who have been made useless. We must do everything that our collective wisdom allows to make peace with nature and with ourselves, and to promote social and economic progress.

These are some of the problems with which engineering must come to grips. Of course, they are not solely engineering problems; the crucial decisions to be made with respect to them are not engineering decisions at all—they are political. But the role of engineering in the solution of these problems is very important. It is an awareness of this role, and a challenge to do something about it, that must be made a part of the education of an engineer. This is the kind of ingredient that can make engineering education alive and relevant.

Where is the time to come from?

Charles Stark Draper, chairman of the National Inventors Council, in speaking from the rich experience of his impressions of creative engineering, sees creative enterprise as the capability of generating novel and effective means to achieve an important, socially relevant end. Among the characteristics of a creative engineer, he lists the ability to recognize key problems, bring relevant knowledge to bear upon them, conceive effective solutions, and carry these through the innovative process. Engineering education should nurture these traits; but it fails to do so in many respects, especially in the way it is structured. Draper sees this structure as an array of fragmented disciplines that developed into fiefdoms for narrow experts as knowledge grew, over the centuries, to prodigious dimensions. However, he finds this situation untenable in light of the possibilities afforded by the rapid processing capabilities of electronic digital computers. They are freeing us from the boundaries of formal disciplines, and permitting us to cut across them with facility. Computers promise to be the most profound educational tool ever devised by man.[14]

Masses of data, which until now necessarily consumed so much of an engineering student's time, can be handled by computers. Indeed, data that were previously beyond hope of handling can now be easily processed. The consequences of theories and assumptions can be tested in minutes.

The computer thus offers a new freedom that is of enormous importance to education: it can release the student from drudgery and enable him to utilize his creative powers along lines of excellence. Changes of substance and technique proposed in the educational process evoke the question: "Where in the curriculum is there time to put these new ideas into effect?"[15] Computers can give us time, if we will accept the offer.

The enemies

No one has spoken more eloquently about the enemies of creativity than Prof. John Stedman of the University of Wisconsin. The "enemies," as he sees them, are mostly undercover forces of obstruction, disparagement, and criticism. Coupled with attitudes of inattention and lack of understanding, they are formidable obstacles to invention and innovation. How are these obstacles to be counteracted? Stedman says the first thing to do is to precondition engineering students: tell them what to expect and how to deal with anticreative forces.

What should be done about the other end of the problem, which asks, "How do societies and institutions cause creative individuals to flower?" So far as is known, Jacob Rabinow was first to express the nutshell answer to the question: "Just love them and what they do." Rabinow's observation is profound; its meaning to education is that the "loving" should begin in the classroom.

There are numerous economic and sociological factors that influence invention and innovation within an organization: remuneration, bonuses, prestige, personal satisfaction, available resources (equipment, assistants, time), caution or timidity on the part of supervisors, jealousy on the part of others, innovative phobias on the part of management, opposition on the part of labor. On a national scale there may be lack of healthy competition in an industry, misguided patent policies, unsympathetic tax and other laws, bureaucracy, and, finally, the dead hand of prestige embodied in the scientific and technological establishment.[16]

Many of these influences are unreasonable and can be discerned by looking closely at ourselves—and found quite easily, of course, by assessing others. "The finger points at human nature," Senator Philip Hart has observed. "We should constantly remind ourselves of these less praiseworthy attributes or characteristics and be on our guard against [them]."[17] Understanding the nature of these human failings should be a very important part of education for innovation.

Critics ranked in rows

At a recent conference on the subject, many key observations and proposals were made to improve creative engineering education.* They have been classified under six themes, to be summarized here—but first let us understand the spirit in which the comments were made.

There have been so many criticisms of the educational system that educators must look with dismay upon the seemingly endless orations. On the one hand, they are faced with the immediate problems of education, and, on the other, with an army of critics. Certainly not all of the volumes of criticism are

*Major contributors to the conference included Robert Banks, the Ford Foundation, Mexico; Prof. William Bollay, Stanford University; Prof. Ray Bolz, Case Institute of Technology, Prof. Robert Dean, Jr., Dartmouth College; Prof. Charles Draper, Massachusetts Institute of Technology; J. Herbert Hollomon, University of Oklahoma; Richard Morse, Alfred P. Sloan School of Management, Massachusetts Institute of Technology; Emanuel Piore, International Business Machines Corporation; Prof. John Stedman, University of Wisconsin School of Law; Prof. Calvin Taylor, University of Utah; Prof. Richard Teare, Carnegie Institute of Technology.

valueless; many are sincere efforts to help further the aims of engineering education. In any case, the following proposals are meant to be constructive.

Inventors and innovators

The first of the six themes has already been introduced; it is that invention and innovation are the essence of creative engineering. They are essential to technological, social, and economic progress and, consequently, are vital to the achievement of domestic and international goals. The development of the inventive and innovative potential of engineering students should therefore be an active concern of government, industry, universities, and, in the broadest sense, society. This proposition is not novel, and is perhaps obvious to all who are concerned about engineering education. It is not so evident to the public, however[18]; and it is simple enough to be easily disregarded, as it appears to have been following the orbiting of the first man-made satellite by the Soviet Union. Although this was a spectacular feat of engineering innovation, which caused tremors in American educational circles, strangely enough it was hailed in the United States as a "scientific" miracle, and the educational reforms thereby generated in American engineering schools were science-oriented—which brings us to the next theme.

Orphans and dreary dissertations

The second theme is that the art of creative engineering has been orphaned in the engineering schools. There are exceptions, of course; but we are concerned here with the educational system in general. In the years since the Second World War, there has developed in the engineering educational system—not only in undergraduate, but in graduate schools—a regrettable and unnecessary schism between the realms of science and engineering. Paradoxically, in the schools of engineering, the art of engineering has been largely neglected. The stress has been on analysis rather than synthesis, on the abstract rather than the messy alternatives of the real world. As has been noted, the subjects typically chosen for graduate engineering theses generally have little to do with the use of technology to solve real social and industrial problems. One looks almost in vain for Ph.D. engineering theses dealing with such problems as environmental pollution and urban transportation, or with the fulfillment of an industrial need. One finds, instead, "dreary dissertations" (Prof. William Arrowsmith's description) isolated from contemporary realities. In engineering schools, why should not the *art* of engineering be given at least equal status with science?

Charles Kettering, the first chairman of the National Inventors Council, once noted that the pneumatic tire is one of the greatest inventions, yet goes unmentioned in the textbooks used in engineering schools. "Why?" he was asked. "Because there are no formulas for it," he replied. "Formulas are plentiful for low-pressure steam boilers, however; consequently, engineering students have to study these, even though they are no longer used."

It is true that engineering increasingly requires a sophisticated scientific base; but an engineer is supposed to be more than a mobile repository of knowledge, adept at attacking single-answer problems. "Where, anywhere in life," Edwin Land once asked, "is a person given this curious sequence of prepared talks and prepared questions, to which the answers are already known?"

The purpose of engineering is too often ignored in the educational system. Engineering students should understand and appreciate that their profession is the art of extending man's capabilities. Graduate students, especially, should be instilled with this sense of mission—and, at the same time, with a sense of professional pride in the contributions of engineering to the progress of mankind. The task might be easier if graduate students were encouraged to undertake, in their thesis work, the solution of engineering problems that require creative design.

Many educators believe that administrative changes might also facilitate a "return" of graduate students to engineering. They propose that the control of engineering theses be placed in the engineering schools, and not in the graduate schools of arts and sciences, whose mission and calling are different from, though relevant to, those of the engineering fraternity.

But most important, a student learning engineering should be permitted to *behave* like an engineer, and that is what the third theme is about.[19]

Learning how to swim

The third theme says, in essence, that the best way to develop and foster inventive and innovative talents in engineering students is to involve them in projects and experiences that stimulate and require such talents. As Professor Draper has expressed it: "I don't know how you can find out if a guy can swim unless you throw him in the water." The creative requisites of invention and innovation, including entrepreneurship, can be developed in an educational environment that requires and encourages these talents. There are numerous critical areas of knowledge concerning the inventive and innovative processes that can be taught to engineering students; or, through demonstration, at least brought meaningfully to their attention.

New interdisciplinary arrangements among the schools of engineering, law, business, and the social sciences—with important inputs from industry and government—would do much to enhance the teaching of engineering skills. Qualified individuals, experienced in the intricacies of invention and innovation, could be brought into the academic environment, both as full-time faculty and as consultants. They could help broaden the scope of teaching, contribute to the skills of the faculty, and enlarge the student's acquaintance with the real world of engineering.

The following are examples of ways in which engineering students could be given an opportunity to become involved in projects and experiences that develop and foster appropriate talents.

1. A specified number of the nation's outstanding inventors should be honored as such annually and, through appropriate additional inducements, brought to the universities as "masters" to work with students on special design projects.

2. An analog of the medical internship and the legal clerkship should be considered for engineering education.

3. Traveling fellowships for teachers who have distinguished themselves in the teaching of invention and innovation should be established, enabling them to demonstrate and introduce their techniques at universities throughout the country.

4. Specialists in the use of creative problem-solving techniques should be retained by engineering schools to give courses to teachers and selected students on the use of these techniques.

5. At least one faculty member per campus should be free to devote a substantial part of his academic year to the organization and introduction of teaching programs designed to stimulate creativity.

6. Design laboratories should be established at engineering schools that will enable students to work on personally selected engineering problems, thereby gaining experience in the planning, design, development, testing, and evaluation of a new product or process.

7. A mechanism should be established to bring promising student inventions to the attention of potential users. Students would thus be given an opportunity to do something with their conceptions, and would learn that invention is merely a part of the innovative process.

Understanding technological change

The fourth theme concerns the need for greater understanding of the processes of technological change. Although the United States annually commits enormous resources to these processes, there is much that needs to be learned about their nature, consequences, and determinants. We know little, for instance, about the influence of tax and other Federal policies on science.

Moreover, existing knowledge about the world of technological change is not only rudimentary, it is disordered. It is compartmentalized in technical, economic, sociological, legal, and other literature. It really needs to be correlated in a comprehensive, multidisciplinary program, so that key gaps can be identified, and research undertaken with respect to them. Models that are more representative of the inventive and innovative processes should be devised, ranging from simple descriptive representations to computer and game simulations. Because this program of research would be of value to numerous communities of interest, it should be undertaken by the government in cooperation with the universities, industry, and the professions. Suitable steps should be taken, through seminars, publications, and consultation, to assure maximum communication and exchange of knowledge and insights developed under the program.

