VOLUME 6 ☐ NUMBER 5 IEEE spectrum MAY 1969

IEEE spectrum

features

35 **Spectral lines: Publication problems**
The most remarkable thing about the Institute's publications program is the willingness of busy experts to act as reviewers. Only rarely does a man refuse the task, and then usually because he feels he is not qualified for the manuscript in question

36 **Physical problems and limits in computer logic**
Robert W. Keyes
The reduction in the dimensions of circuits and devices via transistors and ICs may have increased logic speeds by three orders of magnitude, but the thermal problems resulting from increased power densities indicate the speed limit is only an order of magnitude away

46 **Art and technology**
II. A call for collaboration
Nilo Lindgren
Really the best reason for an engineer's entertaining the idea of joining an artist in a collaborative project is based on the pleasure principle—it will be good fun

57 **Analytical uses of energy balances**
William A. Vogely
Energy balance analyses involving the areas of forecasting, resource adequacy, resource efficiency, and consumption shift indicate that, while market growth is the major contributor to change, increased efficiency and new uses also play their parts

64 **Integrated circuits for television receivers**
Eizi Sugata, Toshihiko Namekawa
Optimum density of integration for black-and-white and color television receivers is a major goal of various groups in Japan, representing set and component manufacturers, as well as several educational institutions

79 **Magnetic traveling fields for metallurgical processes**
Yngve Sundberg
Electromagnetic stirrers have proved to be a valuable aid to the metallurgist and steel manufacturer in that they provide improved homogenization as well as temperature equalization, rapid and complete chemical reactions, and effective degassing

89 **Electricity and weather modification**
II. Plans and programs, prospects and purpose
Seymour Tilson
International research programs analogous to the IGY may help close the increasingly dangerous gap between theory and technique—and thereby help close developing gaps between weather modification programs and public credibility, priorities, and support

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

the cover

That positively charged water droplet being ejected into the bright sky by an air bubble bursting at sea is not as lonely as it looks. Collectively, such droplets carry positive charge into the air at an oceanwide rate of 160 amperes—not much, but enough along with other factors to upset classical theories of atmospheric electricity. See page 89.

departments

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

BOARD OF DIRECTORS, 1969

F. K. Willenbrock, *President*
J. V. N. Granger, *Vice President*
J. H. Mulligan, Jr., *Vice President Technical Activities*
M. E. Van Valkenburg, *Vice President, Publication Activities*
R. W. Sears, *Secretary*
Harold Chestnut, *Treasurer*
S. W. Herwald, *Junior Past President*
W. K. MacAdam, *Senior Past President*

G. J. Andrews
H. P. Bruncke
Werner Buchholz
E. E. David, Jr.
R. G. Elliott
F. A. Hawley
L. C. Hedrick
Hubert Heffner
D. M. Hodgin

D. G. Lampard
H. R. Mimno
D. C. Ports
C. F. Savage
George Sinclair

W. T. Sumerlin
R. H. Tanner
J. G. Truxal
R. P. Wellinger
J. R. Whinnery

A. N. Goldsmith, *Editor Emeritus and Director Emeritus*
Haraden Pratt, *Director Emeritus*
E. B. Robertson, *Director Emeritus*

HEADQUARTERS STAFF

Donald G. Fink, *General Manager*

John L. Callahan, *Staff Consultant*
Richard M. Emberson, *Director, Technical Services*
Ralph H. Flynn, *Director, Publishing Services*
Elwood K. Gannett, *Director, Editorial Services*
William J. Keyes, *Director, Administrative Services*
John M. Kinn, *Director, Educational Services*
Leon Podolsky, *Staff Consultant*
Betty J. Stillman, *Administrative Assistant to the General Manager*
Howard E. Tompkins, *Director, Information Services*

Committees and Staff Secretaries

Awards Board: Una B. Lennon
Educational Activities Board: J. M. Kinn
Fellow: Emily Sirjane
Finance: W. J. Keyes
History: W. R. Crone
Information Services: H. E. Tompkins
Information Systems Advisory: H. E. Tompkins
Internal Communications: Audrey L. van Dort
Intersociety Relations: J. M. Kinn
Life Member Fund: W. J. Keyes
Long Range Planning: W. J. Keyes
Membership and Transfers: Emily Sirjane
Nomination and Appointments: Emily Sirjane
Professional Relations: J. M. Kinn
Publications Board: E. K. Gannett
Sections: Emily Sirjane
Standards: J. J. Anderson
Student Branches: Emma White
Technical Activities Board: R. M. Emberson
Translated Journals: A. A. McKenzie

IEEE SPECTRUM is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: 345 East 47 Street, New York, N.Y. Cable address: ITRIPLEE. Telephone: 212—752-6800. Published at 20th and Northampton Sts., Easton, Pa. 18042 and additional offices. **Change of address** can be made effective with any particular issue if received by IEEE Headquarters by the 10th of the preceding month. **Annual subscription:** IEEE members, first subscription $3.00 included in dues. Single copies $1.00. Prices for nonmember subscriptions available on request. **Editorial correspondence** should be addressed to IEEE SPECTRUM at IEEE Headquarters. **Advertising correspondence** should be addressed to IEEE Advertising Department, at IEEE Headquarters. Telephone: 212—752-6800.

Responsibility for the contents of papers published rests upon the authors, and not the IEEE or its members. All republication rights, including translations, are reserved by the IEEE. Abstracting is permitted with mention of source.

Second-class postage paid at Easton, Pa., and additional offices. Printed in U.S.A.

BPA

OTHER IEEE PUBLICATIONS: IEEE also publishes the Proceedings of the IEEE, the IEEE Student Journal, and more than 30 Transactions for IEEE Groups with specialized interests within the electrical and electronics field. Manuscripts for any IEEE publication should be sent to the editor of that publication whose name and address are shown on page 10A of the January 1969 issue. When in doubt, send the manuscript to E. K. Gannett, Editorial Services, at IEEE Headquarters, for forwarding to the correct party.

The newly developed narrow bandwidth elliptic-function filter is discussed in this article by Bob Sleven, Irv Rubinstein and Fred Hinte, members of our Filter Technology and Applications Department. If you are interested in the technical details of this development, write us and we will send them to you.

Narrow Bandwidth Microwave Elliptic-Function Filter

The elliptic-function filter has been receiving considerable attention recently because of its extremely sharp selectivity. The realization of this sharp selectivity has been reported in the literature for wide bandwidth applications (references 1, 2, 3, and 4). However, it is for narrow bandwidths (from 5 percent to a fraction of a percent) where loss becomes significant that the elliptic-function filter offers one of its most important advantages over other filter types—lower loss. This article discusses the newly developed narrow-bandwidth elliptic-function filter first reported at the 1969 IEEE G-MMT International Microwave Symposium (reference 5).

The design of this filter, which is based on the lowpass prototype, is simple to obtain and is readily realizable in printed or other TEM transmission line. The low-loss feature of the elliptic-function filter has been verified both theoretically and experimentally.

Figure 1 shows a theoretical elliptic-function response. It is the stopband ripple that makes this response shape different from the conventional ones. It is also this factor that led us to believe that the elliptic-function filter might have lower loss. For example, if in a practical application the specified maximum rejection were 60 dB, the elliptic-function filter could be designed to have a stopband ripple with a minimum attenuation of 60 dB. The conventional filter could also provide 60 dB attenuation at the required frequency, but as the frequency deviates further from center frequency the attenuation continues to increase. For this example, attenuation in excess of 60 dB represents an overdesign and is paid for with added insertion loss. The data listed in the table below shows that the difference in insertion loss between elliptic-function and conventional filters is significant. The characteristics of three different type seven-resonator filters are compared for a ten-to-one ratio of unloaded Q (Q_u) to loaded Q (Q_L). This ratio could represent a 1% fractional bandwidth and a Q_u of 1000.

	Elliptic-function	Equal-element	Chebyshev
Loss (dB)	2.70	4.05	4.35
BW 60 dB BW 3 dB	2.19	2.44	2.38

Comparison of Filter Loss and Selectivity

In addition to having the lowest loss, the elliptic-function filter also provides the sharpest selectivity.

The first step in the design of the elliptic function bandpass filter is to convert the low-pass prototype to a bandpass prototype with bandpass resonators tuned to f_o and pairs of bandstop resonators tuned above and below f_o. The bandstop resonators provide the stopband ripple shown in Figure 1. To convert the bandpass prototype to a realizable microwave circuit, several steps are involved. In the prototype there are both shunt and series connected resonant circuits. In a microwave transmission line filter only one type of connection can be used. This is accomplished with impedance inverters. To set the impedance of the various circuits at a reasonable level, input and output capacitive gap impedance inverters are used.

At this point the circuit is realizable on paper but not in practice. The resonators of each bandstop pair are connected in the circuit at the same point. When the physical circuit was built, interaction or stray coupling occurred between these resonators and the elliptic-function response could not be achieved. The solution to this problem was a circuit suggested by Geffe (Ref. 6) many years ago for a different purpose. Using the Geffe transform the interferring resonators could be separated but still provide the desired transfer function.

With this last problem solved, a large number of N=3 and N=5 elliptic-function filters have been successfully fabricated. Although the measured losses have been slightly higher than calculated, they have in all cases been substantially lower than achievable with conventional filters of the same rejection bandwidth.

The printed microwave channelizer shown in Figure 2 is an example of a practical application of elliptic-function filters. In this case, a loss savings of about 3 dB resulted from the use of elliptic-function filters.

The work described in this article was sponsored by Rome Air Development Center, Rome, New York.

References

1. M. C. Horton and R. J. Wenzel, "Realization of Microwave Filters with Equal Ripple in Both Pass and Stop Bands," presented at the International Symposium on Generalized Networks, Polytechnic Institute of Brooklyn, Brooklyn, New York, April 1966.
2. R. Levy and I. Whiteley, "Synthesis of Distributed Elliptic-Function Filters from Lumped-Constant Prototypes," IEEE Transactions on Microwave Theory.
3. M. C. Horton and R. J. Wenzel, "The Digital Elliptic-Filter—A Compact Sharp-Cutoff Design for Wide Bandstop or Bandpass Requirements, "IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-15, p. 307-314, May 1967.
4. J. D. Rhodes, "The Stepped Digital Elliptic Filter," to be published in the IEEE Transactions on Microwave Theory and Techniques.
5. I. Rubinstein, R. L. Sleven, A. F. Hinte, "Narrow Bandwidth Elliptic-Function Filters," 1969 IEEE G-MTT International Microwave Symposium, Dallas, Texas, May 1969.
6. P. Geffe, "Simplified Modern Filter Design," John Rider Publishers, Inc., New York, 1963.

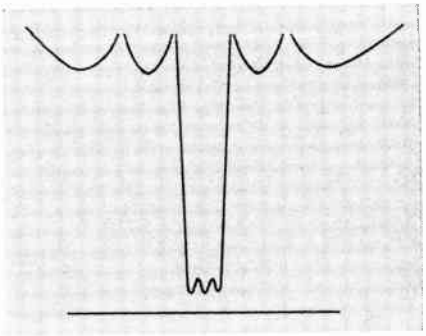


FIGURE 1. Elliptic-function Filter Response.

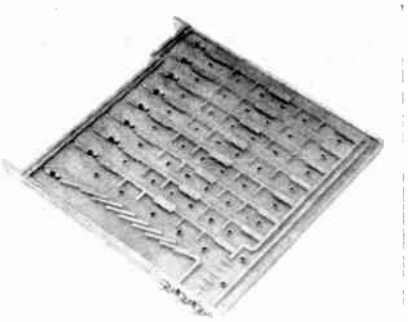


FIGURE 2. Elliptic-function Channelizer.

A complete bound set of our thirteenth series of articles is available on request. Write to Harold Hechtman at AIL for your set.

Forum

Readers are invited to comment in this department on material previously published in IEEE SPECTRUM; on the policies and operations of the IEEE; and on technical, economic, or social matters of interest to the electrical and electronics engineering profession.

Computer-controlled traffic

Gordon D. Friedlander's February article, "Computer-Controlled Vehicular Traffic," concentrates on the use of computers as means for improving the efficiency of traffic flow. While a well-designed computer system can undoubtedly provide for more efficient *flow* of traffic it must by necessity falter when confronted with an impossible storage problem. When an excessive number of vehicles converge on a downtown area with a limited number of parking spaces the traffic problem becomes severe and its resolution lies outside the realm of computer control. Shortage of parking spaces was the subject of a *Wall Street Journal* article on April 24, 1967.

Parking congestion has been caused by the provision of city-subsidized free or low-cost parking spaces. Eliminating the subsidy and charging for parking spaces on the basis of free-market prices would resolve in addition to the parking congestion many other big-city problems. Wolf Dreverman has focused on parking and traffic related city problems in an article in the German news magazine *Der Spiegel*, no. 16, 1967.

Dreverman points out that low cost curbside and city-lot parking has attracted a large number of cars and along with them pollution, noise, parking, and traffic problems. In addition, the "availability" of low cost parking has diverted many passengers from mass transportation systems with the consequent atrophy of rapid transit.

The suggested solution is to raise parking fees to the level at which approximately 10 percent of the parking spaces remain vacant. This constantly available storage buffer would eliminate cars searching for parking spaces from streets meant for traffic flow. Putting the burden for the cost of parking on the user, instead of the taxpayer, by raising parking fees to the level determined by supply and demand would eliminate the parking problem. The increased and true cost of commuting to the city would be apparent and would become a revitalizing stimulant for mass transit systems. To the extent that parking prices would be higher than necessary for the long run they would stimulate the construction of parking facilities by private enterprise. Once the excess demand for parking spaces has been eliminated computers can be used to control both the flow and the storage of automobiles.

Ernst F. Germann
Austin, Tex.

Engineering ethics

This letter is prompted by our recent involvement in a day-long meeting of the Ethics Committee of the Engineers' Council for Professional Development, where some 25 representatives of the ECPD member societies, and other concerned individuals, reviewed the attitudes of the societies with respect to the ECPD Canons of Ethics of Engineers.

It was reported that ten of the 12 member societies of ECPD have adopted the Canons either in whole or in part, one society has the matter under active consideration, and only one has tabled action on the question of adopting the Canons. We regret that this last society is the IEEE; and we also regret that IEEE has recently reduced its Committee on Professional Relations to an *ad hoc* basis.

The ECPD committee recognized the need to keep the Canons under review and to consider strengthening them if it was judged that this could result in a statement more in keeping with the expanded role of technology in our modern society. To this end a subcommittee was appointed to recommend on a course of action. The IEEE representative is on the subcommittee.

In order to stimulate the IEEE input into this important matter we hope to arrange a session on ethics at the forthcoming NEREM conference in Boston in November.

We recognize that there is a wide diversity of experience in regard to professional ethics between the various segments of the engineering profession. Civil and consulting engineers have traditionally been alert to the need for an open espousal of an ethical code. The American Institute of Electrical Engineers played a leading role in establishing a code for the engineering profession as a whole. The main reason that a profession adopts a code is to place on record its dedication to the public welfare—which alone makes it deserving of the public trust.

A second aspect of the codes of conduct that the professions have adopted is the application of moral principles to the specific relationships and problems that relate to the profession concerned. A review of the codes convinces one that such statements are not trite, but do indeed provide helpful guidance particularly to the young engineer, or professional person. The codes also give formal support to the fact that a professional man, even as an employee, has the right and the duty to point out unethical practices whenever they occur.

In many ways the public is crying out for ethical leadership. Reflecting a sensitivity to the desires of our society, a research stoppage was held at M.I.T. on March 4, "To convey to our students the hope that they will devote themselves to bringing the benefits of science and technology to mankind."

Garret Hardin has recently pointed out [1] that there is NO TECHNICAL SOLUTION to the problem of over population. We are faced with moral problems which call for nations and individuals to act for the public good.

AN IEEE CONTINUING EDUCATION SERVICE

COMPUTER PROGRAMMING FOR ELECTRICAL ENGINEERS

COURSE DESCRIPTION: The lectures will provide a tutorial description of the use of digital computer programs as an aid to the practicing EE in solving many of his day to day problems.

FIRST DAY:	Introduction to Programming Digital Computers The Basics of a Programming Language Special Tools of the FORTRAN Language Developing an Algorithm
SECOND DAY:	Some Advanced Tools of FORTRAN Calculations and Data Plotting With A Digital Computer Use of Library of Subprograms

Problem solutions include velocity of an electron beam, calculation of resultant EMF, a nonlinear electronic circuit, transient circuit response, Newton-Raphson technique, numerical integration, functional plotting, general network analysis, and others.

COURSE PURPOSE: This course is intended to equip attendees with the knowledge to begin to write some of their own programs and to discuss programming problems with experienced programmers.