Improving the educational climate

The fifth theme is: If any of the suggested programs are to be successfully undertaken, the *climate* for creative engineering education will have to be improved. It is not enough for academic administrators simply to approve of creativity in education; there must be a favorable environment for both students and teachers. Merely to proclaim the virtues of creativity is familiar tokenism. A favorable environment requires positive support, academic freedom, and sympathetic accommodation for innovative faculty members.

The environment for the student is equally important. In most engineering schools, grading practices do not encourage novel departures from, or challenges to, material absorbed in class and through study. Feedback is expected; the closer the student adheres to what is "correct" or accepted, the better the grades—and the better the grades, the more prodigious and passionate the courtships of recruiters at school's end. In short, the grading systems customarily employed in engineering education tend primarily to measure retentive prowess rather than creativity. And grades are a critical determinant of a student's attitudes and propensities.

Creativity, like invention and innovation, is risky. If the system discourages it, why gamble? After years of conditioning, a student may well ask himself: "Who wants creativity?"

The system of grading, as it is commonly practiced in the engineering schools, needs to be evaluated and its positive and negative effects assessed. One recent development, which appears to be gaining favor in many schools, is the option to take at least one course annually on a "pass–fail" basis. This scheme tends to reduce the riskiness of being creative. Even in courses where a "pass–fail" system would be inappropriate or too frighteningly new, the creative performance of a student should be taken into account. For instance, the constructive challenge of presuppositions and orthodoxy, the continual search for better ways of doing things, the observation of the familiar in new lights, and the ability to suggest alternative solutions and not just one "right" answer all are indications of a talent for healthy dissent and creative initiative. Indeed, teachers should strive to identify and encourage traits of originality and inventiveness at all levels of engineering education. If awards and other forms of recognition were given to outstandingly creative students and, particularly, to those responsible for inventive solutions to problems, these would do much to encourage creativity.

Supporting innovation in education

The sixth and last theme takes note of the need for greater cooperation among the universities, industry, the foundations, the professional associations, and government. Their cooperation is essential to the development and support of creative engineering education. They need to identify and communicate problems and opportunities.

If, as is commonly assumed, industry and government agencies need and want creative engineers, they should so inform the universities. First, however, they should appraise their own needs. Are they really committed to the disruptive process of innovation? Do they really encourage it?

If they truly want inventors and innovators, they should support programs and institutions that produce such individuals. The most penetrating influence in the academic world is not exhortation, but tangible support.[20] Tangible, sympathetic support is an essential element of the innovative process—in engineering education, no less than elsewhere. It gives organizations the "will to think," as William Shockley has observed.

Conclusion

We have concentrated on six principal themes or propositions: (1) inventors and innovators are the vital elements of technological change, which is the business of creative engineering; (2) the art of creative engineering has been orphaned in the engineering schools; (3) the creative requisites of invention and innovation can be encouraged in an educational environment that requires and encourages these talents, and an understanding of these processes can be taught; (4) research on the processes of technological change, including a synthesis of existing knowledge, is needed and should be undertaken on a comprehensive, multidisciplinary basis by government, industry, and universities; (5) improving the climate for creative engineering education requires positive inducements to faculty and students; and (6) greater cooperation among the universities, industry, the foundations, professional groups, and government is essential to the development and support of creative engineering education.

Perhaps the most appropriate way to conclude this essay is to paraphrase the Victorian economist, Alfred Marshall: Every short statement about the needs of education is misleading, with the possible exception of this one.

In any case, what is proposed by the six themes we have explored is not an abandonment of science or of the need to master the structure of scientific and engineering knowledge. It is not the establishment of a design-oriented monolith to supplant the predominantly science-oriented regime that now exists in most engineering schools. Rather, what is suggested is a recognition of the essential nature and role of engineering, the learning of engineering by behaving like an engineer, and an increased emphasis on the excitement of discovering for oneself. In short, what is proposed is that engineering education be kept alive and relevant—that it truly be education for innovation.

This article is based on the book *Education for Innovation*, to be published by Pergamon Press, New York, N.Y., early in 1968.

REFERENCES

1. Whitehead, A. N., *The Aims of Education and Other Essays*. New York: Macmillan, 1959.
2. Bruner, J., "The conditions of creativity," in *Contemporary Approaches to Creative Thinking* H. E. Gruber *et al.*, eds. New York: Atherton Press, 1962, p. 3.
3. MacKinnon, D. W., "Fostering creativity in engineering students," *J. Eng. Ed.*, vol. 52, p. 130, Dec. 1961.
4. Coler, M. A., "The two creativities," *Chem. Eng. News*, vol. 44, p. 73, Aug. 15, 1966.
5. "Technological innovation: its environment and management," U.S. Government Printing Office, Washington, D.C., 1967.
6. Denison, E., "The sources of economic growth in the United States," Committee for Economic Development, Washington, D.C., 1962.
7. Kendrick, J., "Productivity trends in the United States," National Bureau of Economic Research, Washington, D.C., 1961.
8. Solow, R., "Technical change and the aggregate production function," *Rev. Econ. Statist.*, Aug. 1967.
9. Eiseley, Loren, *The Immense Journey*. New York: Random House, 1957, p. 89.
10. Close, Upton, *The Revolt of Asia*. New York: Putnam, 1927, p. 232.
11. Toynbee, A. J., *Change and Habit*. London: Oxford University Press, 1966, pp. 200–201.
12. Gardner, J. W., *Self-Renewal*. New York: Harper & Row, 1963, p. 57.
13. McLuhan, Marshall, *Understanding Media: The Extensions of Man*. New York: McGraw-Hill, 1964.
14. "Computers in higher education," U.S. Government Printing Office, Washington, D.C., 1967.
15. "Interim report of the Committee on Goals of Engineering Education," American Society for Engineering Education, Washington, D.C., 1967.
16. Dudinstev, Valdimir, *Not by Bread Alone*. New York: Dutton, 1957.
17. U. S. Senate Subcommittee on Antitrust and Monopoly, "Concentration, invention, and innovation," U.S. Government Printing Office, Washington, D.C., 1965, p. 1332.
18. Blade, Mary-Frances, "Creativity in engineering," in *Essays on Creativity in the Sciences*, M. E. Coler, ed. New York: New York University Press, 1963, pp. 112–113.
19. Bruner, J. S., *The Process of Education*. Cambridge, Mass.: Harvard University Press, 1960, p. 14.
20. Pierce, J. R., "What are we doing to engineering education?" *Science*, vol. 139, p. 397, July 23, 1965.

Digital control of shaft speed and position

For precise speed control of rotating shafts, digital techniques are proving to be extremely useful. Digital circuits, because of their binary nature, enjoy relative freedom from variations in supply voltages and component values

R. A. Millar *Huntec Limited*

Where speed accuracies within 0.1 percent are desired over a wide speed range, the digital approach has proved superior to the analog method. This article describes a typical hybrid digital/analog servo system used to control the speed of a dc motor. A vital part of the system is the comparator, which must combine the functions of frequency and phase discrimination and perform the transition smoothly. It is also shown how precise synchronism between two or more rotating systems, with respect to both speed and position, can be maintained, giving the effect of a "synchronous link."

Precise control of shaft speed is frequently specified in research, industrial, and military applications in which important parameters are critically dependent on speed. A common application in industry is web control, such as a paper drive, where maintenance of the exact speed relationship between various rollers is essential for correct web tension, since a change of 0.1 percent of the absolute speed of one roller can cause a pileup or tearing of the paper. In research a centrifuge might be required to generate precisely programmed accelerations for the evaluation of the accuracy of accelerometers. Military photoreconnaissance requires the precise advancement of film at slow rates in order to eliminate jitter and image smearing, which could cause loss of detail and resolution. In the computer field it is sometimes desirable to synchronize several magnetic drums with each other, in order to increase data storage, access time, and stability of display.

Where only one speed is required, the simple approach is to use an ac synchronous motor fed from a power source of stable frequency. In order to operate over a wide range of speed, the volts-per-hertz relationship must be kept constant, and for this purpose there are now available various forms of silicon controlled rectifier electronic frequency inverters. Without closed-loop control, however, the speed accuracy is limited by motor hunting.

One form of conventional feedback speed control compares a direct voltage from a tachogenerator (which provides the analog signal of speed) with a reference voltage. Another form uses an ac signal from the tachogenerator and feeds it to a frequency discriminator. In either case the error signal is amplified to control power to the motor. With a large amplification the speed error of the velocity servo is small, but it cannot be zero.

Recent technology in the field of automatic speed control has shown increasing emphasis on the use of digital techniques for the derivation of the error signal within the servo loop. In fact, where speed accuracies of better than 0.1 percent are desired over a wide speed range, the digital approach has proved superior to the analog method, the latter being limited by problems of drift and stability relating to reference voltages, summing networks, and tuned circuits. The binary nature of digital circuits offers the inherent advantage of relative freedom from variations in supply voltages and component values. Also, the region of servo control is confined to a discrete "bit" of shaft position and theoretically can be made very small.

The digital technique can be conveniently extended to situations involving more than one rotating system, even when they are placed remotely from each other. Precise synchronism, with respect to both speed and position, can

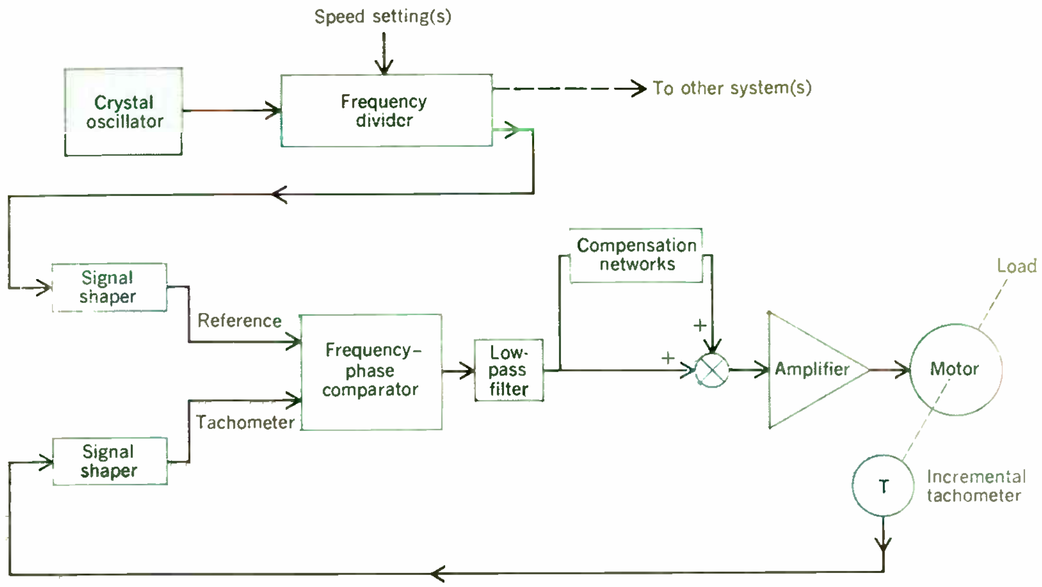
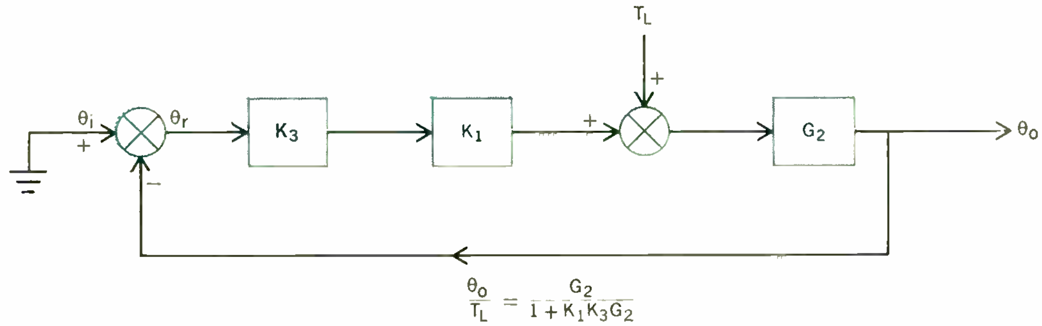


FIGURE 1. Block diagram of hybrid digital/analog speed-control servo system.