COURSE MATERIALS: Those to be furnished include IEEE Study Outline and Course Notes with bibliography. Student should purchase "FORTRAN With Engineering Applications" by D. D. McCracken, John Wiley and Sons.

COURSE INSTRUCTORS:

Dr. Monroe M. Newborn—B.E.E. Renssalaer Polytechnic Institute
M.S. and Ph.D. Ohio State University

Bradley Bealle—BA Columbia University

PROBLEM SOLVING FOR ELECTRICAL ENGINEERS USING TIME SHARED COMPUTERS

COURSE DESCRIPTION: The lectures will consist of a tutorial description of how to use time shared computers to solve practical EE problems. Stress will be on "conversation" with the computer as a powerful engineering aid.

FIRST DAY:	Advantages of Time Sharing a Digital Computer Operating Statements of a Time Shared Computer System Problem Solution Methodology Mathematical Operations in Time Sharing
SECOND DAY:	Time Sharing Techniques for: Circuit Design and Analysis System Design and Analysis Statistical Analysis Available Time Shared Systems

COURSE PURPOSE: This course is intended to equip attendees with the knowledge to begin to solve some of their own practical problems on a time shared computer system.

COURSE MATERIALS: Those to be furnished include a Study Outline and Course Notes with bibliography, and a reading list.

COURSE INSTRUCTORS:

Paul D. Oyer—B.S. and A.M. Indiana University

John T. McAuley—BEE College of the City of New York
MSEE Polytechnic Institute of Brooklyn

ADDITIONAL COURSE BENEFITS

Learn how to—

- —Improve techniques of problem solving
- —Talk the same language with computer personnel
- —Recognize a problem the computer should solve
- —Understand FORTRAN
- —Use a computer from your own office
- —Use newer and faster methods
- —Solve actual EE problems on a time sharing computer

June 10 & 11, 1969
Chicago, Illinois
O'Hare Inn
Mannheim & Higgins Road
Des Plaines, Illinois

To: Education Registration IEEE, 345 East 47 St., N.Y. 10017

ENCLOSED IS MY CHECK (OR MONEY ORDER) TO COVER MY REGISTRATION IN THE COURSE.

CHECK ONE

☐ Problem Solving for Electrical Engineers Using Time Shared Computers

☐ Computer Programming for Electrical Engineers

Name (Please print or type)

Position or Title

Address Phone

City State Zip Code

Please make check payable to IEEE.

FEE: $100 Members; $115 Non-members (for either course). Non-members will have the option of applying the differential toward IEEE membership dues if application is received in IEEE within three months following the completion of the course. Fee includes packaged Study Outline, Course Notes and lunch each day.

Up-to-date

from

Van Nostrand Reinhold Co.

THIN FILM TECHNOLOGY

Bell Telephone Laboratory Series

Prepared under the direction of Robert W. Berry, Peter M. Hall, Murray T. Harris.

This important new book brings you the essential background you need to execute a Thin Film program or pursue other developments in this field. The book guides you through the various methods by which films are formed . . . through the properties the materials require for Thin Film applications . . . to the many practical considerations in the production of electronic circuitry. Of immediate value to anyone concerned with this technology, the book *must* be examined to be appreciated.

706 pp., 6"x9", $15.00

HANDBOOK OF MATERIALS AND TECHNIQUES for VACUUM DEVICES

By Walter H. Kohl

This new, comprehensive successor to the "bible" of the electronics industry, "Materials and Techniques for Electron Tubes" provides necessary information on all materials essential to the vacuum device industry. New chapters cover titanium, zirconium, hafnium, and their alloys; structures and coatings; secondary emission; and voltage breakdown. High-temperature properties receive added attention. Metallurgical considerations are stressed.

636 pp., 7"x10", $30.00

For 10 Days Free Examination

Mail This Coupon Today

Van Nostrand Reinhold Company, Dept. S-569
120 Alexander Street,
Princeton, N.J., 08540

Please send me books checked below for 10 days free examination. At the end of that time, I will send you my remittance or return the book(s) and owe nothing.

☐ **THIN FILM TECHNOLOGY, by Berry, Hall and Harris, $15.00**

☐ **HANDBOOK OF MATERIALS AND TECHNIQUES for VACUUM DEVICES, by Kohl, $30.00**

Name

Address

City State Zip

Circle No. 5 on Reader Service Card

Hardin says, "We must exorcise the spirit of Adam Smith. . . . who popularized the idea that an individual who intends only his own gain is led by an invisible hand to promote the public interest." The morality of some acts is a function of the state of development of the society at the time it is performed. Using our rivers as cesspools does not harm the general public under frontier conditions—the same behavior is intolerable in a metropolis.

Thus, ethical standards are vastly important and cannot be taken for granted. The open promulgation of ethical standards by professional men is demanded by the public and is an essential part of the application of technology in our times.

E. Howard Holt, Chairman
ECPD Committee on Ethics
El Paso, Tex.

Philip L. Alger, Member
ECPD Committee on Ethics
Schenectady, N.Y.

REFERENCE

1. *Science*, Dec. 13, 1968, pp. 1243–1248.

Corrections noted

Two errors have been revealed concerning an article in April's IEEE SPECTRUM by Timothy J. Healy entitled "Convolution Revisited."

On page 91, the equation in the left-hand column contains two sets of inequalities, the second of which is incorrect. The equation should read

$$g(x) = e^{-x} \qquad 0 \leq x \leq \infty$$
$$= 0 \qquad -\infty \leq x < 0$$

On page 93, the first equation after Eq. (21) should read

$$H(s) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(\lambda) g(x-\lambda)\, d\lambda \right] e^{-i2\pi xs}\, dx$$

CHANGE OF ADDRESS

Change of address can be made effective with any particular issue if received by IEEE Headquarters by the 10th of the preceding month. Please attach magazine address label here, print your new address below. Be certain that you include your IEEE member number and new ZIP Code. Mail this form to: Coding Department, The Institute of Electrical and Electronics Engineers, Inc., 345 E. 47 St., New York, N. Y. 10017.

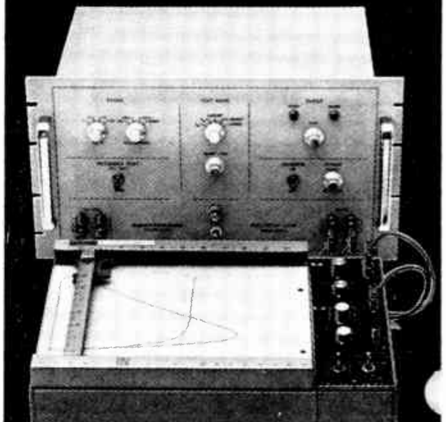
Name

IEEE Member Number

Address

City State ZIP Code

Electronic Load tests solar cells and panels automatically, delivers X-Y readouts...



here.

No other testing device offers the accuracy and simplicity of our solid-state Model D-550 Electronic Load in testing solar cells and solar cell assemblies to 250 watts automatically. Use it as a portable laboratory unit or with a standard 19" panel for rack mounting to obtain readings of open circuit voltage, short circuit current and voltage versus current and relative power versus voltage curves. Complete with Moseley 135 X-Y recorder or may be ordered without recorder. Model D-550-10 with 500 watt capacity also available.

Write today for specifications, prices and further information. Spectrolab, 12484 Gladstone Ave., Sylmar, Calif. 91342. Telephone (213) EMpire 5-4611.

Spectrolab

A Division of textron Inc.

Circle No. 6 on Reader Service Card

VOLUME 6 □ NUMBER 5 IEEE spectrum MAY 1969

Spectral lines

Publication problems. Last month's editorial ran lightly over the present pattern of journal publication in the IEEE, remarking incidentally that (contrary to rumor) unsolicited manuscripts are very much welcomed by IEEE SPECTRUM. Having outlined the scope of the program, we can take a look this month at two of the problems it generates—the selection and the editing of so many pages of highly technical material.

The selection process varies with the journal. The most remarkable thing about it is the willingness of busy experts to act as reviewers. Only rarely does a man refuse this necessarily thankless task, and then usually on the ground that for the manuscript in question, he is not qualified. Without this spirit of willingness to review, the whole publication program of the Institute would fall apart.

A manuscript submitted to the PROCEEDINGS is sent to the member of the Editorial Board who is best acquainted with the field that it treats. He suggests some reviewers well versed in that particular subject. The Managing Editor collects critiques from three of these. If—as usually happens—there is a consensus, he acts on it. In difficult cases, he appeals to the member of the Editorial Board for help. The decision may be to accept, reject, or ask for revision: when the manuscript is of good technical quality but is of interest only to specialists in a single area served by a Group, it is referred to a TRANSACTIONS. For an invited paper, the author himself may nominate reviewers.

Letters for the PROCEEDINGS are sent to but one reviewer, who gives no critique; he says Yes or No. Because he gives only a binary judgment, he can give it with a minimum of delay—often by telephone. On the few occasions when the author of a Letter has appealed from the lone reviewer's judgment to the Editorial Board, it has been clear that the reviewer was right. One such appeal that I was involved in took excruciating amounts of time for all concerned, and the result was a published Letter that just was not worth the man-hours of nonauthor time that went into it. If the author of a Letter does not compose a terse and clear account of what he has to say, and does not find and acknowledge the accomplishments of those who have worked the area before he did, he must not expect that the editorial people will do these things for him: his Letter should be rejected, and the chances are now overwhelming that it will be.

A manuscript submitted to IEEE SPECTRUM is forwarded to the Editor. For a manuscript that shows enough promise, he selects a reviewer or asks a member of the Editorial Board to do so. When the review is unfavorable but the Editor feels that the manuscript may nevertheless have some merit, he sends it to a second reviewer. Two adverse reviews result in rejection or in a request for extensive revision.

If the review endorses the technical accuracy of the manuscript (perhaps with some reservations that call for revision) the Editor decides whether the article now seems to have sufficient appeal for SPECTRUM. This is admittedly a guessing game, and the odds are strongly influenced not only by the warmth or coolness of the review, but also by the other manuscripts that are on hand. The incumbent Editor rejects the concept of "the average reader" of SPECTRUM; he pictures a population of nonaverage readers, and tries to maximize the probability that each of them will find something of interest in every issue.

After selection comes editing. About 30 years ago, public-school teaching of English in the U.S. gradually began homing in on a newly lighted beacon: "The principal objective of creative expression in school is the individual pupil's joyous realization of the values of his own experiences. . . Some study of principles and even practice exercises are needed, but such academic procedures must be kept in their proper place as accessories to the life experience."[1] The academic procedures have been kept in their place so thoroughly that the U.S. engineer under 40 who writes well is a rarity.

In science and in engineering, writing well means writing clearly. The cardinal sin is ambiguity, and U.S. engineers wallow in it. Those who are conscious of it try to clean things up with a lather of redundancy. To provide more effective kinds of purification, as well as to give aid to authors from other countries, the Institute has a large staff of copy editors. The problem arises: How do you train a girl with a major in English to eliminate the ambiguities in a manuscript on circuit theory, without making the text unambiguously wrong? It is a measure of the tenacity of the Editorial Department that this problem has been mastered.

There remain the problems of production, distribution, and indexing of the journals. These are nearly the same for all the journals; they will be surveyed in next month's "Spectral lines," which will deal also with the special problems generated by the TRANSACTIONS.

J. J. G. McCue

---

[1] Curriculum Commission of the National Council of Teachers of English, *An Experience Curriculum in English.* New York: D. Appleton-Century Company, Inc., 1935.

Physical problems and

Transistorized computer logic has made steady progress toward higher speeds by reducing the dimensions of circuits and devices. However, even though circuit and device speeds have increased by three orders of magnitude, voltage, current, and power levels have remained about the same. Power dissipation at increasingly higher current densities seems to be leading to difficult thermal problems that will eventually limit the progress of logical circuitry toward higher speeds. This article examines these physical problems and limits, and offers the most directly effective solution to the dilemma—lower operating temperature.

The modern high-speed computer was made possible by the introduction of the transistor into electronic technology. The replacement of tubes by transistors in electronic computers about 1957 immediately increased the logical power of computers by an order of magnitude, and steady progress has ensued since that date. The transistor is much faster than the tube and, in addition, uses much less power. Figure 1 shows how logic delay has been decreased by technological progress during the years since the introduction of the transistor into computer technology. Delay, the time between the application of input signals to a logical circuit and the emergence of the logical result, is a measure of circuit speed.

The lower power consumed by the transistor is at least as important as its greater speed. It permits many more elements to be packed into a given space and thus makes much larger computers possible. The combined increase in speed and in size, as measured in terms of number of circuits, has led to enormous advances in the rate at which a computer can carry out instructions.[1]

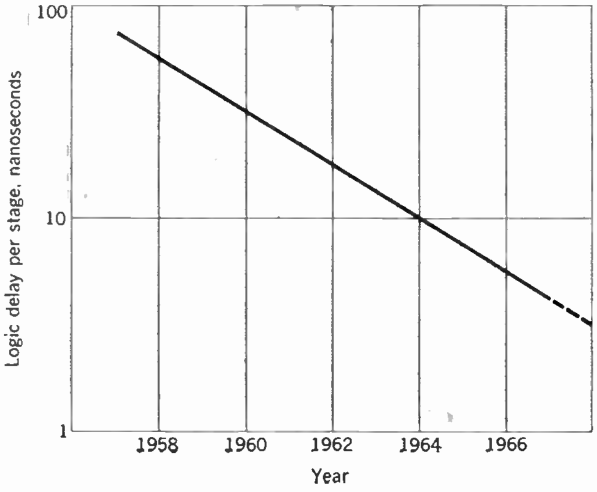
Nevertheless, there are still many things that computers cannot do. There are problems so enormous that the most powerful machines cannot solve them on a useful time scale.[1,2] Other problems, whose size cannot be measured so quantitatively, seem far too complex for the computer, even though they are susceptible to attack by the human mind. Although conceptual advances will undoubtedly contribute to the solution of these problems, there is clearly a need for far more powerful machines. Thus we are led to ask: How long will progress such as that illustrated in Fig. 1 continue? What limits are there on the eventual performance of computers?

Fundamental limits

We would like to think that we could come up with an answer to these questions, which is in some sense "ultimate." I am not convinced that there is any such thing as an "ultimate" limit. In fact, finding ways to surmount those obstacles that, at the present, seem to be the limits is what technology is all about. For example, some so-called limits simply boil down to an assumption that we cannot fabricate anything smaller than a wavelength of light. Obviously, as technology gives us the prowess to fabricate things on smaller distance scales—for example, by replacing the use of optical methods of defining structures by electron-beam methods[3]—the distance scale will change and therefore the limits will also change.

We would like to find limits that are independent of technology and are derived from fundamental principles, such as quantum mechanics, thermodynamics, and information theory. A primary example of a limit of this kind is the requirement that energy be created in a logical operation. This is a result of the fact that

FIGURE 1. Logic delay of fastest available computers versus the year.



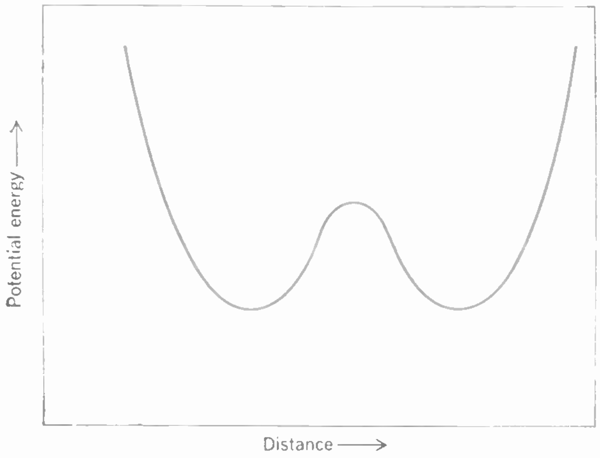
limits in computer logic

The ever-increasing computer speeds and ever-diminishing circuit dimensions brought about by the technological explosion might lead one to believe that the process could continue forever; however, obstacles will be encountered

Robert W. Keyes *IBM Thomas J. Watson Research Center*

logic destroys information. Fewer bits of information come out of an AND gate, for example, than go in. The applicability of this idea to computers was first recognized and justified in some detail by Landauer.[1] An example of Landauer's ideas is illustrated by Fig. 2, which shows a bistable potential well that contains one particle. A bit of information can be represented by a statement of which half of the well contains the particle. If we don't know which well the particle is in, the system has an entropy $k \log 2$, where k is Boltzmann's constant. If, now, we perform some logical operation that results in the particle being in a particular, known well, then the entropy of the system becomes zero. The destruction of the entropy of this well must be accompanied by a creation of heat $kT \log 2$. All of this has been argued and illustrated in much greater detail by Landauer. However, even this limit is, in a sense, technological, because it can be lowered by cooling the computer.