FIGURE 2. Position servo with disturbing torque.



be maintained between the systems, resulting in the effect of a synchronous link. By varying speed ratios the electronic equivalent of a variable gear box is also realized.

The speed-control system

Figure 1 illustrates a typical hybrid digital/analog servo system used to control the speed of a dc motor. The command and feedback control inputs take the form of two pulse trains, which are designated "reference" and "tachometer" respectively. The function of the frequency-phase comparator is to compare these two pulse trains and generate an error signal that, when suitably amplified, causes the motor to rotate in such a manner as to phase-lock the tachometer pulse train to the reference pulse train.

At this point the action of the system resembles that of an ac synchronous motor with the unique features of a large permissible frequency variation and a large number of poles. Another aspect of the stystem is that of a phase-locked oscillator[1] whose output frequency, that of the tachometer, is locked to the input reference frequency.

The reference pulse train is obtained from a digital frequency synthesizer[2] whose "clock" can be a crystal, a tuning fork, or a relaxation type of oscillator. Since the servo locks the tachometer frequency to that of the reference, it is the stability of the clock that governs the long-term speed stability of the system. It is quite feasible nowadays to obtain crystals with stabilities of a few parts per million per day. The simple frequency synthesizer of Fig. 1 shows a crystal oscillator feeding binary frequency dividers whose outputs are selected by the speed-setting (command) switch.

The feedback pulse train is generated from an incremental-type optical or magnetic tachometer, which is coupled to the output shaft. Incremental optical encoders have been made with line densities of over 10 000 per shaft revolution. The tachometer incremental density N and the required output revolutions per minute Ω together define the reference frequency needed, as follows:

$$\text{Reference frequency} = \frac{\Omega N}{60} \text{ hertz} \qquad (1)$$

The hybrid part of the system includes a low-pass filter (to remove the reference-frequency component), conventional amplifiers, and compensation networks.

Since the type of servo action, during phase lock, is one of controlling the instantaneous position of the rotating shaft, the system is inherently unstable, and thus error-rate damping[3] is required. In addition, integral control is used to reduce the steady-state error.

A block diagram of the position-type servo is shown in Fig. 2 with input $\theta_i = 0$; T_L is the load torque disturbance. K_1G_2 represents the motor transfer function, with

$$K_1 = \frac{K_D}{K_E}$$

where K_D is the damping constant, expressed in newton-centimeter-seconds per radian, and K_E is the induced EMF constant, in volt-seconds per radian.

$$G_2 = \frac{1}{K_D s(sT + 1)} \text{ rad/N·cm}$$

where s is the Laplace transform of d/dt, in seconds^{-1}, and T = mechanical time constant, in seconds, given by

$$T = \frac{1}{K_D} \times \text{inertia}$$

and K_3, in volts per radian, is the gain of the phase comparator and amplifiers.

From the block diagram,

$$(K_1K_3\theta_r + T_L)G_2 = \theta_o$$
$$\theta_r = -\theta_o$$

Combining these equations, we get the position response of the system:

$$\frac{\theta_o}{T_L} = \frac{G_2}{1 + K_1K_3G_2} \text{ rad/N·cm loading} \qquad (2)$$
$$= \frac{1}{K_1K_3} \text{ for low- and mid-band frequencies}$$

The product K_1K_3 gives a measure of the "stiffness" of servo action.

The advantages of a direct-drive dc motor[4] are of special interest here since the control system accuracies can impose high performance demands on the motor. The use of direct coupling eliminates problems of geared servos relating to backlash, resilience in the gear train, and high rotor inertia. The direct-drive motor is designed to have high torque-to-inertia ratio, fast response, freedom from cogging, and high torque at low speeds without overheating. It can be used at speeds ranging from almost zero to several thousand revolutions per minute.

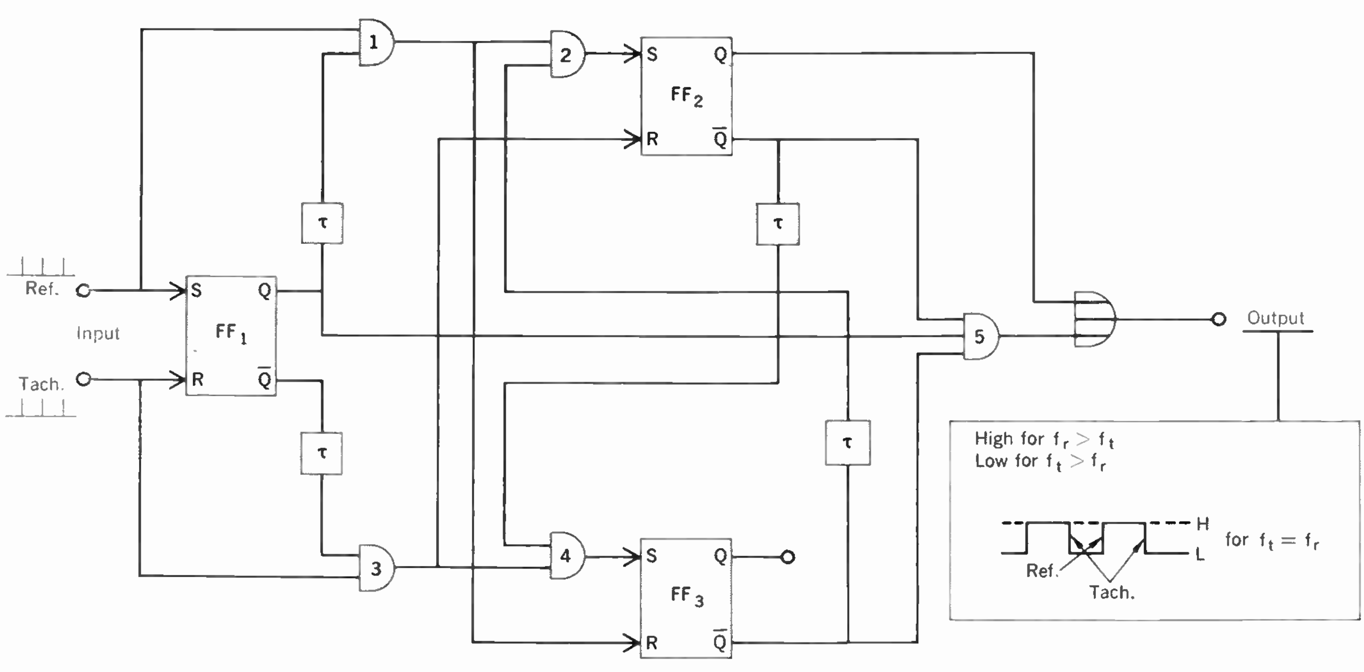
The comparator

The error signal is derived from a digital comparator, which must combine the functions of frequency and phase discrimination and perform the transition in a continuous manner. Moreover, the comparator must not permit the system to "lock" onto a multiple or submultiple of the reference frequency.

Various methods of achieving this error signal have been described elsewhere. Some techniques involve the use of a reversible counterregister,[5] which is driven up by the reference pulses and down by the tachometer pulses. The net accumulated difference in count is continuously monitored by a digital-to-analog converter, which delivers a proportional error signal to the drive amplifier. The steady-state gain is made quite high, with a small error count sufficient to saturate the drive amplifier; thus, the net count is practically zero at the required speed. Since the counter is in effect acting as an integrator, any difference in pulse rate, no matter how small, will be detected and corrective action will be applied.

A useful frequency synthesizer for such a system is a "binary multiplier," set to deliver at its output some

FIGURE 3. Logic diagram of digital frequency-phase comparator.



exact fraction of the input pulse rate by selective mixing of the binary-stage outputs. For example, a five-binary-stage multiplier of this type can produce 32 different speed ratios, variable in increments of 1/32.

The logic diagram of a relatively simple type of comparator, comprising three gated reset–set flip-flops, is shown in Fig. 3. The operation of the comparator logic is characterized by three states:

1. Frequency of the reference (f_r) greater than that of the tachometer (f_t). The Q output of FF_1 enables a reference pulse to reset FF_3 via AND gate 1. The Q output of FF_3 in turn enables AND gate 2 to pass a reference pulse to set FF_2. The Q output of this flip-flop passes through the OR gate and establishes the comparator output at "high," signaling that the motor speed is below that of the command setting. This condition causes full positive acceleration to be applied to the motor.

2. Frequency of tachometer greater than that of the reference. The Q output of FF_2 enables a tachometer pulse to reset FF_2 via AND gate 3. The $\bar{Q}$ output of FF_2 in turn enables AND gate 4 to pass a tachometer pulse to set FF_3. The comparator output is now maintained at "low," signaling that the motor speed is above that of the command setting. This condition causes full negative acceleration to be applied to the motor.

3. Frequency of reference equal to that of the tachometer. Both FF_2 and FF_3 would have been reset. The Q outputs will enable AND gate 5 to pass the Q state of FF_1, which is set by a reference pulse and reset by a tachometer pulse. This signal, after filtering, drives the motor at the correct speed.

The $f_r = f_t$ condition is actually obtained during the transition from $f_r > f_t$ to $f_t > f_r$ or vice versa. It could be said that the closed-loop operation defines the states $f_r > f_t$ and $f_t > f_r$ as being unstable regions and $f_r = f_t$ as the stable region wherein the system pulls the motor into lock with the reference. The system cannot lock onto multiples or submultiples of the reference, since these imply states 1 or 2 just described.

By observing the comparator output waveform for the state $f_r = f_t$, one can obtain a measure of the instantaneous angular positional accuracy of the control system. This is given by the peak amount of jitter of the edge due to the tachometer pulse. If the observed jitter is one pth of one reference cycle and the tachometer incremental density is given by N, then the instantaneous positional accuracy is equal to

$$\pm\frac{1}{P} \times \frac{1}{N} \times 360 \times 3600 \text{ seconds of arc} \qquad (3)$$

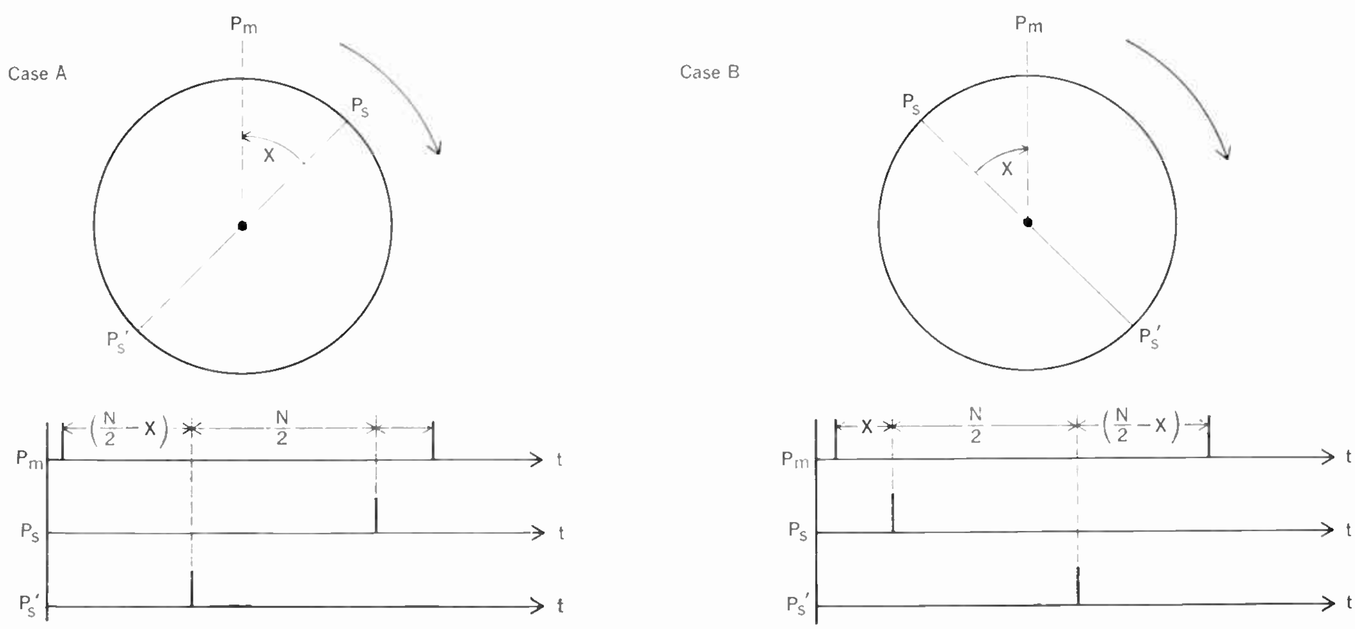
Factors such as servo stiffness, tachometer accuracy, mechanical coupling, and torque disturbances influence the ultimate precision of the position control. For very low speeds, consideration must be given to the input frequency because the information from the signal inputs is obtained on a sampling basis, and thus if the frequency is too low a significant lag will be introduced into the system. If the logic response time is adequate, the maximum speed is limited by the maximum "rpm" rating of the motor or the load.

It should be noted that the digital logic of the control system can be easily implemented by integrated circuits and that the comparator could even be fabricated in monolithic form.

Speed and position synchronization

In applications involving two or more rotating systems, the individual speed controllers can be driven from a common frequency synthesizer. Independent speed settings can be made for each rotating shaft, which will then be rotating in exact speed synchronism with respect to the other(s). Alternatively, one system can act as the master and its tachometer output can be used as the reference signal for the other systems (slaves).

FIGURE 4. Timing diagram of index pulses.



N = number of tachometer increments per revolution

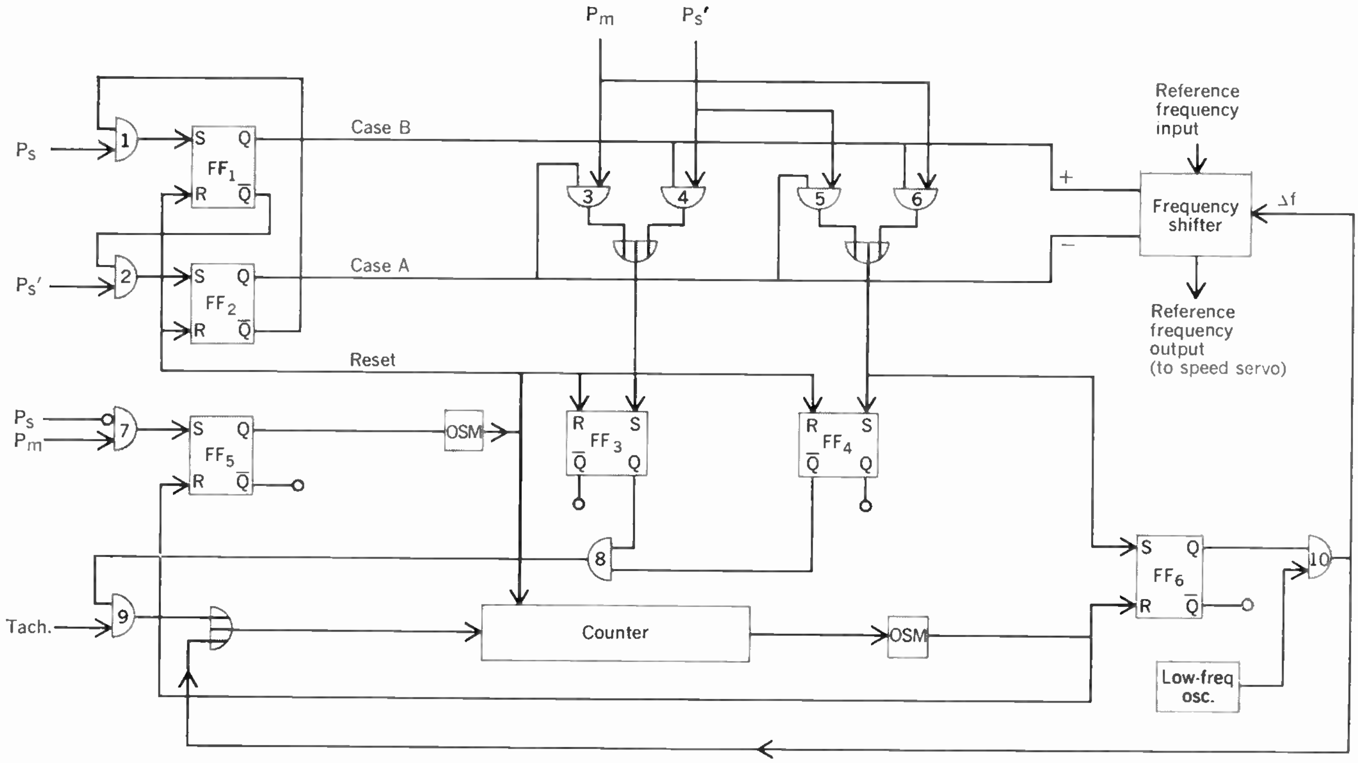


FIGURE 5. Logic diagram of position synchronization system.

Further, it is possible to increment one rotating shaft clockwise or counterclockwise, with respect to another, simply by phase-shifting the reference input signal (to that servo system) continuously through 360 degrees. This operation, analogous to rotor slip, can also be described as frequency shifting; when performed in controlled amounts, it results in precise position control.

The following scheme was designed for a specific situation in which the pulse rate of the master index pulses was very low compared with the incrementing rate.

Figure 4 is a timing diagram of the various index pulses concerned. P_m and P_s are pulses picked off the rotating shafts of the master and the slave respectively. P_s' is also a slave pulse, but it is spaced 180 mechanical degrees from the pulse P_s. In case A or B the quantity X is the amount of difference (lead or lag) between P_m and P_s in terms of tachometer increments. The requirement then is to shift or increment the slave shaft until P_m and P_s coincide. From the timing sequences shown it is apparent that if the interval ($\frac{1}{2}N - X$) is selected and stored in a modulo $N/2$ counter, the phase-shifting operation can be started and each 360-degree shift counted and added to the stored number until zero state is reached —that is, until X pulses are fed into the counter. When this state is reached, the phase-shifting operation can be terminated; P_s should then be in coincidence with P_m. Pulse P_s' is used to determine the sense of direction, clockwise or counterclockwise, that the slave shaft must be incremented to achieve synchronism with the master.

Figure 5 illustrates the logic of the position synchronization operation. Gate 7 performs the "coincidence test" between pulses P_s and P_m, and if there is an "out-of-coincidence" condition, FF_5 is set. The Q output triggers the one-shot multivibrator (OSM), which resets the counter and flip-flops 1, 2, 3, and 4. FF_1, FF_2, and gates 1 and 2 function as the lag–lead discriminator with a lock-out feature. The Q outputs of the flip-flops enable gate pairs 3 and 5 or 4 and 6 to pass the P_m and P_s' pulses, as the case may be. For case A, the P_m pulse sets FF_3, which starts the tachometer count, and the P_s' pulse sets FF_4, which stops the count. For case B, the P_s' pulse starts the count and the P_m pulse stops it.