FIGURE 2. Information can be stored in a particle contained in the potential well shown. The information consists of a statement of which half of the well the particle is in.



A very similar result can be obtained by considering the problem of thermal noise in the computer. Electrical noise per unit of bandwidth is $4kT$. The bandwidth must, however, be something of the order of magnitude of the reciprocal of the delay time. Thus if one assumes that the power required to operate a logic circuit must be large compared with the thermal noise power, one again obtains the result that the energy per logical operation must be of the order of kT.

Another fundamental idea is that, if a meaningful signal is confined to a time less than τ, it seems reasonable to expect that the energy used must be larger than the uncertainty energy h/τ. The same magnitude of energy results if quantum noise rather than thermal noise is regarded as setting the power level. This quantum limit, like the thermal limit, is quite small for any kind of a system we might think of today.

Voltage level and impedance level

Things like limits such as those that have just been described are a great many orders of magnitude away from any actual computer element and don't provide information that is useful for understanding performance limits of real computers. The basic physics that I think really determines the power level of modern electromagnetic computers is comprised of the following two observations.[5]

First, logical functions are basically nonlinear functions, and computer circuitry depends on nonlinear phenomena to perform them. Now, nonlinear electrical phenomena only occur in devices—in particular, at semiconductor junctions—when the electric poten-

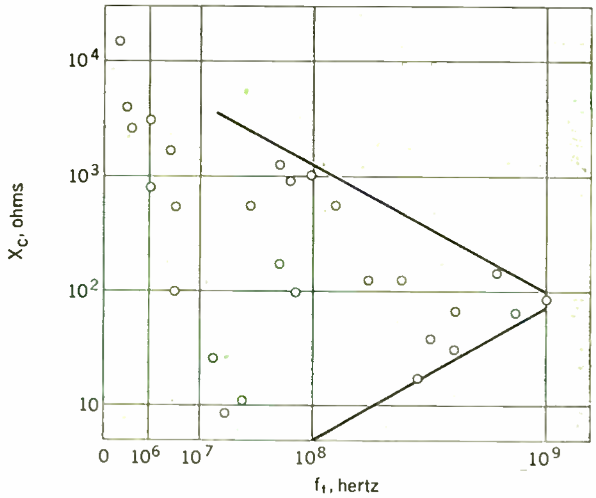


FIGURE 3. Collector impedance as a function of characteristic frequency for a selection of transistors showing the convergence of the impedance level to $0.3\,Z_0$. (Data from Johnson.[12])

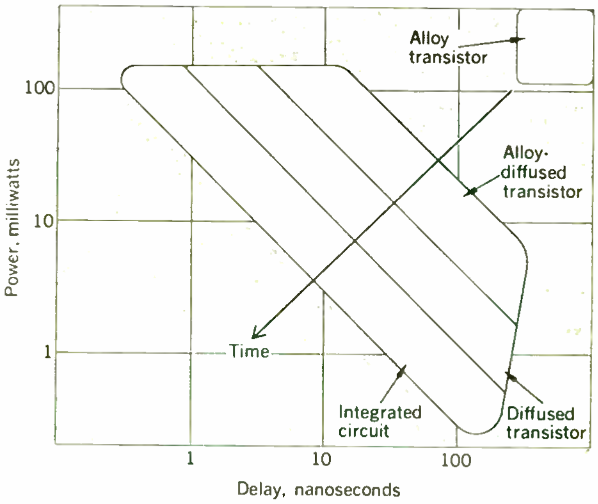


FIGURE 4. Power dissipated in logical circuitry as a function of logic delay. The arrow shows the direction of evolution of this plot with time as newer technologies are introduced into logical circuitry.

tial applied to them is changed by an amount large compared with the thermal voltage, kT/q. Here, q is the charge of the electron. Although this statement is well known, intuitively apparent, and widely accepted, it is hard to justify rigorously. The most relevant theoretical result that I know of here is one of Bernard and Callen,[6] adapted to the present case by Gunn.[7] These authors have shown that there is a proportionality between the nonlinearity of a device, as measured by the second derivative of the current with respect to the voltage, and the shot noise of the device. That is, if the current is expanded as a function of voltage to the first order of nonlinearity,

$$i = gv + \tfrac{1}{2}hv^2$$

then $h/g = \theta q/kT$, where θ is the fraction of full shot noise exhibited by the device.

This result shifts the burden of the limit upon the nonlinearity to shot noise. It is apparent that the nonlinearity and the shot noise could be increased if the charge per particle q could be made larger than the charge on an electron. Thus, larger nonlinearity might be expected from effects involving cooperation between electrons. However, even superconductive Josephson devices, which are based on effects that involve cooperation between an enormous number of electrons, require voltages greater than k/q times the superconducting transition temperature and, therefore, greater than the temperature.[8]

It is possible to find theoretical mechanisms that yield very nonlinear dependence of current on voltage.[9] However, the current in semiconductor junctions can contain contributions arising from many mechanisms and the mixing of the various contributions washes out the nonlinearity of any one of them. Series resistance and physical inhomogeneity further degrade the nonlinearity.[9–11] Although values of h/g slightly greater than q/kT have, in fact, been observed in certain tunnel diodes,[9] it seems clear that the attainment of nonlinearity on a scale less than the thermal voltage will be quite difficult. The basic feature is that a clearly defined nonlinear response to an electric signal requires that the energy a single electron can obtain from the signal voltage must be larger than the thermal energy. Even nature apparently finds something of the sort to be true; the voltage level at which neurons operate is about $3kT/q$. A much larger factor of safety seems necessary in most logical circuitry, perhaps because of the great redundancy of the physiological system. In any case, it appears that kT/q is a realistic limit for the scale of nonlinearity of computing elements for the foreseeable future.

The second observation that leads to the limit being described is that logic circuits have to communicate with one another at very high rates of speed. Communication takes place via assemblies of conductors that must be thought of as transmission lines. Efficient transfer of signals between elements of logic and the transmission lines demands that their impedances be similar. The impedance of a transmission line, however, is always similar to the basic unit of impedance Z_0, the impedance of free space, which has a value of 377 ohms. As an example, the approach of high-speed circuity to an impedance level of the order of magnitude Z_0 is further illustrated in Fig. 3. Johnson[12] recently defined and calculated a characteristic impedance X_c for a selection of high-speed transistors (X_c is defined as the impedance of the collector-base capacitance at the transition frequency f^t, at which the gain goes to unity). The values of X_c are plotted against f_t; note the convergence of the impedance level toward $0.3\ Z_0$ with the increase of f_t.

Given the voltage level and the impedance level, the power level is implied:

$$P_0 = (kT/q)^2/Z_0 \qquad (1)$$

Putting the fundamental constants into this formula and setting $T = 300°\text{K}$ gives a power level of about 2 microwatts. Obviously, to obtain very large, reliable, nonlinear effects, it is necessary to use voltages an

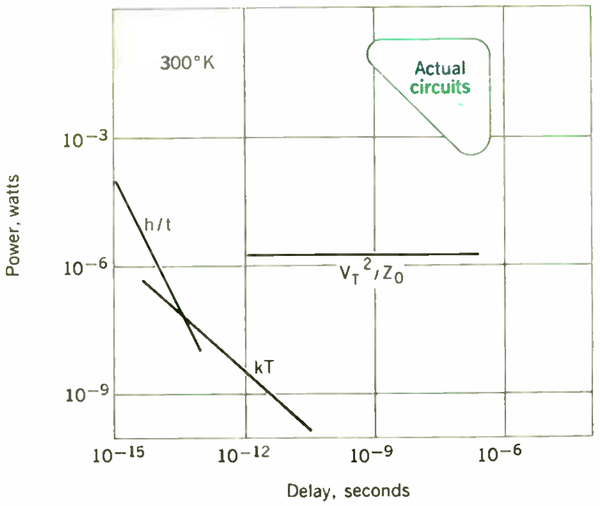


FIGURE 5. Power dissipation vs. delay per logical stage. The performance of actual circuits is reproduced from Fig. 4 in the upper-right-hand corner. Various limitations on the P–t_d relation derived from physical laws are also shown (V_T is the thermal voltage kT/q).

order of magnitude or two larger than the thermal voltage. In addition, it is hard to make transmission lines with an impedance equal to that of free space; in particular, many modern circuit technologies are most suitable for producing lines with a characteristic impedance considerably less than that of free space. Further, circuits that perform logical operations contain several nonlinear elements. Thus we may expect the actual power level to be several orders of magnitude greater than the number derived from Eq. (1), and this is indeed found to be the case. Similar thoughts lead to a current level,

$$i_0 = kT/qZ_0 \tag{2}$$

The most important feature of (1) is not the exact value of the power it gives, but its implication that the power is independent of the speed or delay. It asserts that the power dissipated per logical circuit in high-speed logic is a relatively fundamental quantity and cannot be expected to change with the progress of technology at a fixed temperature. This result is compared with the history of transistorized logic in Fig. 4. Here, power per logical stage is plotted against time delay per logical stage. Four different generations of transistor technology are represented. Each generation resulted in faster logic. The earliest transistorized computers used the alloy transistor. This was soon replaced by the alloy-diffused transistor and then by the all-diffused transistor. The latest generation of logic is based on integrated circuits. The point that Fig. 4 illustrates is that the power dissipation of the fastest logical circuits hasn't changed much over the years, even though the speed has increased by almost three orders of magnitude. Thus, the suggestion of physical theory that the power dissipation of the fastest logical circuitry at a fixed temperature is a physical constant is confirmed by computer history. Figure 4 also illustrates that performance can be traded for power within a given technology. Figure 5 depicts the relation of the facts in Fig. 4 to the various theoretical limits.

The same arguments apply to the current, which is something like the thermal voltage divided by the transmission-line impedance. Extrapolating to the future, we may expect both power levels and current levels to remain about the same as they are now.

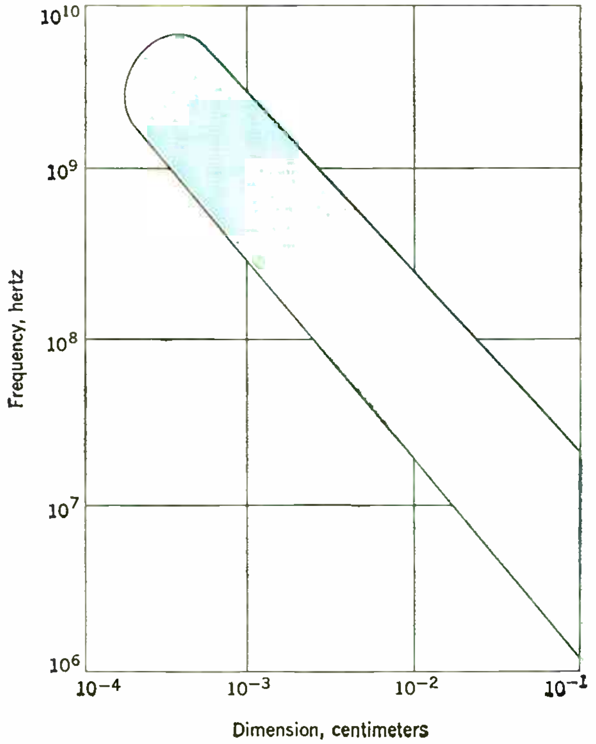
Speed and size

The progress toward higher logical speeds that is apparent in Fig. 4 has been attained essentially by making transistors smaller. One demonstration of this is given in Fig. 6, in which the maximum frequency of oscillation of various transistors has been plotted against some dimension. These two quantities are just about inversely proportional to one another. Although the frequency here is plotted against just one dimension of a transistor, there has been a tendency for all the linear dimensions to scale in proportion. The area has been rapidly reduced and, therefore, the current density and the power density in logic transistors have been rapidly increasing. It appears that this increasing current density and lower power density will create problems that will limit the speeds attainable by continued development along present lines.

The most obvious of these problems are as follows. First the emitter efficiency decreases. Another problem is the base-stretching effect.[13] Many of these problems that limit current density have been alleviated by heavier doping of the semiconductor. In fact, the trend to smaller and faster structures has been accompanied by a trend to heavier doping.

Since the operating voltage of logic circuitry is

FIGURE 6. Characteristic frequency of transistors vs. a linear dimension. Because various technologies have been used for the fabrication of the transistors in question, the linear dimensions involved are not always exactly comparable.



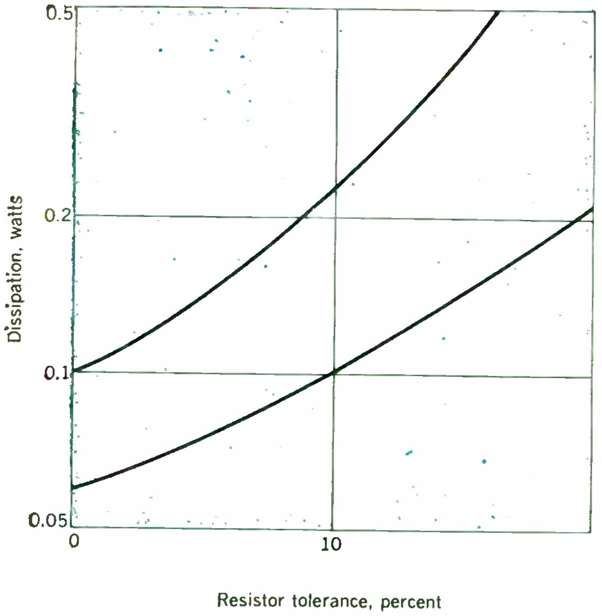


FIGURE 7. An example of the dependence of power dissipation per logical stage on resistor tolerance.[15] The tradeoff between power and number of components is also illustrated; the upper curve refers to a circuit with fewer components than the lower curve (after de Troye).

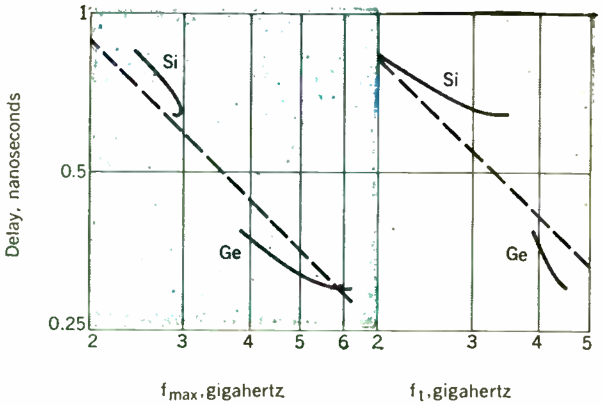


FIGURE 8. Time delay vs. transistor frequency as derived from some measurements of Schlig.[17] Here the frequency and the delay were varied by varying the temperature of a particular circuit. Transistors made in two materials were studied. It is seen that, roughly speaking, the product $t_d f$ is a constant.

fixed by the need for nonlinearity, electric fields increase with decreasing size, and electrical breakdown becomes an increasingly serious problem.

It is known that electromigration phenomena take place at very-high current densities. High-density current causes atoms to move and brings about eventual structural degradation.

Another class of problems that will limit the trend to smaller structures concerns thermal effects. If structures are made smaller, the thermal resistance through which the power must flow away from a device increases and the temperature rise increases.

Unfortunately, it is very hard to put any of this on a really quantitative basis. One reason is that the modern transistor is very complex and cannot be described by the simple one-dimensional models with clearly defined emitter, base, and collector regions that everyone is familiar with. In addition, the modern transistor is so small that it is hard to probe its structure with enough detail to know exactly what one is dealing with. The incompleteness of knowledge also extends into the realm of the physical properties of semiconductors. The electronic structure of very heavily doped semiconductors is incompletely understood and has not yet yielded to theoretical analysis. We can handle the complicated three-dimensional structural features of the transistor by computer simulation, but there is no way to avoid the uncertainties arising from inability to define precisely the structure under consideration and from inadequacy of basic physical theory.

A different kind of obstacle to putting performance limits on a quantitative basis stems from the possibility of trading pure physical performance limits against cost or software. For example, the transistors and the circuits produced by specified technological methods will have a certain variability. Performance can be improved, at the expense of increased cost, by selecting elements for uniformity or high performance. An example, which shows how power dissipation can be traded against component tolerance and against component count, is shown in Fig. 7.[14-16] It seems that resistor variability is somewhat like noise, which must be overcome by electric power. Speed might be increased by allowing a greater temperature rise at a p-n junction, but this kind of speed increase might involve an increased rate of component failure, or an increased probability of error. Increased probability of error might well be tolerated by incorporation of redundancy and error detection. It is apparent that the potential tradeoffs here are too numerous and complex to be susceptible to objective analysis, and that they obscure the quantitative significance of any attempt made to derive performance limits from purely physical reasoning.