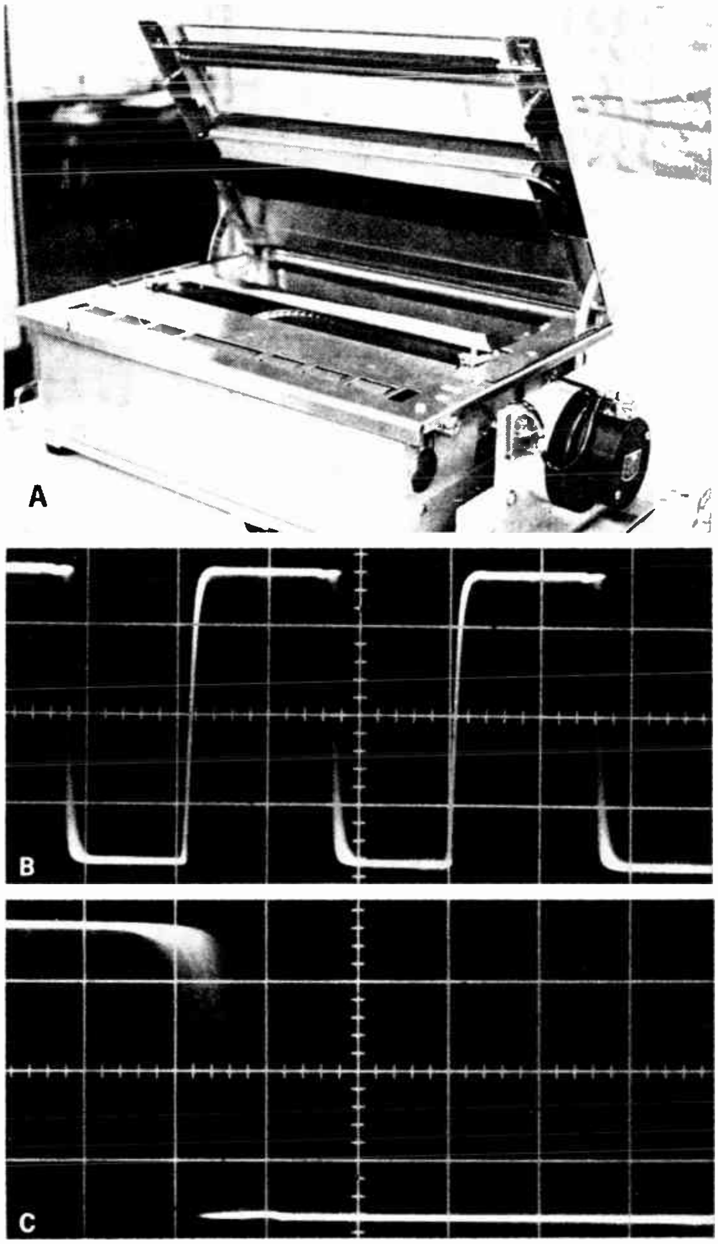
The action of stopping the count, in either case, also sets FF_6, which turns on the frequency shifter via gate 10. Each 360-degree phase shift is counted and added into the counter until the zero state is reached. When this occurs, FF_5 and FF_6 are reset via the "one shot," the frequency shift is terminated, and the coincidence test is resumed.

With slight modification, the system can be designed to position-synchronize slave shafts rotating at multiples or submultiples of the master speed. If the slave is rotating at high speeds, or coincidence is monitored at a high enough rate, one can dispose of the counter and resort to a simpler scheme. The frequency shifter can be implemented, in several ways, by single-sideband generation or digital methods of phase shifting. A novel method is to use a phase-shifting resolver driven by a bidirectional stepping motor.

Applications

Typical uses for the digital speed servo are to be found in such areas as precision drives for tape recorder cap-

FIGURE 6. A—Facsimile-type recorder used to chart ocean subbottom geological horizons. B and C—Oscilloscope traces of comparator output in phase-lock condition.



stans, antenna systems, facsimile recorders, camera shutter controls, computer magnetic storage drums, centrifuges generating precisely programmed accelerations, rolling mills, multicolor rotogravure presses, and machine tools. The settings of speed can be effected manually by switches (calibrated directly in revolutions per minute) or automatically programmed by digital devices such as punched tape, cards, etc. For example, the speed control system can be incorporated into a numerically controlled machine tool scheme wherein a punched tape provides the digital data for work-table positions, cutting tools, and operating speeds.

The digital speed–position control system has direct application in process control wherever it is desired to coordinate various parts of the process along the line. Ratio control of speed by use of frequency dividers produces the effect of mechanical gearing—a particularly useful feature for widely separated operations. One crystal oscillator forms the frequency standard to which the various parts of the process can be precisely related.

By substitution of a variable-frequency oscillator for the fixed one, it is feasible to obtain controlled acceleration whereby all motors, being referenced to the variable-frequency oscillator, smoothly accelerate to operating speed while maintaining exact speed ratios between themselves.

The particular system described in this article has been used for the precise speed control of a facsimile-type recorder in an oceanographic instrument to chart subbottom geological horizons; see Fig. 6(A). The recorder used has a resolution of about 0.02 cm and a sweep about 20 cm long. It was decided to restrict any jitter to one tenth of the recorder resolution. Thus, the jitter or position control specified was

$$\frac{1}{10} \times \frac{0.02}{20} \times 100 = 0.01 \text{ percent}$$

The speed range of interest is 30–960 r/min and the inertia of the recording drum is 0.35 N·cm·s^2.

A crystal oscillator with 0.01 percent long-term stability is used to supply the reference frequency. The tachometer has a line density of 2000 lines per revolution. The maximum torque disturbance anticipated was about 4.2 N·cm. The "stiffness" factor, K_1K_3, was calculated as 85 dB, giving a displacement error of 14 seconds of arc per newton-centimeter. Figures 6(B) and 6(C) are oscilloscope traces of the comparator output in the phase-lock condition. The peak-to-peak time jitter of the tachometer pulse observed (with the drum rotating at 480 r/min and reference frequency = 16 kHz) is about 4 μs on a time base of 5 μs/cm; see Fig. 6(C). Thus the positional accuracy is

$$\frac{4}{62.5} \times \frac{1}{2000} \times 100 = 0.003 \text{ percent}$$

The result has been improved registration and time resolution as compared with a previous drive employing an ac synchronous motor with gearing and clutches. The application also features speed and position synchronization—that is, "slaving" of one recorder to a "master." The P_m and P_s index pulses are synchronized to within a band of three tachometer pulses ($\pm$16 minutes of arc).

Revised text of a paper presented at the International Electronics Conference, Toronto, Ont., Canada, Sept. 22–27, 1967.

REFERENCES

1. Gardner, F. M., *Phase-lock Techniques*. New York: Wiley, 1966.

2. Husson, G., and Sherman, B., "A transistorized frequency synthesizer," *J. Brit. IRE*, Apr. 1961.

3. Thaler, G. J., and Brown, R. G., *Analysis and Design of Feedback Control Systems*. New York: McGraw-Hill, 1960.

4. Aherne, J. M., "The direct-drive dc servo," *Electro-Technol. (New York)*, June 1963.

5. Kent, G. B., "Hybrid digital/analog servomechanisms," *J. Brit. IRE*, Sept. 1963.

IEEE publications

scanning the issues
advance abstracts
translated journals
special publications

Scanning the issues

On Our Kind of Writing. Engineers and scientists probably get bludgeoned and bludgeon themselves more than any other group about their writing ability—which is rather a shame, since the bludgeoning probably does very little good, and ends usually in being just another form of self-torture. I would guess that anyone who really wants to write a bit better, day by day, will do far more for himself by reading things he *enjoys*—novels, stories, plays, comic books, whatever—and let the elements he enjoys come into his own writing through osmosis. One might call this "improving one's writing by the pleasure principle." To do anything creative, one has got to let up and be playful. As Heine advised, suspend your self-criticism until afterwards. Those who approach the act of writing with solemnity and dead seriousness usually end up sounding like solemn asses. One does not need to look far to find "examples of same."

This is not to deny the need to learn the rudiments of the structure of one's language nor does it deny the need to analyze your writing. But why does a man who has been through at least four years of college need to be told that, if he wishes to improve his writing ability, he should spend six months reading Fowler's *Modern English Usage*, a little each day? He surely knows, if the educational system has conditioned him at all, that he's got to master the structure and the subtleties of the structure of his language as best he can. By that time, he surely ought to know, too, that grammar is not all, that grammer is merely the acquired *theory* of how the language functions structurally. However, the theory is a rough approximation, and the really interesting questions involved in writing are not resolved by going to the grammar cookbook of rules and formulas but are embodied in the whole world stream of literature.

Although one is tempted to go on to make many points (e.g., that the stream of technical writing is often turbid and therefore makes the job of finding good examples and good influences for one's own writing more of a chore, but that, nevertheless, there are superb examples of fine writing in this stream, etc.), the main point is that spirit, fun, and play lie at the heart of seriousness. To acknowledge that much is to liberate oneself from some of the bludgeoning we all undergo.

So much for the spirit with which one enters into his "creative outpouring." Suspend criticism *during* but not after the act. Afterwards, there *must be* analysis, criticism, and rewriting; and if one is not used to analyzing his own writing, one needs a few ideas to get started. If you feel you are in this category of benighted and bludgeoned confusion, then you might look up George E. Schindler, Jr., in the current issue of the IEEE TRANSACTIONS ON ENGINEERING WRITING AND SPEECH. His analysis of a dozen characteristics of technical prose is worth reading. It is, in contrast to most such efforts, a lively and sensible contribution, and I should think that if it leads one or two spirited young men to go on to tackle Fowler, Schindler will have done more than his bit for "the cause." And the cause, of course, is the brightening up of the technical literature that we all must read as we pursue our professions. Remember, bludgeoned writers make for bludgeoned readers. That ain't good. It isn't even fun. (G. E. Schindler, "Why Engineers and Scientists Write as They Do—Twelve Characteristics of Their Prose," *IEEE Trans. on Engineering Writing and Speech*, December 1967.)

Aerospace Computers. The current issue of the IEEE TRANSACTIONS ON ELECTRONIC COMPUTERS is a special issue giving us a keyhole on the aerospace computer field. It provides, says L. J. Andrews in a short but provocative foreword, some interesting and important ideas heretofore not publicized. In the majority of the papers, the emphasis is on systems, but more than that, on creative system engineering applied to aerospace computers. Some of the papers cover contemporary computers—bench marks for the present generation. These computers do not embrace the concepts of reliability enhancement and kiloelement components suggested in other systems. Even these papers represent only the first steps in the technological changes anticipated for the near future. With the coming of kiloelement components, the possibilities for lighter, faster, more reliable computers are great, inevitable, and *unimportant*. What is really important, stresses Andrews, is the opportunity for a breakthrough in the Babbage, Hollerrith, and Von Neumann sense. The functional freedom represented by large-scale integration (which is discussed in two papers) will have an impact as great as the vacuum tube and transistor.

Andrews then launches on an argument that one is curious to hear more about. It revolves around the popular notion that computers are only automata, giving the semblance of thinking, but never creating. We have had, says Andrews, to redefine what we mean by "thinking" several times to keep the statement true. And we may soon have to redefine what we mean by "create." Man created the computer, but he seemed to use his own thinking procedures as a model, a single thread of thought processes taken one step at a time. After 20 years of progress our concept of organization has not extended to that of the organization of the office staff, a central supervisor delegating tasks to autonomous workers, file clerks, typists, and messengers. Andrews' hypothesis, he says, is that although man himself and his environment formerly played an intrinsic role in the solution to computation problems, he may now be part of the problem of how to progress. The challenge may not be to evolve the next member of the computer species in our own image, but to create a viable true mutation. (L. J. Andrews, "Foreword on Aerospace Computers Special Issue," *IEEE Trans. on Electronic Computers*, October 1967.)

Speech Research. In a surprising way, research on human speech is becoming of central interest in many different scientific and engineering endeavors. Some of the engineering interests in speech are spelled out in four papers in recent issues of IEEE TRANSACTIONS.

For a general review, the reader may wish to read "Systems for Compressing the Bandwidth of Speech," in the September IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS. The paper, by Bernard Gold and Charles Rader of the M.I.T. Lincoln Laboratory, should be noted particularly because it is the first in a series of papers that these authors are preparing dealing with speech bandwidth compression systems. Although the authors' primary purpose is to acquaint readers with the engineering aspects of channel vocoder design, they are covering a number of areas of speech research and application such as formant vocoders and evaluation of speech systems. Their initial paper summarizes some of the pertinent knowledge of the acoustic properties of speech, which constitutes the foundation from which more recent speech work is taking off. The authors thus include an inventory of English speech sounds, which they describe on several levels. One description is in terms of the positions of the articulators, e.g., tongue, vocal cords, and so on; a second description is on the acoustic level; and the third is in terms of electrical networks and generators. These basic descriptions are necessary for the understanding of speech bandwidth compression systems, which are forms of "analysis-synthesis" systems. As the authors say, the concepts needed to invent speech bandwidth compression systems derive from ideas about human speech production and perception. Of these ideas, they go on, two stand out and appear to encompass much of the significant work that has been done up to now in the field. One, dealing with speech production, is the development of physical analogs of the system of valves and acoustic tubes that comprise the human vocal mechanism to make "talking machines." A bandwidth compression device can be regarded as comprising such a talking machine *plus* another machine that can extract from the acoustic signal a set of control signals that can properly actuate the talking machine. This latter machine, called the synthesizer, and the machine that derives the control signals, called the analyzer, constitute the complete system. The authors discuss several ramifications of such speech bandwidth compression systems.

Another rather general paper, "Speech Processing Techniques and Applications," appears in the same TRANSACTIONS. Its authors, H. F. Olson *et al.*, make their interest in speech analysis and synthesis quite explicit. The objective in communications systems is to provide a savings in the channel capacity required for transmission. The main concern in speech processing by machine, they say, is to extract automatically the information content of the signal, a process that, if it is successfully implemented, could reduce the bandwidth requirements for speech transmission by a factor of more than 1000 to 1. Among other matters, these authors discuss some of the problems involved in speech recognition machines based on the analysis and synthesis of speech elements that can be easily segmented. Only in their conclusion do these authors give a hint of the perplexing depth of the so-called segmentation problem. Diligent research and development programs, they say, must be carried out in order to accomplish the segmentation automatically by machine.

A less general paper in the same issue, "Experimental, Limited Vocabulary, Speech Recognizer," by C. F. Teacher *et al.*, is nonetheless of some interest because of its aim at developing a simple, small, and practical speech recognizer. The device recognizes the spoken digits "oh" through "nine" with an accuracy of 90 percent.

The fourth paper to note, and by no means the least interesting, is "Spelled Speech as an Output for Computers and Reading Machines for the Blind," which appears in the September issue of the IEEE TRANSACTIONS ON HUMAN FACTORS IN ELECTRONICS. The author, Kenneth Ingham, describes the development of a speech display system for the blind built around the TX-O and PDP-1 computers at M. I. T. The system, which permits the application of a variety of sound compression techniques, and in which spelled-speech alphabets and displays were developed, rested its hopes on the finding that if words were spelled aloud rapidly enough, the whole word, rather than the individual letters, would become the unit of perception to the blind person. The various approaches detailed in the paper by Ingham are intended, he says, to provide an immediate, although perhaps temporary, solution to the desperate reading problem of the blind person. (*IEEE Trans.*)

Electrifying the Railroads. Electric propulsion of vehicles—both cars and trains—has long been a dormant concept in the United States, although other countries such as Russia, France, Italy, and Japan have expanded their electrified rail systems substantially in recent years, and England never wholly gave up the electric car as did the United States. Economics, of course, was the dominating factor. In the United States, it was difficult to justify the costs of electrification in view of the existing traffic patterns and market demands, plus the fact that the entire transportation industry was undergoing rapid and radical change. Now, however, there have been significant developments in the past few years that, in their cumulative effect, have caused a dramatic revival of interest in railroad electrification. And air pollution, of course, has brought great public interest in the evolution of practical electric cars. The electric utilities stand to gain much in both instances.

In the midst of a persuasive argument on railroad electrification, Edwin O. George, president of The Detroit Edison Company, remarks that no mode of transportation has been invented that is more economical and efficient than the electrically driven steel wheel on a steel rail.

Among the more important new factors that now favor railroad electrification, George cites the development of the rectifier locomotive that opened the way for application of commercial 60-Hz power with all the attendant benefits and savings.

Among the nontechnical changes have been the U.S. population shifts into concentrated areas, thus raising the prospects for traffic densities in goods (and apparently people) that will be attractive to railroad managers.

On the other side of the picture are the electric utilities, which have been increasing greatly in size and which are therefore coming into a position where they could handle the required load.

This says enough, perhaps, to suggest a future trend to watch for. George goes into much more detail for those who want it. (E. O. George, "Railroad Electrification—Challenge and Opportunity," *IEEE Trans. on Industry and General Applications*, September/October 1967.)

Advance abstracts

The IEEE publications listed and abstracted below will be available shortly. Single copies may be ordered from IEEE, 345 East 47 Street, New York, N.Y. 10017. Prices are listed with the abstracts of each publication; libraries and nonmembers outside the United States and Canada should add $0.50. (M—Members; L—Libraries; NM—Nonmembers.)

Copies of individual articles are not available from IEEE but may be purchased from the Engineering Societies Library at the foregoing address.

Proceedings of the IEEE
IEEE Journal of Quantum Electronics
IEEE Journal of Solid-State Circuits
IEEE Transactions on

Circuit Theory	Industry and General Applications
Education	Microwave Theory and Techniques
Electron Devices	Parts, Materials and Packaging

Proceedings of the IEEE

Vol. 56, no. 1, January 1968
M—$1.00; L—$1.50; NM—$2.00

Distortion and Crosstalk of Linearly Filtered, Angle-Modulated Signals, *E. Bedrosian, S. E. Rice*—An important problem in the theory and practice of receiving angle-modulated signals is the design of the filtering elements that must be employed. It has long been known that filtering introduces distortion and crosstalk into the signal. However, the computation of these effects is difficult. The methods customarily used employ approximations of one kind or another, and the equations used do not apply to all cases of practical interest. Here formulas are presented that enlarge somewhat the domain of cases amenable to calculation. In this analysis, an angle-modulated signal having an arbitrary phase function is applied to a general linear filter, and the phase of the output is expanded in a series having the linearly filtered input as the leading term. The expansion is then specialized to the case of a narrowband signal applied to a narrow, symmetrical, bandpass filter. A spectral analysis is performed by assuming a Gaussian input phase and examining terms through fifth order in the output phase expansion. This leads to the main results given, namely, expressions for the leading terms in the output spectrum. It is argued that these terms represent the principal contribution in the case where the distortion is small. To demonstrate their application to a practical problem, the formulas are used to calculate the distortion and crosstalk produced when an FM signal, having a flat baseband spectrum, is passed through a single-pole filter. This example is of some current interest because such a filter has been employed in the forward path of a feedback FM receiver used for satellite communication. A number of cases are considered, and the results of the computations are plotted.

Small-Signal Behavior of Nonlinear Lumped Networks, *C. A. Desoer, K. K. Wong*—Two theorems concerning the small-signal behavior of nonlinear time-varying networks whose state equations are of the form $\dot{x} = f(x, u, t)$ are developed. The conclusions of the theorems are supported by experiments. The input is of the form $U(t) + u(t)$, where the bias, $U(t)$, is allowed to be time-varying (typically, slowly varying) and $u(t)$ is the small signal. The bias induces a moving operating point $X(t)$. Given some simple assumptions concerning the linearized small-signal equivalent circuit it is shown that provided $u(t)$ is sufficiently small on $[0, \infty)$, the state trajectory about the operating point is bounded on $[0, \infty)$ and tends to zero as $u \rightarrow 0$. The method of proof also shows that this result applies to some distributed circuits. The second theorem shows that the push-pull connection reduces the distortion due to the nonlinearities of both resistors and energy-storing elements. Numerical experiments that support the conclusions of the theory and a design procedure for nonlinear networks to be operated in the small-signal mode are described.

Standing Spin Waves in Ferromagnetic Thin Films, *J. W. Hartwell*—A theory of spin-wave resonance in ferromagnetic thin films is developed in a manner that permits a computation of the real and imaginary parts of the circularly polarized RF fields as well as the power absorption spectrum. The case of the dc field intensity applied normal to the surface of the film is considered, and the saturation magnetization is taken as constant throughout the body of the film. Variations in the dc fields near the surfaces are treated in the boundary conditions for the RF magnetization by considering the unsymmetrical nature of the exchange interaction at the surfaces. The boundary conditions are characterized by a constant for each surface that controls the degree of surface pinning in the RF magnetization. Damping is included in the formulation by means of a phenomenological constant in the spin-wave equation and by simultaneous solution of this equation with Maxwell's equations for a conductor. The results of a computer program are presented showing the roles of the various parameters in determining the power absorption spectrum and the RF fields. A comparison with experimental spectra is made, and an anomalous resonance at field intensities higher than that which is found for the principal resonance is predicted.

Thermal Annealing of Proton-Irradiated Silicon Solar Cells, *B. J. Faraday, R. L. Statler, R. V. Tauke*—Solar cells made from 1.5- and 10-ohm·cm p-type silicon, with silver-titanium evaporated electrodes, were irradiated by 4.6-MeV protons at room temperature to fluences ranging from 1×10^{10} to 1×10^{12} p/cm^2. The photovoltaic current–voltage characteristics, the photovoltaic spectral response, and the minority carrier diffusion length were studied as the solar cells were annealed isochronally to temperatures up to 600°C. The proton radiation damage annealed in two stages, the first occurring between 50°C and 150°C, and the second between 350°C and 450°C. The removal of proton damage in this manner differs markedly from the annealing reported for 1-MeV electron damage, where practically no recovery of the photovoltaic properties is observed below 350°C. At any selected annealing temperature, the 10-ohm·cm cells were observed to recover to a slightly greater degree than the 1.5-ohm·cm type of cells.