Extrapolation

In spite of all these reservations, let us rashly try to extrapolate present trends in computing technology and see where the tendency for power and current densities to increase with circuit speed will become a serious problem. In order to do this, it will be necessary to introduce many rough approximations and drastic simplifications of the complex models that describe the modern transistor. Furthermore, certain extrapolations of transistor technology extend into regions in which the physical models of the transistor may be questioned because the extrapolated dimensions are no greater than such atomic parameters as electron wavelengths and electron mean free paths. Extrapolation into such regimes needs caution.

The first questions we have to ask are: How basic is the reduction of size to the increase of speed? Can it be expressed quantitatively? It seems reasonable and it is indeed the case that computer speed is closely related to transistor speed. An interesting example of this is displayed in Fig. 8, which is based on some recent experiments of Schlig in which he varied the

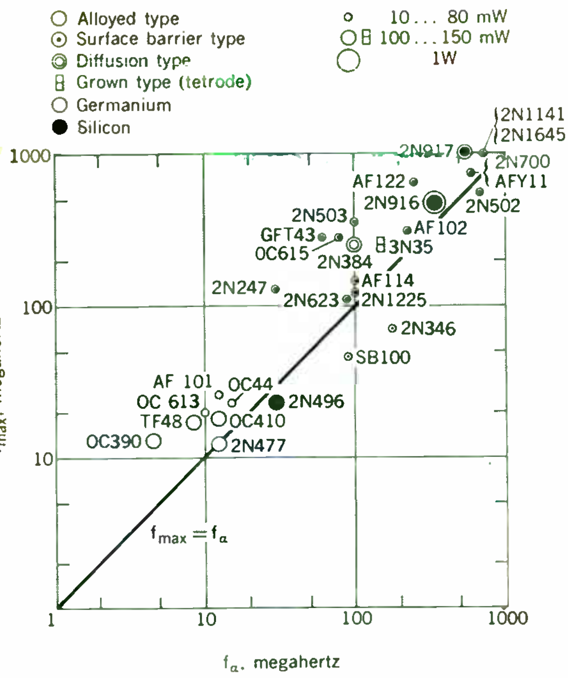


FIGURE 9. Maximum frequency of oscillation f_{max} plotted as a function of the current-amplification cutoff frequency f_α for many kinds of transistors. (Reproduced from Joachim Dosse, "The Transistor," copyright 1964, D. Van Nostrand Company, Inc., Princeton, N.J.)

temperature of logical circuitry and measured both the transistor parameters and the circuit delay.[17] It is seen that, as the temperature is varied and as one goes from one device and material to another, the circuit delay varies approximately inversely as a characteristic frequency of the transistor. The relationship can quantitatively be expressed by saying that the delay times the frequency is a constant that has a value of about two:

$$t_d f = 2 \qquad (3)$$

Although complete characterization of transistor performance is very complicated, has been made the subject of complete books, and requires specification of many parameters, it has been summed up here in the single parameter f. This, of course, is in accord with common usage, which recognizes that some transistors are "faster" than others without entering into detail as to exactly what measure of "fastness" is in question. The impression that measures of speed based on current gain, such as f_a and f_t, are increased in parallel with f_{max}, the maximum frequency of oscillation, as technology progresses may not be gained from elementary textbooks, but the rough equality of these two kinds of measures of speed was clearly recognized and explained long ago by Early.[18] It is responsible for the fact that Eq. (4) may be derived from either part of Fig. 8, and is further illustrated by Fig. 9.[19]

In terms of the device, sizes have to decrease for two reasons. One stems from a property of solids, and is that electrons approach a certain limiting velocity v_L at high electric fields in semiconductors, and thus take a certain minimum amount of time to traverse a fixed distance. The only way to decrease this time is to decrease the distance to be traveled. The second aspect of the size problem is that devices are controlled by charge; in other words, contain capacitors that must be charged. The charging takes place, as we have seen, through impedances of the order of magnitude of Z_0. Thus, charging times can only be decreased by reducing capacitances. Broadly speaking, the capacitance of a structure is reduced by decreasing its area. The extent to which capacitance can be decreased by increasing thickness is limited by the fixed maximum velocity just mentioned. Dimensionally, capacitance is a length and must in the long run change with the length scale.

A quantitive relationship between transistor speed and a linear dimension of a transistor can be extracted from Fig. 6, which shows that the product of the width d and f, a characteristic frequency of the transistor, is a constant. (A similar relationship recently has been demonstrated directly.[20]) The constant s_0 has the dimensions of a velocity and the value (Fig. 6)

$$df = s_0 = 2 \times 10^6 \text{ cm/s} \qquad (4)$$

The form of the relationship between delay time or characteristic frequency and linear dimension can be understood by the following considerations. Charge must be used to change the patterns of electric field and potential in a semiconductor; this is the charge stored in the capacitances of the device. Its magnitude is something like

$$Q = \int \rho A \, dx' = (KA/4\pi x)V$$

Here, ρ is charge density, A is the area of the device, K is the dielectric constant, x is the thickness of an active region of the device, and V is the voltage level. The delay time must be greater than this charge divided by the current, or some kind of RC time:

$$t_D > Z_0(KA/4\pi x) \qquad (5)$$

The delay is also limited by the limiting velocity v_L.

$$t_D = (2/f_t) > 2 \cdot 2\pi x/v_L \qquad (6)$$

Combining (5) and (6) and substituting $Z_0 = 4\pi/c$, the impedance of free space in Gaussian units (c being the velocity of light), yields

$$t_D^2 > 4\pi KA/cv_L \qquad (7)$$

The delay time does indeed come out to be proportional to the square root of the area or to a linear transverse dimension of the structure as suggested by Fig. 6. The characteristic velocity $s_0 = A^{1/2}/t_D$, which relates the delay time and the linear dimension, comes out to be around 10^7 cm/s when v_L is set equal to 5×10^6 cm/s, its value in Ge and Si, according to Eq. (7), within an order of magnitude of the value derived from Fig. 6.

It is also apparent that the gross size of circuitry must be reduced if faster speeds are to be obtained. A significant part of the circuit delay we have been talking about comes from the time for a signal to propagate from one part of the logical circuitry to another. Light can travel 30 cm in one nanosecond in free space and only 7½ or 10 or 13 cm in the structures that contain the high-dielectric-constant materials used in modern logic. Thus the whole scale of computer circuitry must be reduced as speed is increased. The decrease of gross size as speed is increased—required by the finite propagation time—means that the power density and density of heat dissipation increase, independent of the basic device used.

This relation between speed and physical size can be combined with further approximations to give quantitative expression to the thermal problem involved. The power dissipated in the various circuit components causes their temperature to rise. The temperature rise is conceptually caused by the flow of heat through some thermal resistance to some surface at which it can be transferred to a fluid, which eventually carries it out of the computer. The power density at which heat is produced in a circuit element, such as the base of a transistor, is 1000 or more watts per cm^2. The maximum density at which heat can be transferred to fluids is only 100 W/cm^2 or less. This mismatch is accommodated by flow of the heat away from the place where it is produced by conduction through some solid to a surface of much larger area.[21] The desired reduction of the thermal current density can only be achieved if the heat spreads out in three dimensions from its source. The thermal resistance encountered by heat flowing away from a small planar structure into a three-dimensional body is known as "spreading resistance." It may be written in the form

$$R_{th} = \frac{C_2}{2\lambda d} \tag{8}$$

Here, λ is the thermal conductivity of the material and C_2 is a constant that depends on the shape of the planar structure.[21] Of course, there will be other thermal resistances in series with that given by Eq. (8). The importance of the thermal resistance of (8) is that it is inversely proportional to d, the linear dimension. Although we justified (8) by considering spreading resistance, it may also be regarded as a result of dimensional analysis. The temperature rise of the planar structure, the thermal resistance times the power P_m, must not exceed a certain amount, which is conveniently expressed as a fraction C_3 of the temperature itself. Combining (3), (4), and (8) shows that the thermal problem will prevent the time delay from being reduced below

$$t_D > \frac{C_2}{C_3}\,\frac{P_m}{\lambda s_0 T} \tag{9}$$

The essence of the argument here is that, as one progresses to higher and higher speeds by making d smaller, one eventually reaches a point at which the thermal resistance is too large to handle the power.

It is hard to make any but a rough guess at the dimensionless constants involved in (9). Assuming $C_2 = 0.025$, $C_3 = 0.25$, $T = 300°\text{K}$, $P_m = 0.1$ watt, $\lambda = 1$ W/cm·°K, and $s_0 = 2 \times 10^6$ cm/s, Eq. (9) yields $t_D > 2 \times 10^{-11}$ second.

There is another kind of thermal problem that also sets a limit on the gross size of computer circuitry. It arises because the heat generated in computers is eventually transferred to some fluid that carries it away from the machine. There is a maximum density at which power can be transferred to fluids. As a good approximation to this density, we may use the maximum density at which power can be transferred to a fluid by nucleate boiling, about 100 W/cm^2 for fluids around 300°K. If we consider that P_m, the power in the fastest computers, is about 0.1 watt per circuit, then it is apparent that circuits cannot be packed mor closely than 1000 per cm^2. Obviously, this is a gross limit. Heat can be conducted away from small regions of much denser packing by conduction through solids to some place where it can be transferred to a fluid at lower power density. However, this kind of thing can only be carried out to a limited extent in very large machines; in a large, complex system with a three-dimensional structure, the heat must be eventually carried out by the motion of some fluid.

If circuits are packed to a density of 1000/cm^2, the average distance from one circuit to the next is 0.03 cm. Electromagnetic waves traveling with a velocity of 10^{10} cm/s take 3×10^{-12} second to go from a circuit to an adjoining one. Modern computer organization has not solved the problem of implementing logic with each circuit connected only to its near neighbors; many long paths are employed. Experience shows that propagation distances of at least ten times the parameter of the circuit lattice are usually required in known methods for wiring logical circuitry. Thus, it appears that there is a limit on interconnection delays of about

$$t_D^2 > m^2 P_m / Q c_1^2 \tag{10}$$

Here, m is the number of circuit lattice constants in the propagation distance, Q is the maximum rate of heat transfer to a fluid, and c_1 is the velocity of electromagnetic waves on the transmission lines of the circuitry. Using $m = 10$, $P_m = 0.1$ W, $Q = 100$ W/cm^2, and $c_1 = 10^{10}$ cm/s gives $t_D > 3 \times 10^{-11}$ second. Of course, this might be reduced by new concepts in the organization of logical circuitry that reduced m.

Another limit on how small semiconductor elements may be made stems from considerations that have been applied by Johnson to power devices.[12] Johnson's basic relation is

$$V f_t = E_B v_L / 2\pi$$

Here, E_B is the breakdown electric field and v_L is the limiting velocity of carriers in high electric fields. Because of Eq. (3), the delay time t_D is limited. If $E_B = 2 \times 10^5$ V/cm, $v_L = 5 \times 10^6$ cm/s, and $V = 40\,kT/q$,

$$t_D > 2 \times 10^{-11} \text{ second} \tag{11}$$

Note that this limitation has come out to be about the same as the thermal one.

Directions for progress

Since we have stated that the various limits derived are not fundamental or ultimate, but technological, the question naturally arises: What avenues of technological progress will allow the limits to be circumvented? Some of these avenues have been mentioned in passing. Although all of the examples and specific limits described were taken from transistorized logic, man of the considerations described also apply to logical circuitry based on any kind of electromagnetic device. Correspondingly, the technical avenues that might be explored will vary in the breadth of their applicability.

All of the limits specific to transistorized logic alluded to involve material parameters. Obviously then, performance can be improved by the selection of better materials. One way in which a material can be made better for many transistor purposes is by heavier doping, but the extent to which this direction can be exploited is limited by finite solubility of impurities, by the appearance of tunneling effects at high doping levels, and by the decreasing mobility of charge carriers at heavy dopings.

Improvements can also be obtained by finding a new semiconductor material with improved properties. Improved properties in this context are often interpreted simply to mean materials with higher carrier mobilities. It will be observed, however, that the performance limits described in Eqs. (9) and (11) do not involve mobility. They undoubtedly underestimate the importance of mobility, because it is difficult to imagine a device in which the carriers move with the limiting velocity v_L at all times; it is likely that electron motion will be mobility-controlled part of the time. Nevertheless, it is possible to suggest that better figures of merit may also involve such parameters as limiting velocity, breakdown voltage, and thermal conductivity. The introduction of new semiconductors into transistorized logic is, however, an approach that promises only limited progress because the list of possible materials is quite small and the investment required merely to evaluate the physical–chemical feasibility of transistorized technology in a new material is enormous. That is, in addition to possessing a set of favorable physical properties, a new material must also satisfy certain metallurgical and chemical conditions to be useful.

Another line of attack on the limits may be based on reducing the basic power level described by Eq. (1). One way to do this involves trying to increase the impedance level substantially. Raising impedance levels increases capacitance charging times and slows the circuit. Raising impedance at the expense of speed in a given technology becomes possible because the impedance matching requirements become less stringent as speed is reduced. It will be noted in Fig. 4 that there is indeed an inverse tradeoff between power and circuit speed, which is, in fact, largely a reflection of varying impedance levels.[14] The tradeoff can be expressed by the formula Pt_D = constant,[16, 22] where the constant is dependent on the technology. In fact, the constant is about inversely proportional to the capability of the technology as measured by typical f_t's or $f_{\max}$'s, so that one can write approximately

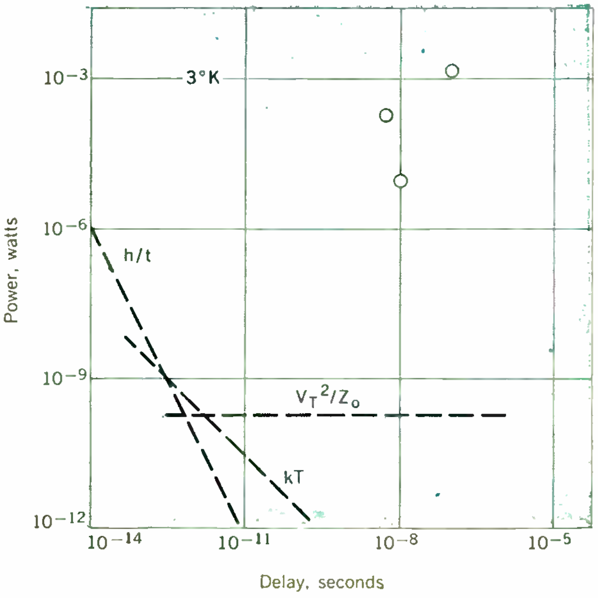
$$Pt_D f = P_0 \qquad (12)$$

The thermal problem described by (9) is not evaded by using the tradeoff possibility expressed by (12) to reduce power as the advance of technology increases f, because (9) involves $P/d \sim Pf$ in the same combination as (12). The tradeoff expressed by (12) can be used to alleviate the thermal problem arising from interconnection delays, however. Combining (10) and (12) yields a weak dependence of t_D on f, $t_D^3 > m^2P_0/Qc_1^2f$; in other words, a slow advance of t_D beyond the limit of Eq. (10) as technology advances (f increases) is not excluded.

It seems obvious, though, that the most straightforward way to decrease the power in Eq. (13) is to lower the temperature. The power level decreases as the square of the temperature, since the voltage level and the current level each decrease in proportion to the temperature. Figure 10 shows how the limits represented in Fig. 5 change when they are evaluated at 3°K instead of 300°K. Since the $(kT)^2/q^2Z_0$ power goes down as the square of the temperature it is relatively less important in comparison with the power based on Landauer's ideas or thermal noise, and based on the uncertainty principle at low temperature. The only limitation that is not changed by lowering the temperature is the quantum one. It seems that the only real ultimate limit is set by the uncertainty principle. Points representing the performance of a few cryogenic logic elements described in the literature are also plotted in Fig. 11. The power involved in these circuits is generally lower than that used in 300°K circuits of comparable speed, but it is even further away from the limits than the 300°K circuits.

The lower power causes the limits on the performance of transistorized logic described by Eqs. (9), (10), and (11) to improve as the temperature is lowered. In fact, the performance improves not only because the power and the current levels go down, but also because the important material properties improve, at least through certain temperature ranges. The increases of mobility and thermal conductivity with decreasing temperature are well known, and fragmentary information suggests that the scattering-limited velocity, which looms large in our limits, also increases. The heat transfer limit of Eq. (10) decreases with decreasing temperature because, as illustrated in Table I and Fig. 11, although the ability of fluids to accept heat by nucleate boiling at a surface decreases linearly with temperature, the power level decreases with the

FIGURE 10. A plot similar to Fig. 5, showing how the theoretical constraints on power and energy change when the temperature is lowered to 3°K. Points that represent the performance of cryogenic circuitry are also included.



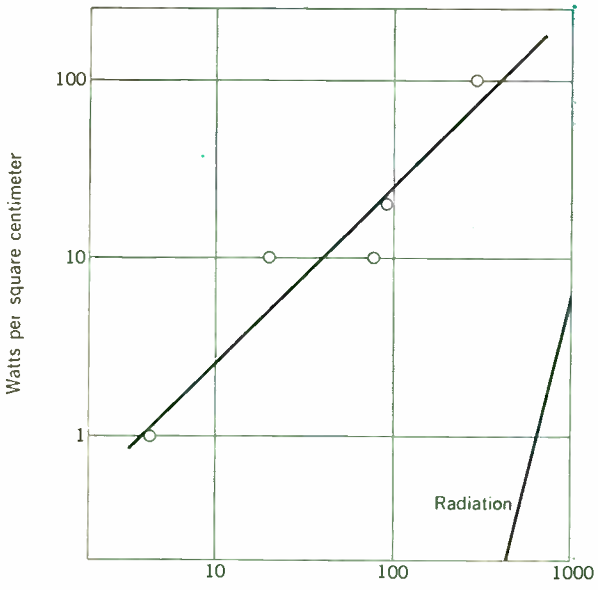


FIGURE 11. Maximum heat transfer rate in nucleate boiling for several cryogenic fluids as a function of their boiling temperatures. Data are from Table I; the equation of the line that represents the data is Q = 0.25T.

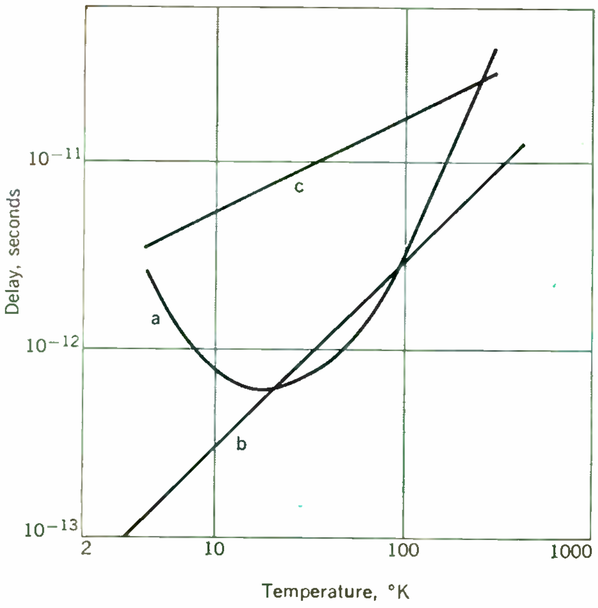


FIGURE 12. The dependence of the limits on delay described by Eq. (9) (curve a, based on the thermal conductivity given by curve 2 of Fig. 13), Eq. (10) (curve c), and Eq. (11) (curve b) as functions of temperature.

square of the temperature. The various limits given by (9), (19), and (11) are plotted against temperature in Fig. 12. A steady improvement down to temperatures of about 20°K can be predicted. Beyond this, however, a serious new problem starts to limit the performance of semiconductor elements. This is the decreasing thermal conductivity of semiconductors below about 20°K, illustrated in Fig. 13.[23, 24] The basic effect is that, in insulating crystals, heat is carried by phonons, and at very low temperatures the number of phonons available to transport heat rapidly decreases. It appears that the usefulness of semiconductors in the helium range will be severely limited by the difficulty of transferring heat through the semiconductor.

Semiconductors may not be useful at very low temperatures for still another reason. Our prediction of improved performance at low temperatures basically results from the fact that the voltage level can be reduced in proportion to the temperature. The reduction of the voltage is possible because the voltage scale on which nonlinearity occurs is proportional to the temperature. It turns out, however, that the nonlinearity of semiconductor junctions does not continue to increase down to the lowest temperatures. Several effects contribute to this reduction of nonlinearity. The nonlinearity on the scale of qv/kT, usually found in semiconductor junctions at 300°K, is a result of the necessity for electrons to be thermally activated over a barrier in order to produce current. As the temperature is lowered, this thermally activated current that crosses the barrier becomes smaller and smaller.

Other contributions to the current, which may not be temperature dependent, become relatively more important. One such other kind of current is tunneling current. (Tunneling current arises from the penetration of a small part of the wave function of an electron, which is predominantly on one side of the junction, into the other side of the junction.[25, 26]) Another is current that crosses the junction through states in the band, the states in the so-called band tail. The nonlinearity is also reduced by compositional fluctuations, which lead to inhomogeneities in the energy gap in the junction plane. All of these effects tend to prevent the nonlinearity of semiconductor junctions from increasing with decreasing temperature below a certain temperature.

I. Scaling of area needed to transfer power dissipated in logical circuitry into a fluid

Fluid	T, °K	Q, * watts/cm²	P, † watts	P/Q, cm²
Typical	300	100	1.8×10^{-6}	2×10^{-8}
O_2	90	20	1.6×10^{-7}	10^{-8}
N_2	77	10	1.2×10^{-7}	10^{-8}
H_2	20	10	8×10^{-9}	10^{-9}
He	4.2	1	3.5×10^{-10}	3.5×10^{-10}

* The maximum rate at which heat can be carried away by nucleate boiling of common cryogenic fluids.[22]
† Calculated from Eq. (1).

Conclusion

Transistorized computer logic has made steady progress toward higher speeds by reducing the dimensions of circuits and devices. However, even though circuit and device speeds have increased by three orders of magnitude since the introduction of the transistor into computer logic, the voltage, current, and power levels have remained about the same. Power *densities* and current *densities* have been increasing rapidly as logical circuitry becomes faster and faster. The dissipation of power at increasingly high densities seems to be leading to difficult thermal problems that

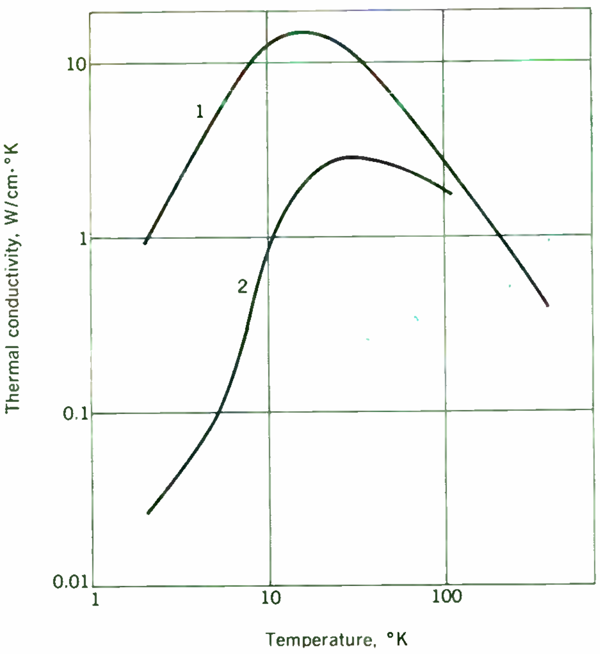


FIGURE 13. The thermal conductivity of n-type germanium as a function of temperature, illustrating the maximum thermal conductivity in the 10°K to 50°K range and the sensitivity of the thermal conductivity of semiconductors to impurities at low temperatures. Sample no. 1 is a pure sample with 10^{13} cm^{-3} impurities, and sample no. 2 is a heavily doped sample with about 10^{19} cm^{-3} Ga acceptor concentration.[23]

eventually will limit the progress of logical circuitry toward higher speeds. An estimate of the limit on speed, based on extrapolation of present technology, indicates that the thermal limits derived in various ways are about the same and that they lie about an order of magnitude beyond the speed of the fastest contemporary circuits. Progress beyond this point can only be made by radical deviations from the current lines of development. The most straightforward new method seems to be lowering the temperature at which the circuitry is operated.

Physical reasoning suggests that the power level of logical circuitry can be decreased in proportion to the square of the temperature as the temperature is lowered. Thus, at lower temperatures, the thermal problems appear at smaller dimensions and higher speeds. Low-temperature operation of semiconductor circuitry seems to be the logical first step in taking advantage of this observation. It appears that semiconductors will cease to be suitable at temperatures in the range 10°K to 50°K but, at lower temperatures, other useful electrical phenomena appear; and these are based on superconductivity.

REFERENCES

1. Kolsky, H., "Some computer aspects of meteorology," *IBM J. Res. Develop.*, vol. 11, pp. 584–600, Nov. 1967.
2. *New York Times*, Sept. 29, 1968.
3. Feynman, R. P., "There's plenty of room at the bottom," in *Miniaturization*, H. D. Gilbert, ed. New York: Reinhold, 1961, pp. 282–296.
4. Landauer, R. W., "Irreversibility and heat generation in the computing process," *IBM J. Res. Develope.* vol. 5, pp. 183–191, 1961.
5. Keyes, R. W., "On power dissipation in semiconducting computer elements," *Proc. IRE*, vol. 50, p. 2485, Dec. 1962.
6. Bernard, W., and Callen H. B., "Irreversible thermodynamics of non-linear processes and noise in driven systems," *Rev. Mod. Phys.*, vol. 31, pp. 1017–1044, 1959.
7. Gunn, J. B., "Thermodynamics of non-linearity and noise in diodes," *J. Appl. Phys.*, vol. 39, pp. 5357–5361, November 1968.
8. Matisoo, J., "The tunneling cryotron—a superconductive logic element based on electron tunneling," *Proc. IEEE*, vol. 55, pp. 172–180, Feb. 1967.
9. Karlovsky, J., "The curvature coefficient of germanium tunnel and backward diodes," *Solid State Electron.*, vol. 10, pp. 1109–1111, Nov. 1967.
10. Sze, S. M., and Ryder, R. M., "The nonlinearity of the reverse current-voltage characteristics of a p-n junction near avalanche breakdown," *Bell System Tech. J.*, vol. 46, pp. 1135–1139, July–Aug. 1967.
11. Cutler, M., and Bath, H. M., "Surface leakage current in rectifiers," *J. Appl. Phys.*, vol. 25, pp. 1440–1441, Dec. 1954.
12. Johnson, E. O., "Physical limitations in frequency and power parameters of transistors," *RCA Rev.*, vol. 26, pp. 163–177, June 1965.
13. Kirk, C. T., Jr., "Theory of transistor cutoff frequency falloff at high current densities," *IRE Trans. Electron Devices*, vol. ED-9, pp. 164–174, Mar. 1962.
14. Wallmark, J. T., "Basic considerations in microelectronics," in *Microelectronics*, E. Keonjian, ed. New York: McGraw-Hill, 1963, pp. 10–96.
15. deTroye, N. C., "Relations between dissipation, fan-out, and resistor tolerances in microminiature digital circuits," in *Micro-Miniaturization*, J. Berghammer, ed. Munich: R. Oldenbourg-Verlag, 1966, pp. 417–421.
16. Suran, J. J., "Circuit considerations relating to microelectronics," *Proc. IRE*, vol. 49, pp. 420–429, 1961.
17. Schlig, E., "Low temperature operation of Ge picosecond logic circuits," *1968 Internat'l Solid State Circuits, Conf. Digest*, pp. 158–159.
18. Early, J. M., "Triode transistors," *Proc. IRE*, vol. 46, pp. 1924–1927, Dec. 1958.
19. Dosse J., *The Transistor.* Princeton: Van Nostrand, 1964.
20. Dill, F. H., Farber, A. S., and Yu, H. N., "Picosecond integrated circuits in germanium and silicon," *IEEE J. Solid-State Circuits*, vol. SC-3, pp. 160–163, June 1968.
21. Keyes, R. W., "Thermal problems of the injection laser," *IBM J. Res. Develop.*, vol. 9, pp. 303–314, 1965.
22. Narud, J. A., and Meyer, C. S., "Characterization of integrated logic circuits," *Proc. IEEE*, vol. 52, pp. 1551–1564, Dec. 1964.
23. Carruthers, J. A., Geballe, L. H., Rosenberg, H. M., and Ziman, J. M., "The thermal conductivity of germanium and silicon between 2 and 300°K," *Proc. Roy. Soc.*, vol. 238A, pp. 502–514, Jan. 29, 1957.
24. Goff, J. F., and Pearlman, N., "Thermal transport properties of n-type Ge, at low temperatures," *Phys. Rev.*, vol. 140, pp. A2151–A2169, Dec. 13, 1965.
25. Zener, C., "A theory of the electrical breakdown of solid dielectrics," *Proc. Roy. Soc.*, vol. 145A, pp. 523–529, 1934.
26. Esaki, L., "New phenomenon in narrow Ge p-n junctions," *Phys. Rev.*, vol. 109, pp. 603–604, 1958.

Robert W. Keyes (SM) studied physics at the University of Chicago, receiving the Ph.D. degree in 1953. He was associated with the Chemistry Division of Argonne National Laboratory from 1946 to 1950 and served in the U.S. Navy for two years. From 1953 to 1960, he was a staff member of the Westinghouse Research Laboratory in Pittsburgh, Pa., where he was involved in research on the properties of semiconductors under high hydrostatic pressure and many other aspects of solid-state physics. Dr. Keyes joined the IBM Research Laboratory in 1960, and has engaged in research and development in fields related to modern electronic technology since then. A Fellow of the American Physical Society, he is also the author of many publications in the engineering and scientific literature.

Art and technology

II. A call for collaboration

Engineers have begun to accumulate experience in various kinds of interdisciplinary ventures as, for instance, with physicians and psychologists. Now they are being asked to collaborate with artists in what could prove to be the most intriguing interdisciplinary adventure of all

Nilo Lindgren *Staff Writer*

As decribed in Part I, the collaboration of art and engineering in the creation of new art works has become a serious preoccupation of many modern artists, who are asking for the help of engineers. A number of organizations now exist whose aim is to facilitate the contact between artists and engineers and to help support them in their collaborative projects. You need not be a Renaissance Man, it is said, to apply for a match with an artist. It won't be all fun and games, although part of it will be, and you might even end up doing something so useless from an engineering point of view, and so right from another point of view, that you could begin wondering why engineering is practiced the way it is—i.e., you might get turned on. This article tells you a bit about what it was like for some others who have already tried, and calls on you to join the action.

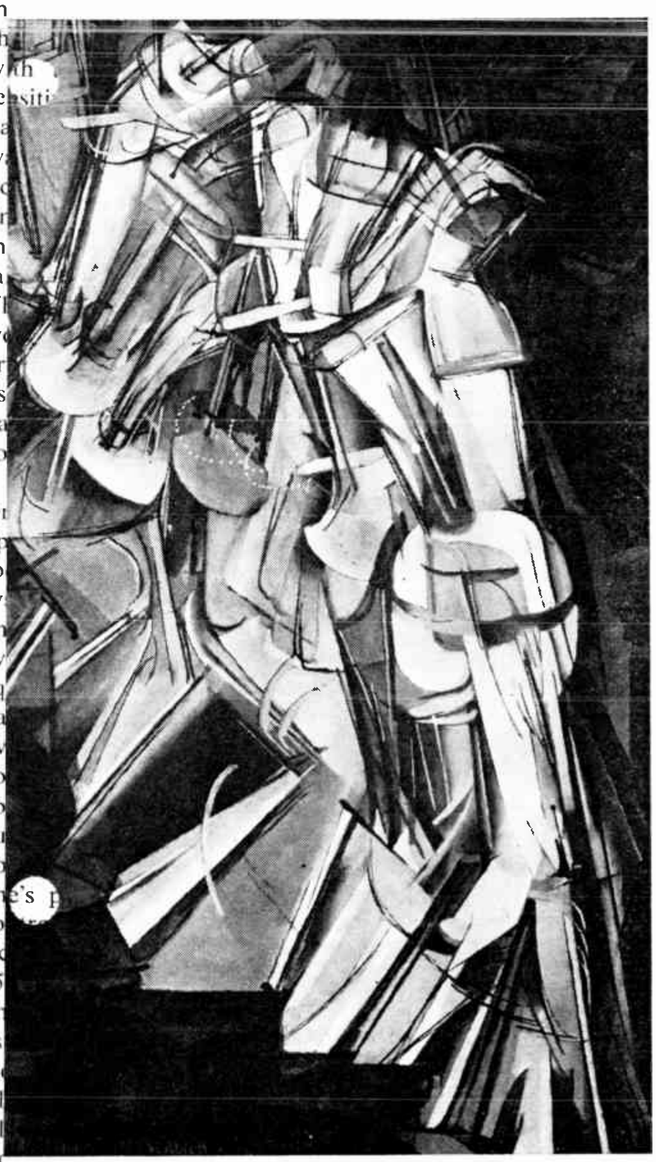
Anybody who has read even a smattering of the art history of this century knows about the famous Armory Show of 1913 in New York. That "International Exhibition of Modern Art," held in the Armory of the 69th Infantry, punctured the United States' esthetic innocence with a revelation of the radical iconographic movements that had emerged in Europe. Containing such works as Marcel Duchamp's "Nude Descending a Staircase, No. 2," the show inspired much controversy. Though, in general, the contemporary press was critical of the show, there were some critics who recognized its importance, and what it heralded for the art of succeeding years. After the 1913 Armory Show, nothing was the same in U.S. art.