Infrared Heterodyne Detection, *M. C. Teich*—Heterodyne experiments have been performed in the middle infrared region of the electromagnetic spectrum using the CO_2 laser as a radiation source. Theoretically optimum operation has been achieved at kilohertz heterodyne frequencies using photoconductive Ge:Cu detectors operated at 4°K, and at kilohertz and megahertz frequencies using $Pb_{1-x}Sn_xSe$ photovoltaic detectors operated at 77°K. In accordance with the theory, the minimum detectable power observed is a factor of $2/\eta$ greater than the theoretically perfect quantum counter, $h\nu\Delta f$. The coefficient $2/\eta$ varies from 5 to 25 for the detectors investigated in this study. A comparison is made between photoconductive and photodiode detectors for heterodyne use in the infrared, and it is concluded that both are useful. Heterodyne detection at 10.6 μm is expected to be useful for communications applications, infrared radar, and heterodyne spectroscopy. It has particular significance because of the high radiation power available from the CO_2 laser, and because of the 8–14-μm atmospheric window.

Transmission-Line Pulse Transformers: Theory and Applications, *R. E. Matick*—The advent of fast-rise-time pulse techniques and their increasing importance brought on by high-speed microminiature circuits and the computer industry has resulted in an increased demand for pulse transformers of various types. The basic idea of constructing transmission-line-type transformers has been known and used for a number of years. However, such devices have not gained widespread usage partly because the existence of such devices is not well known, but largely because of a lack of a basic understanding of their operating principles in terms of elementary fundamentals as well as their capabilities and limitations. The basic ideas of transmission-line transformers are developed step by step from ordinary transmission-line theory, emphasizing pulse response rather than ac excitation as is usually the case. Both impedance transformers and balanced-to-unbalanced (balun) transformers, including inverters, are considered with physical insights into their operation. Several fundamental concepts are developed and explored in detail (without mathematics) since they have a strong bearing on practical applications. Also presented are new information and pulse measurements that will be useful in the design and applications of such devices, showing their capabilities and hitherto unexplored limitations, as derived from the fundamental concepts. The work reported is partly supplementary to other published work and is partly new work with the goal of providing a convenient fundamental understanding of these devices and their inherent potential and shortcomings. Although the intention is not to give a detailed design procedure, some approximate calculations and discussion of significant design criteria are included to give some indication of the goals desired.

At Ebasco, engineering is a fine art.

Scientia gratia artis. Orchestral music, with all its intricate tonal mosaics, is no longer the private preserve of the privileged. And power is one of the reasons. Power that makes large-scale outdoor concerts a reality; that unites the performing arts with the culture-hungry world via telecommunications; that reaches out to satisfy man's material needs by moving his cities as well as his mind. □ This is the kind of power you can create at Ebasco —power that demands science closely akin to art. Its tools: the most modern of technologies—nuclear engineering, composite structures, systems design and, more important, human imagination to anticipate man's needs a decade or more in the future. □ If you're an Electrical, Mechanical, Civil, Corrosion or Nuclear Engineer who would like to apply his ingenuity to attaining some of man's constructive goals, why not let us tell you how you can help achieve them at Ebasco... Engineers, Consultants and Constructors to the Electric utility industry. □ Please send your resume to Professional Employment, Department 38A-1, Ebasco Services Incorporated, Two Rector Street, New York, New York 10006. An Equal Opportunity Employer.

EBASCO

An Electric Bond and Share Company

Technical correspondence

W. W. II vignette related

I read with great interest your World War II article in November's IEEE SPECTRUM. Being at NRL in 1936, 1940, and 1941, I was a part of the early days of Navy radar that you wrote about in your article. While among the first group of officers to receive radar training, I was assigned to Pearl Harbor in November of 1941, and was aboard the *U.S.S. Pennsylvania* (BB-38) on that fateful morning (after leaving the *Lexington*).

You tell the story of the first "combat" use of the CXAM-1 on board the *U.S.S. Lexington*. Some things in the background never get into the history books, so it may be of interest to you if I relate some of the events leading up to this first success.

Just before Pearl Harbor Day, the *U.S.S. Lexington* had gone to Midway and was returning to Pearl Harbor at the time of the attack.

The CXAM-1 radar was out of commission and had defied all attempts on the part of the radiomen and electricians to repair it. When the ship arrived in Pearl Harbor, I was ordered to report immediately. It was soon determined that the antenna-train motor was defective and would have to be removed. With the help of a couple of strong sailors, the motor was removed from the pedestal and lowered 125 feet to the deck. In total darkness, and while breathing stack gases, this was quite a chore. The motor was dismantled and the trouble traced to a defective oil seal that had resulted in burning up the armature, the heavy short circuit thereby demagnetizing the permanent magnet pole pieces. Fortunately, we had a spare armature and, to provide a shunt field, I wound the pole pieces full of magnet wire. After connecting the reassembled motor to the ship's direct current, using lamp bulbs as a field rheostat, it took off and was put back onto the pedestal.

Operation was resumed by dawn and, when we left Pearl Harbor that morning, we had an operating radar. I next put the amplidyne back together and, by hooking up the control field in a Wheatstone bridge circuit, I was able to get CW or CCW rotation and stop the antenna at the bridge null for range measurement. Finally, after putting the servo amplifier back together, full normal antenna rotation was resumed. The receiver LO drifted so badly that I rebuilt it, replaced a focus pot, and tuned up the duplexer. The CXAM-1 was now operating at peak performance and we were able to track our CAP as far out as 80 miles.

Captain Sherman told me I had *one* job on board the *Lexington*, and that was to keep the CXAM-1 operating. *I* felt I had another job, and that was to teach his crew to do the same. An intensive training period followed, and by Christmas of 1941 this goal was achieved. The very successful detection of Japanese bombers and their near total destruction were certainly in large measure due to the peak performance of the CXAM-1.

One of the *Lexington* lookouts told me that on the 6th of December he had sighted a plane. Discussions in plot concluded that it could not be possible, since no other carriers were in the vicinity. I wonder what would have happened if the CXAM-1 had been in operation? I have also wondered if the *Lexington* was actually scouted by a Japanese plane. Perhaps if an oil seal hadn't failed, events might have been different.

Incidently, the September issue of *USNI Proceedings* has an excellent article on shipborne radar.

I returned to the Bureau of Ships in mid-1943 and was assigned to duty in the Radar Design Branch, NEL, and ONR until my retirement in 1956. I am now manager of the Surveillance Radar Department at Raytheon.

May I add, as a final comment on your article, "Well done."

I. L. McNally
Commander, USN (retired)
Wayland, Mass.

More nerve communication

This is in reference to Professor Wei's remarks concerning my letter, "Communicating on Nerves" (see IEEE SPECTRUM, *Technical correspondence*, pp. 164–165, August 1967).

Membrane phenomena are extremely complex, since the membrane is believed to be the site of most, if not all, living processes.[1] The ion transport process modeled by Professor Wei is a housekeeping function separate from the action potential source[2] discussed in my letter. It is the action potential and its behavior that appears to be involved in the function of communication in nerves, and not the transport process.

I would like to support Professor Wei's observation that electron transport might also play a role in membrane transport, as well as in the source of the action potential I described. The two phenomena both utilize adenosine triphosphate (ATP) as an energy source, along with most other functions in the living system. This similarity would give rise to the same types of charged particles in both phenomena.

The device-type thinking commented upon by Professor Wei arises from the necessity for describing the phenomena in both wave-mechanical and classical terms from physics.[3] As may be seen from Ref. 3 and other references quoted in my prior letter, communication phenomena in the nervous system have always been described with device-type thinking on both the microscopic wave-mechanical and macroscopic classical-physical levels. On the other hand, the housekeeping transport process described by Professor Wei has historically been described as a biological principle.

The ability to describe communication processes in nerves by wave-mechanical terms that are requisite for making artificial neural prostheses is a relatively recent event.[4] It has become possible because of research on the materials of which membranes are composed, at wave-mechanical levels. These properties of common substances are indeed startling in everyday terms. For example, the intrinsic properties of water, the behavior of which I described in my letter, reveal it to be a tetragonal polar liquid crystal with three degrees of freedom. These degrees of freedom give wave-mechanical water an infinite self-repairing or self-organizing property, revealing a source for the order–order transitions of a perpetual motion machine variety as noted by Schroedinger[5] and Leo Szilard.

Since wave-mechanical water is a polar crystal, the movement of charged particles through it seems most closely described by Feynman's polaron theory.[6] Identification of the correct theory is vital in understanding communication functions in nerves, as the entire process is a spatial-temporal phenomenon. Polaron theory shows that the comparable time delays, or temporal aspect of the phenomenon, could never be achieved with any but polar liquid crystals such as water. Efforts to duplicate the phenomenon in the nervous system with solid-state devices have failed because the numerous time delays required are too bulky for practical achievement in solid systems, where the mobility of electrons is near the speed of light in an intrinsic material. The correct time delays and thermodynamic properties are built in at the atomic and molecular or wave-mechanical levels only in water.

Robert F. Edwards
University of California
Los Angeles, Calif.

(This investigation was supported in part by a Public Health Service Special Research Fellowship.)

REFERENCES

1. Jacob, F., *Science*, vol. 152, p. 1470, 1966.
2. Chance, B., Estabrook, R., and Williamson, J. R., *Control of Energy Metabolism*. New York: Academic Press, 1965, pp. 377–379.
3. Gabor, D., "Communication theory and physics," *Phil. Mag.*, vol. 7, no. 41, p. 1161, 1950.
4. Edwards, R. F., dissertation presently in preparation.
5. Schroedinger, E., *What is Life?* New York: Anchor Books, 1956.
6. Browne, F. C., "Experiments with the Polaron," in *Polarons and Excitons*, C. G. Kuper and G. D. Whitfield, eds. New York: Plenum Press, 1963.

GVT addenda

Mr. Friedlander's coverage of the gravity-vacuum tunnel (GVT) in September's IEEE SPECTRUM is generally accurate and to the point. The following may help to clear up confusion in the minds of many IEEE members who are tracking GVT's progress.

1. Passenger comfort limit is typically 0.14 *g* rather than 1.4 *g*; GVT is adhering to this criterion whenever there are standees, and exceeding it slightly when everyone has a seat.

2. Urban applications of GVT [stage lengths of 0.5 to 10 miles (0.8 to 16 kilometers)] do not require the cross ducts, cross valves, or auxiliary vent valves described in the article. These are used only for the greater stage lengths encountered in intercity travel.