The second Armory show, "9 Evenings: Theater and Engineering," held October 13 through October 23, 1966, may have performed a similarly significant role. The performances, created jointly by artists and engineers in the first such collaboration on a major scale, were as much as anything the creation of environments. They were not, in any case, evenings of theater in any traditional sense. Moreover, their creation served both as a manifesto and as an experimental probe.[1]

Actually the experiment in art and technology that grew into the nine evenings of theater and engineering at the Armory in New York was originally scheduled as part of a festival of art and technology in Stockholm; but the project to go to Sweden grounded on the very issue that, according to some interpreters, imparted to the nine evenings their unique significance and their claim to success in the face of acerbic critical response.

At issue was the relationship between the artists and engineers. Stockholm wanted to supply engineering support in Sweden; the artists were, in effect, to give in their orders and the engineers would carry them out. But the New York group opted for what constitutes not only a different mode of doing business but a fundamental difference in esthetic outlook as well. The artists wanted the engineers to be in on the groundwork and planning, to engage in an ongoing dialogue with the artists, from the very beginning. To engineer Billy Klüver and artist Robert Rauschenberg, the two major driving forces of the 9 Evenings, the exploration of the *process* of collaboration constituted the fundamental experiment. It was even suggested at one point that the audience be allowed to come to the Armory during the days to see the preparatory work in progress.

Thus, the upshot of the dispute with Stockholm on the relationship between artists and engineers—which art critic Brian O'Doherty noted at the time as a "prophetic friction!"—was to lead to a pioneering venture. The New York artists wanted, as O'Doherty wrote, "to work with the technicians—follow their



MARCEL DUCHAMP'S "Nude Descending a Staircase, No. 2," 1912 (Arensberg Collection, Philadelphia Museum of Art).

thinking, involve them in the artists' ways of thinking. This makes great demands on patience, mutual understanding, and above all, time. Ideally, it would acquaint the artists with their new materials, since art often arises out of the nature of materials themselves. This democratic idea is especially typical of Rauschenberg's equal respect for all materials, which in this case included people."[2]

Rauschenberg himself speaks of the waste of possibilities in using technology only to carry out some preplanned work, for the technology is then like a *dead* medium rather than an encounter. "I do think there are," he says, "definitely two attitudes toward technology, and the one of using technology as a servant I find very dull and outdated. It seems to me closed-minded." However, in an "open esthetic investigation, where it almost doesn't matter whether one makes art out of it or not, in which an art object may not even be produced," the purpose is not to move in a single direction. Instead of taking the materials and just executing something preconceived, if "from the beginning, if you have some curiosity about what would happen if you could do *this*, and then you share that curiosity with someone else and make him curious, then that's the beginning of the piece . . and it grows and grows . . . the whole world then is the medium, if you can relate in your esthetic to something other than the finished product."[3]

In this light, it is possible to think of the 9 Evenings as an experiment in which not only artists and engineers were engaged, but in which the audience was engaged as well, an experiment in which the consequences and the significance would not become visible until later—perhaps only in successive works. As O'Doherty wrote later, the evenings "received, on the whole, an appalling press—based mainly on the justifiable irritation of interminable delays, technical failures of the most basic sort, and long, dead spaces between, and sometimes in the middle of, pieces. Yet, as such irritations faded away, one is left with startlingly persistent residual images, and strong hints of an alternative theater that has been lagging in its post-Happenings penumbra between art and theater. . . ."[2]

And, with time, the afterlife of the 9 Evenings grows stronger. One critic who saw the evenings as a "total disaster" now concedes that the experiment stands as a kind of "inspired announcement."

The '9 Evenings'—the collaborators

The artists involved in the evenings included composer John Cage, painter and playwright Öyvind Fahlström, painter and choreographer Alex Hay, choreographer Deborah Hay, choreographer Lucinda Childs, artist Robert Whitman who did some of the earliest "Happenings,"[4] choreographer Steve Paxton, composer David Tudor, choreographer Yvonne Rainer, and many others who contributed in performing, handling, and even helping build equipment. The engineers, recruited from Bell Labs, included: Per Biorn, who works in semiconductor research and who designed and built a ground effect machine for Child's "Vehicle"; Cecil Coker, who has worked on synthetic speech computers; Peter Hirsch, who works in underwater sound and who built an 80-kHz Doppler sonar for Lucinda Childs; Harold Hodges, who works in laser research and who designed and built an antimissile missile and floating snowflakes for Fahlström's "Kisses Sweeter Than Wine"; Jim McGee, who works on holograms but built programming drums used dur-

Tried
conversation (engineers and artists).
Found it didn't work. At the last
minute, our profound differences
(different attitudes toward time?)
threatened performance. What changed
matters, made conversation possible,
produced cooperation, reinstated
one's desire for continuity etc.,
were things, dumb inanimate things (once
in our hands they generated thought,
speech, action).

The '9 Evenings' as performances

As performances the 9 Evenings differed from one another as much as the artists who created them. In Rauschenberg's "Open Score," for instance, imagine the action. It takes place in the enormous half-barrel-shaped space. The Armory floor becomes a tennis court, and as the two players hit the ball back and forth, its sound against the racket is transmitted from radios inside the racket handles, and amplified through speakers. As the play proceeds, the lights illuminating the court are extinguished one by one each time the ball is struck, until the stage is dark, at which time the game ends. At that point, hundreds of people move onto the Armory floor where they are illuminated with infrared lights; the crowd scene in the darkness is picked up with infrared sensitive television cameras, then projected on three large screens, so that the audience feels the presence of a large crowd without being able to see it except on the reproduced screen images. When the lights finally go on, the huge crowd of "actors" bows to the audience seated on the bleachers.

In Whitman's piece, "Two Holes of Water," the Armory space, in relative darkness, gradually fills up with automobiles driven in slowly, sheathed in plastic, announced by loud sounds as of warfare and political speeches. The cars maneuver into positions around movie screens that appear one by one around the Armory space, until one is viewing simultaneously about ten screens—showing animals running, penguins frolicking, planes taking off and landing backwards, cells dividing, a man and woman dressing and undressing, caressing, silently mouthed television advertisements, and landscapes going by—accompanied by the sounds of bombs exploding, aircraft, political harangues, typewriters clacking, and so on. One sees live actors on the stage being televised and, simultaneously, their televised images being viewed by the anonymous spectators in the cars, who are also being observed by the audience. This induces in the observer a gestalt of subjective responses, normally reserved for these separate images, which, in their radical juxtaposition in one's feelings, produce an insight into one's modes of "pattern recognition," of one's modes of taking elements of this technological world.

Whitman's was a beautifully choreographed piece, simultaneously both amusing and deeply moving, and riddled at last with the sad sense of how the available space of our world is filling up with the feverish productivity of technology.

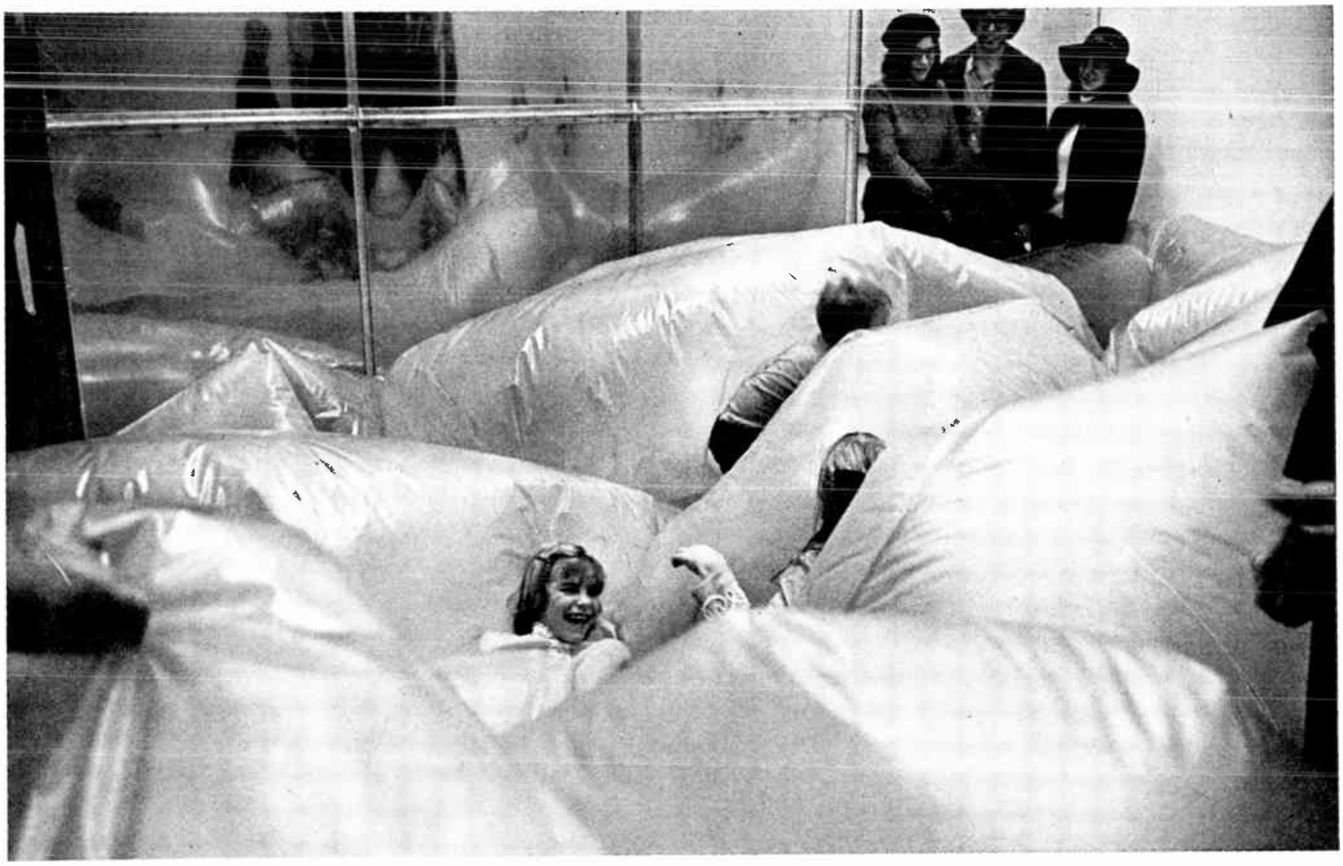
More than this cannot be described here; the photographs on these pages suggest a bit more. For better or worse, the 9 Evenings now belong to the history of art and theater. They belong as well to the history of technology, not because they were technologically momentous, but because they were a foretaste—a first step toward a perhaps vigorous involvement and engagement of recently separate endeavors.

Not everything worked during the 9 Evenings, the "downtime" was considerable, the technology was, as Klüver admitted, peanuts; and the Evenings were still, as John Cage suggests, like the early movies in which the traditional stage, the new camera, the literary content, and the acting were identifiable components not yet integrated. But for those who participated, the Evenings were an achievement full of anguish and joy. For them, the experiment was a success—they had discovered, and came away convinced, that they could work together, and that a lot of energy has been wasted in the past in sustaining a supposed conflict between cultural specialists.

Collaborative organizations

There is now much evidence that the movements of certain artists toward engineers and technology, and toward the engineering industries, seems to be occurring spontaneously in many places. Almost every day, one hears of shows in art and technology and of groups and alliances being formed to advance the "new combine."

A conspicuous new feature of the Cambridge intellectual landscape, for instance, is a Center for Advanced Visual Studies at the Massachusetts Institute of Technology under the direction of the distinguished artist and professor Gyorgy Kepes. The new Center, housed almost symbolically in what used to be M.I.T.'s student cooperative store, aims at the collaboration of artists with scientists and engineers through the appointment of artists as Fellows of the Center for periods of two or three years. The first group of Fellows includes William Garnett, an aerial photographer; Vassilakis Takis, a Greek sculptor of kinetic devices controlled by magnets; Harold Tovish, the producer of recent mirage-like optical sculptures; Otto Piene, a German artist known for his light sculptures, light ballet, and multimedia performances—in Düsseldorf he belonged to *Group Zero* whose members often did art works of a transitory nature and stressing their "phenomenological" character; Jack W. Burnham, a light and kinetic sculptor; and John Whitney, who has made computer movies. The artist–Fellows not only engage in seminars and meetings with M.I.T. scientists, but they work together as well on ways in which to use light in large-scale environmental forms. A monumental sculpture of light in Boston Harbor is one of the projects under consideration.



THIS PIECE, an inflatable plastic environment by Pedro Lujan, whose "message" is clearly understood by the younger set, and the two pieces shown on the next page, were included in a large exhibition of about 120 worker involving collaborations between artists and engineers and using technical materials and processes. The exhibition, "Some More Beginnings," held at The Brooklyn Museum this winter, included all the works resulting from a competition sponsored by Experiments in Art and Technology, New York.

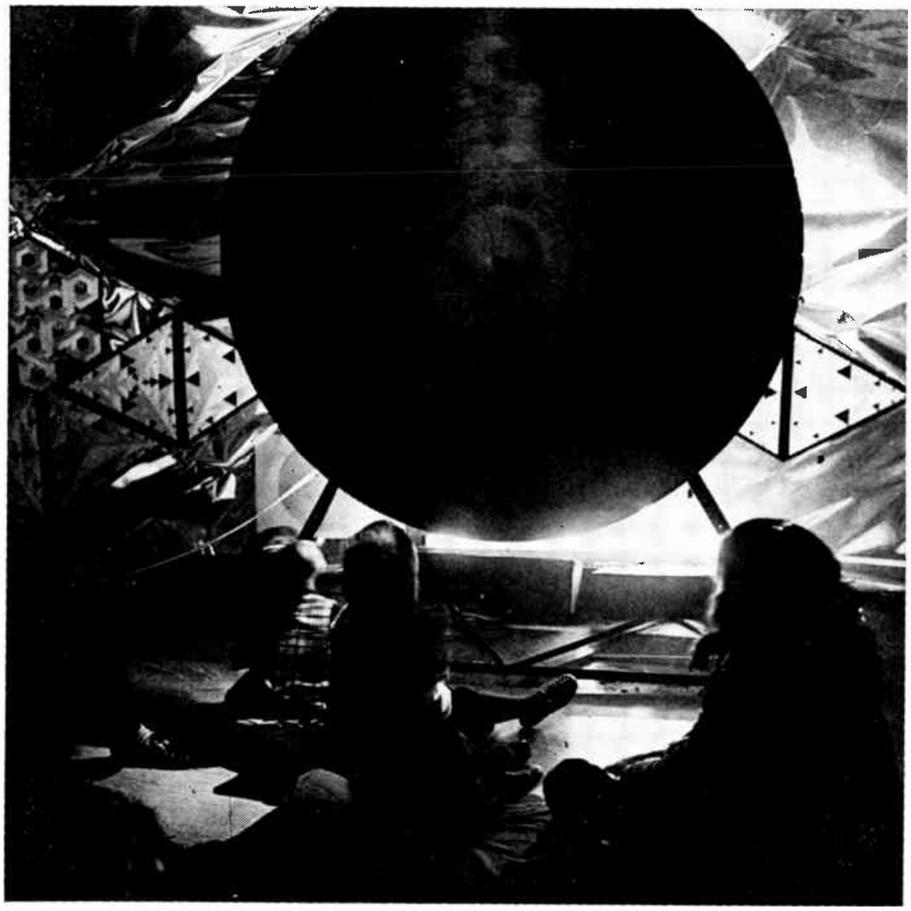
From Kansas City a year ago came news of another venture like the 9 Evenings—the creation of "The Magic Theatre" at the Nelson Gallery of Art, a collaborative venture of the museum, artists, the Kansas City Performing Arts Foundation, and more than 40 local and out-of-state industries, which contributed nearly a half-million dollars of man-hours and materials.[5] Eight enviromental art works, which required a massive technical collaboration, were created by artists Stanley Landsman, Howard Jones, James Seawright, Charles Ross, Robert Whitman, Stephen Antonakos, Boyd Mefferd, and Terry Riley.

Kansas City's "Magic Theatre" title is said to have been selected from Hermann Hesse's novel *Steppenwolf* but what has not been mentioned is that Hesse's "Magic Theatre" also had a subtitle—"Entrance Not for Everybody"—which is apt, because all the ventures mentioned thus far were collaborations of rather exclusive groups.

But an organization devoted to the collaboration of artists and engineers for whom entrance *is* for everybody now does exist—any artist and any engineer can join it free merely by letting the organization know that he is interested in experimental collaboration; and any industry, company, foundation, or whatever, may support it if it wishes by allowing its engineers to work collaboratively with artists, by allowing artists to come into its R&D and production facilities, by donating equipment and components, or by giving money, preferably in thousand-dollar chunks every year, and sustaining such support over many years. The new organization, called E.A.T. (Experiments in Art and Technology), is prominent not only for its joyful democratic principles (conducted in a free-spirited and slightly anarchic atmosphere where each person does what he does best); it is prominent, too, because of its scale of ambition, its breadth of contracts, and its seriousness of purpose in creating a healthy freeway between artists and engineers not only so that they can find one another in the first place but also by supporting the ongoing work-dialogues that are established. Moreover, any engineer or artist who wishes to contact E.A.T. to establish a collaboration need in no way be convinced that he is a "Renaissance Man." The collaboration may even go better and be more exciting if he isn't.

The need for some form of organization to facilitate engineer–artist contacts had been demonstrated in many ways. A growing number of artists everywhere are hungry for engineering help. For instance, even before the formation of E.A.T., a relatively unadvertised meeting, asking for artists and engineers who might be interested in collaborative projects, brought a turnout of over 300 artists, 75 of whom brought

"ULTIMATE PAINTING," 1968 (right), by Clard Svenson and Larry Lard, is a Mylar stretched zonahedron dome environment. Inside the dome, a painted 1.5-meter-diameter disk spins, lit by four strobe lights, alternated with ultrasonic illumination in 20-minute intervals. The piece "Speak That I May See," 1968 (below), by Roberta Phillips and Copthorne MacDonald, is a life-size plaster and electronic sculpture. On its cathode-ray-tube face, one sees time-varying patterns generated by stereophonic music. All of the pieces in the E.A.T.–Brooklyn Museum exhibition were categorized as to whether they were planar images, reliefs, constructions, environments, processes, performances, and as to whether they generated sensory stimuli through light, sound, or heat sources, whether they incorporated moving parts, how their images were generated, whether they were interactive with the spectators or with the environment, and as to whether or not people were essential to the piece as in games and theater.



with them pending projects for which they wished immediate engineering assistance. Rauschenberg says that on his travels to South America and in Europe, the situation is the same. The young artists do not even manifest much interest in pop art; it is about art and technology that they ask.

Thus, it was clear that a practical pathway needed to be opened up so that artists could get to engineers, could get into industry, to get scientific and engineering information, and that real financial support was needed for artists who wished to work in an inherently more expensive medium. The chief question was how to reach the engineers and let them know how much they were wanted, because the artists could not hope to have access to the most sophisticated and advanced technology without gaining access to engineers.

Probably more than any other single factor, the preconceptions that engineers and industrialists held about the character of art and the character of artists had to be unfrozen. For it is undeniable that for most of us who have been nurtured on earlier painters—those of us who have grown up loving the works of Uccello, Bosch, Brueghel, Goya, Hiroshige, Monet, Picasso—and excited by works created on a highly personalized basis, there is an impediment to perceiving that the artist's interest in technology is both profound and pervasive. Moreover, in this respect, the formation of an *organization* presumably to advance the cause of artists may strike one as antithetical.

These and other considerations led the founders of E.A.T. to formulate a philosophy of collaboration that strikes out questions of esthetics as much as possible. In their matchings of artists and engineers, they stress the *process* of collaboration and experimentation rather than the results a particular artist aims for; they aim at involving the artist with the relevant forces shaping the technological world rather than concerning themselves with the set of esthetic convictions that guide him. Consequently, just as all kinds of specialists are welcome under the umbrella of IEEE, so all kind of artists and engineers are welcome in E.A.T. The organization thus hopes for a rich growth through the avoidance of artistic cliques or factions.

In its short career, E.A.T. has already assembled a list of remarkable achievements. It has attracted about 6000 members (including roughly 2000 engineers and 2000 artists) and made about 500 matchings in collaborative projects. It has organized about 35 local groups in the United States. Correspondent groups have been formed in Argentina, Australia, Brazil, England, France, Germany, Holland, Italy, Sweden, and Switzerland. In addition, E.A.T. has opened regional offices in Los Angeles and Tokyo, and a group in Canada is setting up a national group there. Moreover, E.A.T. has sponsored a competition and several awards for an *engineer's* contribution to a collaboration, with the jury drawn from scientists and engineers. Some of the works, which appeared in a major exhibition of all the entries at The Brooklyn Museum, are shown on these pages. One of the most recent projects undertaken is the orchestration of an international collaboration for Expo 70. By 1980, E.A.T.'s officers expect a $10 million operating budget, with several million going to the support of large-scale projects, and several million more going into services for artist and engineer members. Present support for the organization comes from the Rockefeller Brothers Fund, the National Endowment for the Arts, industry, and private individuals.

Experiments in Art and Technology can be contacted at 235 Park Avenue South, New York, N.Y., 10003. There has also been some thought about the possibility of forming an art-and-technology group within IEEE. Expressions of interest or concern sent to IEEE Headquarters might galvanize the issue one way or another.

A model for collaboration?

Despite the foregoing "advertisements," one must recognize that we have been describing pioneering experiments that have been going on in a rarefied atmosphere. We see nothing yet resembling the unity that prevailed in the fine and mechanical arts in centuries past. It is probably a fact that the art-and-technology movement at this time occupies the attention of only a minority of artists and undoubtedly an even smaller minority of engineers. Moreover, in terms of completed pieces in the art-and-engineering genre, there is not yet much to sing about; collaboration exists more in the realm of possibilities than actualities.

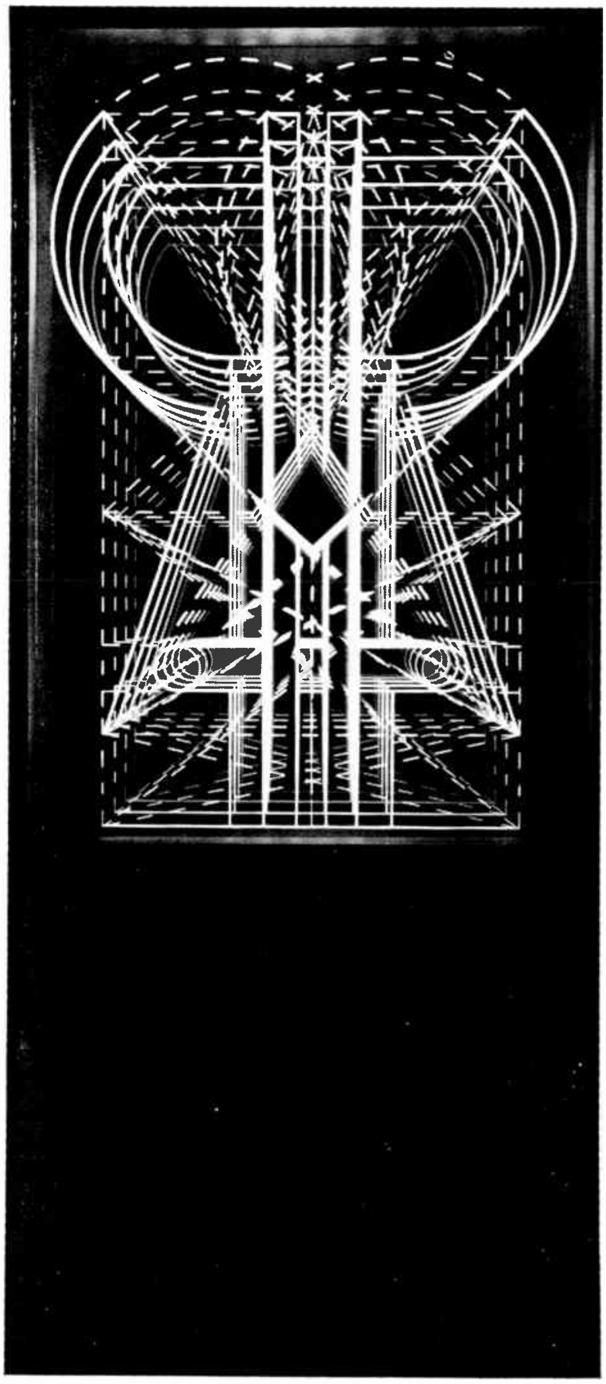
Thus, at this stage, unless one is inclined to condemn the idea of artistic and engineering collaboration outright, it is probably wise to focus one's attention on the process of collaboration. The sympathetic onlooker might say: let the experiment find its own limits and constraints; let it stake out its special territories through trial and error; let it grow as it will without worrying overly much about where specific works belong as far as esthetics or engineering are concerned. Nevertheless, in this situation, one might wish to know if there have been other collaborative situations that bear any resemblances to that of art and technology, and whether or not one can learn anything from an examination of such situations.

Two such parallel or analogous alliances come immediately to mind—the collaborations between engineers and physicians and collaborations between engineers and psychologists, both of which led to the organization of groups within IEEE (the Bio-Medical Engineering Group and the Human Factors in Electronics Group, now called the Man–Machine Systems Group).

Engineering's serious alliance with medicine began not many years ago, generally in the form of a collaboration between individual physicians and engineers. Some of these collaborations worked, some didn't; most were undersupported; and most resulted in various kinds of assistive instrumentation for physicians. (More deep-going, sensitive, and truer collaborations, as in the doctor–engineer developments of heart pacemakers and artificial organs, have really got going in the past few years.) However, it often turned out that the expectations of the physicians and the engineers were disappointed. The really complex systemic problems confronted by doctors, in which many interdependent variables had to be taken into account (if they were known or understood at all) would not yield to the kind of component engineering a single engineer might carry out in his basement in his spare time. Moreover, there were resentments polarizing around the question of who was to hold the governing shares in the partnership. The usual case in those days was that the doctor insisted, undoubtedly rightly (the patient's condition being his sole responsibility), that he was to have final say on equipment development. The medical position was that the engineer would never come to understand the medical problems in their true complexity. Thus, engineers who had been attracted to work that they felt had much social and moral reinforcement to recommend it (and who in some cases were deliberately moving *away* from work on weapons systems at the cost of sharp cuts in income) now found themselves in the sometimes frustrating role of mere servant–technicians. They felt injured by the subservient positions in which physicians held them, and insulted by their failure, in spite of their sacrifices, to be respected as true professional men. In any case, such collaborations, whether they failed or succeeded, sharpened questions about the nature of engineering professionalism. Who was the more professional? Was it the doctor, working alone and personally with individual patients, bearing the final responsibility for the success or failure of his treatment; or the engineer, usually working on a large coordinated team, and usually within the framework of a paternalistic industrial situation or establishment, on supersystems where the responsibility is shared out among many individuals? Or were the two situations really comparable? If the physical system (computer, missile, space vehicle, or what have you) or its components failed to meet schedule, or failed in performance, delivery or use could be postponed with consequences neither so severe nor disquieting as those encountered by a physician or surgeon.

Despite such issues and other subsequent arguments, however, the collaboration of medicine and engineering has steadily gained ground and adherents; financial and physical support, though never large, has increased and seems now about to take off to substantially greater levels; new biomedical engineering departments have been established at many universities; teams of specialists of all kinds from the engineering and life sciences have made real and impressive advances that would have been out of reach of engineers and doctors alone; and, of late, it has been possible to discern that biomedical engineering has begun to take on a character of its own. No longer pure engineering, no longer pure medicine, in the traditional sense, the field is claimed by some of its spokesmen to be no longer a hybrid. Although one still hears overtones of the old arguments and expressions of disillusionment, and complaints about communication barriers, biomedical engineers of the younger generation say that the integration is tacitly complete and that the biomedical engineering field is in the process of building its own scientific basis. The situation now, they often say, is that you pose the biomedical problem you wish to solve, and then you seek for the tools and methods, from engineering or medicine, wherever they may be, in order to solve the problem.

"BARMECIDE," 1968, by Stanley Landsman, is a container with a light source, fixed images, and reflective surfaces that allow one to see the images out to infinity. The piece was in an exhibition, "Light: Object and Image," at the Whitney Museum of American Art. (Leo Castelli Gallery, New York.)



And in these tasks, life scientists, physical scientists, physicians, and engineers all work on a more equal footing.

A somewhat similar integration of psychologists and engineers is discernible in the field once called human engineering, then human factors in engineering, and now man–machine systems.

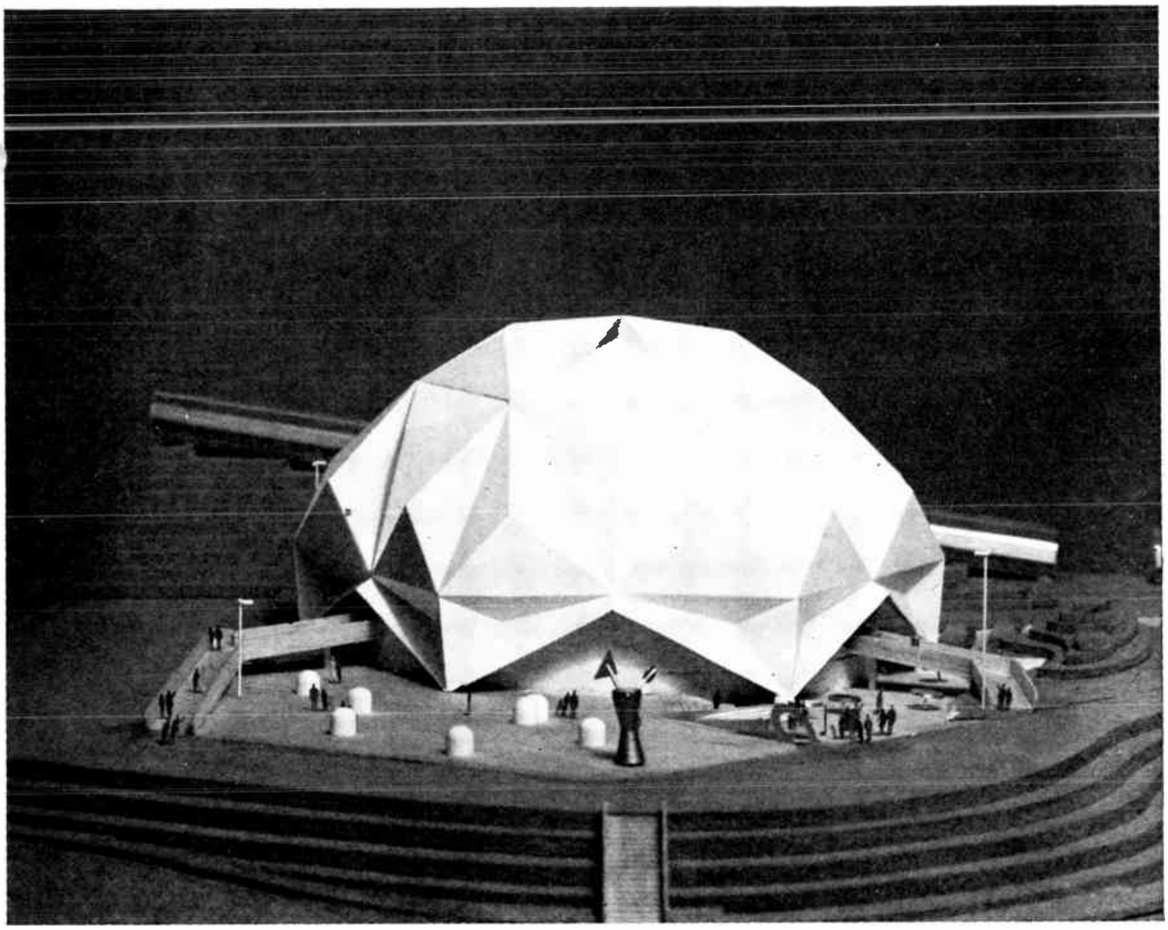
This brief description is not meant to minimize the difficulties that have existed and that still exist in the growth of such collaborations, but their example gives one a basis for projecting that some similar stages of evolution could come about in art and technology. The parallel can by no means be exact, however, or carried through to a trustworthy conclusion, especially as one begins to weigh the unique character and history of the artist in society.