3. The "dynamic flotation" concept, described in the early stages of GVT evolution, has been replaced by spring suspension of the tubes. Springs afford many of the advantages of water flotation but are more suitable for sloping stretches, small tunnel sections, avoidance of corrosion, and predictable dynamic behavior.

4. Current thinking for the northeast corridor employs speeds not over 400 mi/h (640 km/h). Even so, thanks to GVT's remarkable acceleration characteristic, it is quite practical to travel from downtown New York to downtown Washington (or Boston) with seven intermediate stops and an elapsed time of 58 minutes.

5. Concerns about "tunneling in faulted areas" may be overstated. New York has some 80 miles (130 km) of bedrock water tunnels at depths of 400 to 1100 feet (120 to 330 meters), comparable in cross-sectional area to the GVT tunnel; these cost typically 2 million dollars a mile ($1.25 million a kilometer). [Our corresponding estimate is $3.3 million a mile ($2 million/km).] The Montreal metro's bedrock tunnels, which account for 70 percent of the entire system, were completed recently at a cost of 1.3 million dollars a mile ($800 thousand/km). These tunnels are some 40 percent larger than the GVT tunnels, even though the GVT car will be more spacious than the metro car. Machine tunneling will be dandy someday; meanwhile, conventional construction of deep tunnels is much less expensive than most cynics will admit.

It may be a sign of your members' alertness that more IEEE chapters have GVT programs than any other professional society.

L. K. Edwards
Tube Transit Corp.
Palo Alto, Calif.

Indoor climate control

I enjoyed the excellent article in the August issue of IEEE SPECTRUM by H. L. Laube on "Economics of Indoor Climate Control." However, there are some very important points he did not mention concerning air conditioning in regard to the nature of the air that is introduced.

Every time air is rolled through a duct it takes on a positive charge that has been proved to be detrimental to man and animal physiology. Various authorities from both the U.S.A. and abroad have confirmed this fact, and the affliction is usually manifested in the general statement, "I contract a sore throat and headache whenever I am confined to an air-conditioned building."

With the advent of completely enclosed air-conditioned buildings, this situation becomes serious over a period of time, medical authorities having very definitely uncovered evidence of these effects.

My great interest in this article is

attributed to the fact that my company has for years been involved in supplying material that counters this problem by introducing negative ions, upon which vertebrates and plants thrive, into the environment. It has been shown conclusively that negative ions contribute to our well-being, whereas positive ions have a very deleterious effect.

Gerald C. Ansell
United States Radium Corp.
North Hollywood, Calif.

Mr. Ansell's comments with respect to the desirability, for health purposes, of a preponderance of negative ions within enclosed spaces is appreciated. Some years ago I had the pleasure of reviewing research work being done on the West Coast in this area by one of its ardent supporters, the late Mr. Wesley Hicks.

The desirability of increasing the negative-ion count in occupied spaces has many proponents. However, it has not become general practice to incorporate this function in air-conditioning equipment.

In the case of decentralized all-electric systems, the outdoor ventilating air normally travels less than two feet through the equipment before entering the conditioned space. Although I have no knowledge of tests that have been run to confirm this viewpoint, it is entirely possible that this short traveling distance may result in the removal of fewer negative ions than is the case with systems that employ long lengths of duct work.

H. L. Laube
The Singer Company
Auburn, N.Y.

While I think highly of H. L. Laube's article "Economics of Indoor Climate Control" (SPECTRUM, August 1967), I feel that the dehumidifying systems that are usually incorporated with heating and air-conditioning plants have been completely missed.

Dehumidification can be classed as comfortable to the individual, as well as essential to certain aspects of industry. In high-class printing, for example, excess humidity can be disastrous, and, in privately owned buildings, dehumidification could be an extra luxury.

However, to include a dehumidifying system hardly entails any other extra capital, if heating has been included in the main plant. In a central system, drying of the recirculating air can be achieved by the heating coil (be it hot water or electrical), which can be in full operation while the cooling process is still going on. A conclusion is that more power is needed to overcome the drying process, which inevitably warms the recirculating air.

The extra cost involved for dehumidification could be quite considerable, especially where large areas are involved and a low humidity factor is required. It would be wise to make known this extra future running cost to the user, otherwise his calculated break-even point could well be unreal.

C. Grima
St. Paul's Press Ltd.
Zeitun, Malta

Mr. Grima is correct in stressing the importance of dehumidification during the cooling season, both for human comfort and for certain industrial processes such as color printing.

In the United States, it has long been standard practice to combine the cooling and the dehumidification functions. The decentralized all-electric system described in the paper has very favorable dehumidifying characteristics. Although some systems have a fixed amount of dehumidification, the decentralized system automatically increases its dehumidifying capability as the relative humidity within the conditioned space increases, irrespective of the amount of humidity generated in or entering the conditioned space.

For simplicity, my paper referred only to heating and cooling and not only omitted such important subjects as dehumidification, but also air filtering, circulation, and ventilation. This, however, should not be interpreted to mean that I consider dehumidification unimportant—in fact, the opposite is the case.

H. L. Laube

The boon of contention

I have just completed reading an article in the September 1967 issue of IEEE SPECTRUM entitled "Railway vs. Highway—the Zoom of Things to Come." I feel that papers of this type are of interest to the readers because they relate what technology will permit in terms of high-speed ground-transportation systems. However, introductions similar to that at the beginning of this article do absolutely nothing to mitigate the problem you have discussed—which was the lack of cooperation between the advocators of high-speed ground systems and those of highway transportation. As a matter of fact, such discussions tend to add fuel to the fire.

As a matter of interest, economic cost-effectiveness studies of even elementary sophistication clearly show that the use of private autos on freeways is the most efficient and the least costly "mass" transportation system ever devised and constructed. Comparisons with systems such as those discussed show the same results and indicate that such systems are of relatively little value in solving the ground-transportation problem.

Until transportation planners are willing to address themselves to the total transportation problem of getting persons and goods from origin to ultimate destination, very little improvement will be accomplished. Merely expending multibillions of dollars on either more freeways or on extremely expensive and ineffective high-speed ground systems will not suffice. If the IEEE wishes to stimulate thinking in this direction, it would be worthwhile to discuss other alternatives, such as the combination of publicly owned small, short-range, nonassigned vehicles to be used in conjunction with high-speed ground systems. This could provide a solution to the problem of getting to and from the bulk carrier systems. Similarly, automated highways, where vehicles are electronically controlled and linked, need to be considered.

It is obvious and absurd to even the casual observer to design a high-speed ground system without consideration of how one gets to and from it. People do not torture themselves driving in bumper-to-bumper traffic if more attractive alternatives exist. Unfortunately, for the public at least, so called "mass-transit" advocates have no interest in building such a system. A trip which requires starting in a private car, parking it, walking to a station, waiting for a high-speed bulk carrier, walking from it to other existing, low-speed public ground systems, waiting, riding, maybe transferring a couple of times, and then ultimately walking to the destination—often right past the parking lot, where he could have parked his own car—is not considered to be an advance in comfort, usually saves no time, and often requires more time than driving the entire distance in bumper-to-bumper freeway traffic.

As a final example, consider the problem of travel between the San Fernando Valley in Los Angeles and Washington, D.C., by private and public transportation, using jet travel between the respective airports. The air portion of the trip requires about 4 hours and 25 minutes, and the ground portion requires a minimum of one and one-half hours in Washington, D.C., and one and one-half to two hours in Los Angeles. These times are for covering about 30 miles on each end; however, the actual travel time of the 30 miles is 30 minutes by car or public airport transit. The balance of the time is used in waiting for baggage, buses, checking in, ticket collection, walking, and parking, none of which can be helped by any of the systems presented by your article. In fact, if any of these were used to get to the Los Angeles Airport, it would undoubtedly add another 15 to 20 minutes to the trip, since it would require at least one and possibly two additional modal changes to complete the trip.

Much more could easily be added to that already presented; however, instead of extending the length of this letter, I would be willing, if you are interested, to discuss the total ground transportation problem with someone from your staff. There are others on the U.C.L.A. campus who would similarly be interested in discussing the economics of transportation systems of the type presented in your article and their lack of desirability.

W. W. Mosher, Jr.
University of California
Los Angeles, Calif.

First of all, I do not think my introductory paragraph will have one iota of influence on the attitudes of many of the private car or mass-rapid-transit "purists." These battle lines are rather sharply drawn; and I have heard too many of Robert Moses' tirades against rail rapid transit to feel that I should apologize for my rather mild statement of the situation.

Further, I do not agree with your contention that all of the mass-transit advocates have no interest in building a balanced transportation concept, with a well-coordinated network of high-speed rail transit, expressways, and feeder bus systems. The General Electric Company, for example, has published a comprehensive booklet advocating just such systems. And a number of our major cities—notably Cleveland and Chicago—are moving in exactly this direction.

As originally scheduled, an article on the balanced transportation concept, by Dr. Edward L. Michaels of the

University of Houston, was to be published in the September issue to complement and supplement my piece. In fact, if you will refer to page 63 of my article, under the subhead "Rapid transit, U.S.A.," you may note that I touch upon the balanced transportation thesis; and it was here that I had intended a cross reference to Michaels' article. However, the publication of Dr. Michael's article was postponed until the December issue (pp. 87–91).

Gordon D. Friedlander

Solution for power failures?

Recently, there was a 90-minute electric power failure here, not unlike the one that New Yorkers experienced. The necessity for decentralization in the production of electric energy is becoming increasingly obvious, and I believe that there certainly exist ways of minimizing the costly effects of electrical failures.

All automobiles and other motorized vehicles are potential suppliers of electric power. They use 12-volt systems, primarily because of battery requirements, and cars that have ac generators usually have the filters incorporated in them. If new cars were provided with 110-volt ac generators, with separate step-down transformers and filters for the battery, much of the auxiliary equipment and accessories, including 1500 watts for a household, could then be supplied at the higher voltage. What is required is a variable-speed, constant-voltage alternator that can be used to replace the generators now in use, or to keep them separate from the filter, and sell step-up transformers for use in case of necessity.

Hospitals and industrial establishments that require more electric power should have their own generators, which could be harnessed to the output of several automobiles, via a treadmill, in emergencies. The rear wheels would be used to couple the car engine to the rotor of the generator, five cars being sufficient to supply 250 or more horsepower.

Many Canadian homes have separate outside electrical outlets for each car to prevent the engines from freezing at night. These could be used to reverse the flow of electric energy, and have the car generator supply the household. In this way, the effects of electrical failures could be minimized. The human and monetary savings will be well worth the effort.

Eric Weissman
Aspen Institute
Aspen, Colo.