It is only when one once again reassesses what is common knowledge—namely, that technology, in less than a generation, has moved to a central role in our society, and that *it* now holds a charismatic force, both magical and threatening, in the minds of most people; that technology occupies a position of authority the Church, the Monarchy, the State, once held—that the interest of artists in this crucial and governing phenomenon becomes more plain. The history of art shows that the artists worked off the tensions set up within society; artists created successively the powerful iconography for Church, for Monarchy, and for social and humanist revolutions. The artists' interest in the instrument that has succeeded those entities seems, then, quite natural. It seems, in fact, to fulfill a requirement that Gyorgy Kepes spoke about at the dedication of the new Center for Advanced Visual Studies at M.I.T.—it answers "historical necessity."

Epilogue—some hard questions

As one considers the prospect of art and technology, there arise many questions that are perhaps unanswerable to everyone's satisfaction, but to which we all might give thought.

Is it art? There will be those who, nurtured on traditional modes of art (illusionistic painting and sculpture, a music made of notes, dance executed by people and not by objects, writing scaled on a sociological and psychological illusionism), will find that the emergence of art grounded in technology causes feelings of pain. Works that move, light up, buzz, whir, that are ordered by the artist from a factory,



AT EXPO 70 in Osaka, Japan, an international collaboration between six artists (four U.S. and two Japanese), four assisting artists, and 12 engineers, will design a unique environment for the Pepsi-Cola pavilion shown here in a model. The dome structure, to be surrounded by a persisting "engineered" cloud, will house a 28-meter-diameter spherical mirror. Through the use of complex optical, sound, and radio effects, the visitor will be able to move freely through the pavilion to find different sensory experiences as he chooses. E.A.T. is administering the design and construction of the pavilion and is orchestrating the work of the collaborating artists and engineers.

music composed of whistles, rumblings, industrial noises, and long silences, and so on, may arouse repugnance, suspicion, incredulity, discomfort, hostility, and rage—in short, a spectrum of feelings that might be expressed by the formula: "But do they mean to call that stuff *art*?"

Although we must abide here with little more than the observation that the art of this century has undergone successive radical transformations, it may be amusing to note that art critics, too, sometimes share the sentiments of the "great unwashed." O'Doherty, for instance, in his collection of insightful essays, writes that there is a "deep unease among critics, historians, and some artists as to what art is all about—a more difficult problem than usual nowadays when anyone with any honesty has to admit he just doesn't know." [2] There is no dearth of "explanations" about what is happening. A recent work by Jack Burnham, *Beyond Modern Sculpture*,[6] probes the reasons behind why sculpture has undergone such radical transformations. Burnham, trained in both art and engineering, stresses the transfiguring role of technology in moving sculpture away from its traditional, inanimate craft-oriented modes and toward its replacement with life-simulating systems through the use of technology. Another recent work, *The Science of Art*,[7] by Robert Mueller, an engineer turned artist, takes a cybernetic-oriented look at the relation of art and science. In *Against Interpretation*,[8] the collected critical pieces of Susan Sontag, one can read a provocative and persuasive "interpretation" of the thrust of recent artistic events, especially as they center about New York. To catch the lyric voice of one of the major influences of these events, one should read John Cage's recent book, *A Year From Monday*.[9] One can trace the roots of modern artistic interest and concern with new materials, real objects, machines, and so on in W. S. Rubin's *Dada, Surrealism, and Their Heritage*,[10] which includes a good treatment of such artists as Max Ernst and Kurt Schwitters. Serious students of the engineering-and-art phenomenon of our century will wish to study the *Vision + Value* series of volumes [11] edited by Professor Gyorgy Kepes of M.I.T., as well as his landmark book, *The New Landscape in Art and Science*.[12] When it was published in 1956, Kepes says, some art magazines refused to review it because they said that art and science were unmixable entities.

These books represent just a sampling.

What's in it for me? This is perhaps the crucial question that the engineer will ask, and it really ought not be answered on the intellectual level. The best answers will come only after the engineer actually engages in a project with an artist. Essentially, the message here is that an artist has a lot to offer an engineer (and his industry) if the engineer just keeps his eyes open and learns how to translate the artist's message.

As he reflects on it, an engineer may discover many reasons for *not* concerning himself with artists and their aims: Such projects may seem silly, gimmicky, useless, difficult, unrewarding, child's play, a waste. It may seem, too, that it would take a stroke of luck to find an artist with whom one would be compatible in a working relationship.

The most respectable reasons are that at the level at which the artist perforce encounters science and technology it is likely that the technical problems will not be exciting enough to engage the engineer's interest; moreover, any engineer worth his salt is likely already expending all his time and energy on a project to which he is committed. Likewise, for the industrialist who considers giving support to artist–engineer work on company time on artistic projects, the prospect of an interesting return, other than a vague and intangible prestige, may seem marginal.

Despite such possibilities, however, collaboration has many appeals, which Klüver has pointed out on numerous occasions: new types of people to work with; new horizons of ideas; a different way of looking at engineering; the satisfaction of carrying something through outside of the goal structures of a profit-making organization, and the accompanying sense of special responsibility that one may miss in a usual engineering team effort; the opportunity to learn about art at first hand; the opportunity to learn how artists may work differently from engineers; and the contrary experience, finding out how much the techniques and processes of art and engineering may be alike.

One of the most serious and most intriguing possibilities is that engineering and engineers will find brand new goals. Engineering goals have never been fixed; in its deepening interaction with society, modern technology has had imposed on it new goals—Cold War systems, space systems, biomedical investigations—and it undoubtedly will continue to find new goals, as in transportation, and in ways of confronting the urban crises. The promise that art holds for engineering in this respect seems strong indeed.

In the clutch, however, each engineer must find his own unique answers. Engineer Fred Waldhauer, for instance, who participated in the 9 Evenings and in other collaborations, says that he got over his idea that art was a "precious thing," and that he felt "freer" as a result of his close association with artists. He feels that artists "take more chances" in their search for limits, and that they do not seem as daunted by the possibility of failure or of adverse criticism as are engineers. Moreover, the artist and the engineer working with him learn to tolerate a higher level of uncertainty. Rauschenberg, too, stresses this point: in any really critical important development, he feels it is important to give up ways of working in which you have been comfortable. To enrich one's environment and one's life, one must take risks. This is why the more arduous and difficult path of collaboration i preferable to the artists' mere exploitation of the given technology. As Waldhauer puts it, "If the artist and engineer understand one another, then the artist can use the *real* technology rather than just using it as a found object."

Engineer Leonard Robinson, with a long career behind him in systems engineering, says that although he paid little heed to art during his engineering career, he found that in collaboration he was "turned on" for the first time in his life. It is clear from his interviews and writings that he was extraordinarily impressed by the character and the commitment of the artists with whom he worked—"they were beautifully patient, believing and understanding." Robinson says of the 9 Evenings that if from the engineering point of view, they were "crude and the results imperfect, this should encourage other engineers and artists to come forward to improve upon that which was started. When they do this, the real purpose of the 9 Evenings will have been realized."

Finally, many of the engineers who collaborated exclaim over the unexpected fun of working with artists—the sheer joy of the enterprise—which, after all the practical, ethical, philosophical, and other lofty reasons and arguments are weighed, seems like the most sensible and satisfying reason of all.

The writer acknowledges the pleasure and stimulation he ha experienced in his discussions about art and technology with the founders and others associated with E.A.T: Billy Klüver, Robert Rauschenberg, Francis Mason, Julie Martin, and Susan Hartnett. And with many others, all of whom cannot be listed, the delight and indebtedness has been no less: Öyvind Falhström, Hans Haacke, Hilton Kramer, Robert Mueller, Max Neuhaus, Brian O'Doherty, Leonard Robinson, Harold Tovish, Fred Waldhauer, Simone Whitman.

REFERENCES

1. Whitman, S., "Theater and engineering: An experiment—1. Notes by a participant," and Klüver, B., "2. Notes by an engineer," *Artforum*, vol. 5, pp. 26–33, Feb. 1967.

2. O'Doherty, B., *Object and Ideas: An Art Critic's Journal 1961–1967.* New York: Simon & Schuster, 1967.

3. Rauschenberg, R., Private communication.

4. Kirby, M., *Happenings.* Scripts and productions by Jim Dine, Red Grooms, Allen Kaprow, Claes Oldenburg, and Robert Whitman. New York: Dutton, 1965.

5. Glueck, G., "Magic unrealism in Kansas City," *The New York Times*, p. 58, May 27, 1968.

6. Burnham, J., *Beyond Modern Sculpture: The Effects of Science and Technology on the Sculpture of This Century.* New York: Braziller, 1968.

7. Mueller, R. E., *The Science of Art: The Cybernetics of Creative Communication.* New York: Day, 1967.

8. Sontag, S., *Against Interpretation.* New York: Dell, 1965.

9. Cage, J., *A Year From Monday: New Lectures and Writings.* Middletown, Conn.: Wesleyan University Press, 1968.

10. Rubin, W. S., *Dada, Surrealism, and Their Heritage.* New York: The Museum of Modern Art, 1968.

11. Kepes, G., *Vision + Value Series: Education of Vision; Structure in Art and in Science; The Nature and Art of Motion; Module, Proportion, Symmetry, Rhythm; The Man-Made Object: Sign, Image, Symbol,* 6 volumes, New York: Braziller, 1965–1966.

12. Kepes, G., *The New Landscape in Art and Science.* Chicago: Paul Theobald, 1962.

Analytical uses of energy balances

Energy balances have provided information concerning forecasting, resource adequacy and efficiency, and consumption shifts that might otherwise remain inaccessible

William A. Vogely *U.S. Department of the Interior*

Energy balances necessitate energy requirements and sources to be expressed in commensurate units—usually heat. If these balances are complete, the flow of energy through the economy can be traced in such units, and the transformation functions of one form or source of energy to another can be specified. Thus, energy balances permit analysis of the entire energy sector on a whole, rather than on a commodity by commodity, basis. As such, they are capable of being used analytically to solve a number of problems that would otherwise be beyond comprehensive analysis. In this article, four such analytical uses are described and applied.

As a prelude, and in order to lend perspective and balance to the data tabulated in this article, a history of the consumption of energy resources in the United States for the five-year periods from 1950 to 1965 is given in Table I.

Forecasting

The use of energy balances as a forecasting tool is illustrated in Table II. Forecasts are made based on the differential demands for energy by end-use categories, and the forecasts of the sources of this energy are made based upon specific technologic trends that are occurring in the energy sector. The forecasts contained in Table II were made by relating energy demands by sector to general economic variables reflecting the growth of demand in these sectors, and the supply distribution was based upon assumptions about nuclear energy and technology; specifically, that no major technological breakthroughs would occur other than the advent of nuclear energy.

Another useful characteristic of the energy-balance approach is that it permits contingency forecasting in addition to probability forecasting. Under contingency forecasting, one can state a specific contingency—for example, the advent of the electric automobile to replace the internal-combustion engine in the United States by the year 2000—and can trace the impact of this contingency to energy sources and energy forms. Such contingency forecasting is useful in identifying potential problems and in pointing to efficient solutions to these problems. Table III shows the transportation sector on an all-electric basis. Note the change in total gross energy inputs (Fig. 1), and the implication for energy sources (Fig. 2).

Resource adequacy

The characteristic of contingency forecasting allows energy balances to be used to test the outside limits of resource adequacy. The first example of this use of energy balances is contained in "Energy R&D and National Progress," published by the U.S. Government in 1965. There the energy model of an all-electric economy was developed using the technological substitution functions between energy in the form of electricity, and energy as a direct fuel. Using this model and the resulting energy balance, the total requirements of source fuels are estimated on the basis of a single source fuel for the generation of the requisite electricity. Outside limits were thus established on fuel use and these were compared with the resource base to determine adequacy (Fig. 3). The analysis is reproduced in Table IV.

The reason for taking this approach to resource adequacy is the high degree of substitutability that exists between energy sources. Such substitutability makes it impossible to project demands for a single energy source independent of the demands for all energy sources taken together. As stated in the energy R&D report, "The extreme degree of substitutability makes it almost pointless to consider projections for specific fuels in the period to [the year] 2000." Therefore, to get an approach to an adequacy evaluation, simple all-fuel models are constructed.

Resource efficiency

The energy balance actually observed in the economy is one that results from the institutional and economic forces operating upon that economy. In a free economy, where such decisions are made based upon the relative cost and consumer demands for energy sources, the energy balance that occurs will reflect these decisions efficiently. However, there is another conception of efficiency that the energy balance approach allows us to quantify and to describe.

The transformation of various energy sources into final usable form requires different gross energy inputs. As shown in Table II, the gross input to the U. S. economy for the year 1966 was 13 560 × 10^{12} kilogram calories (56 600 × 10^{12} kilojoules). This is the energy content of the coals, natural gas, and petroleum consumed directly, plus the electricity generated by water converted at central power station efficiency, and nuclear generation taken at the caloric equivalent of the kilowatthours produced.

I. U.S. energy balance—past consumption of energy resources for fuel and power as direct fuels and purchased electricity (10^{12} kilogram calories)[1]

Consuming Sectors	Anthracite	Bituminous Coal and Lignite	Natural Gas (dry)	Petroleum and Natural Gas (liquid)	Hydropower	Nuclear Power	Total Gross Energy Inputs	Utility Electricity Generated and Distributed	Total Sector Energy Inputs
Household and commercial[2]									
1950	166	568	414	656	—	—	1 804	138	1 942
1955	83	364	718	853	—	—	2 018	215	2 233
1960	43	214	1 076	1 056	—	—	2 389	318	2 707
1965	20	138	1 394	1 195	—	—	2 747	489	3 236
Industrials[3]									
1950	32	1 469	939	558	—	—	2 998	141	3 139
1955	13	1 461	1 244	694	—	—	3 412	254	3 666
1960	14	1 221	1 584	674	—	—	3 493	329	3 822
1965	12	1 421	1 937	711	—	—	4 081	413	4 494
Transportation[4]									
1950	5	424	33	1 710	—	—	2 172	6	2 178
1955	3	117	64	2 296	—	—	2 480	5	2 485
1960	2	21	91	2 614	—	—	2 728	5	2 733
1965	—	5	131	3 070	—	—	3 206	5	3 211
Electricity generated, utilities[5]									
1950	23	538	164	167	403	—	1 295	285	—
1955	21	857	301	129	377	—	1 685	474	—
1960	18	1 055	450	142	447	1	2 113	652	—
1965	14	1 468	605	187	517	10	2 801	907	—
Miscellaneous and unaccounted for									
1950	29	—	—	85	—	—	114	—	114
1955	31	—	—	125	—	—	156	—	156
1960	36	—	—	132	—	—	168	—	168
1965	37	—	—	128	—	—	165	—	165
Total gross energy									
1950	255	2 999	1 550	3 176	403	—	8 383	—	7 373
1955	151	2 799	2 327	4 097	377	—	9 751	—	8 540
1960	113	2 511	3 201	4 618	447	1	10 891	—	9 430
1965	83	3 032	4 067	5 291	517	10	13 000	—	11 106

[1] Excludes energy sources used for raw materials other than fuel and power. Raw material uses in 1950 were equivalent to 357 × 10^{12} kilogram calories; 1955, 430 × 10^{12} kilogram calories; 1960, 566 × 10^{12} kilogram-calories; and 1965, 668 × 10^{12} kilogram calories.

[2] Energy inputs into the household and commercial sector, used as direct fuels in equipment averaging 75 percent efficiency, and as utility electricity at 100 percent thermal efficiency.

[3] Energy inputs into the industrial sector, used as direct fuels in equipment averaging 75 percent efficiency, and as utility electricity at 100 percent thermal efficiency.

[4] Energy inputs into the transportation sector, used as direct fuels in vehicles with a maximum efficiency of 25 percent (design point gasoline engines at 70 miles per hours) (113 km/h).

[5] Energy inputs into the electric utility sector, used as direct fuels at central power stations (conventional thermal, hydropower, and nuclear power) at 22 percent average efficiency in 1950, 28 percent in 1955, 31 percent in 1960, and 32 percent in 1965.

FIGURE 1. Gross energy inputs: probability forecast in terms of contingency forecast within all-electric transportation sector.

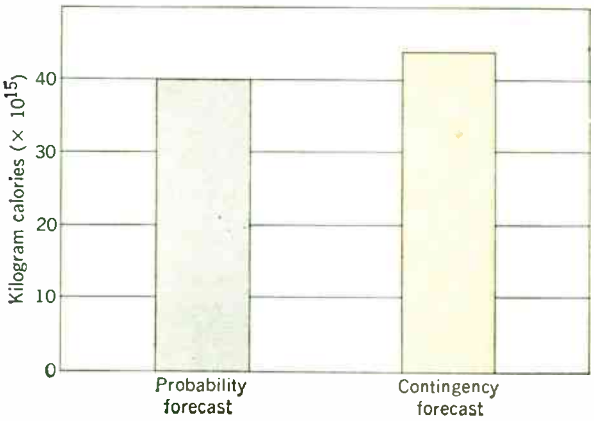
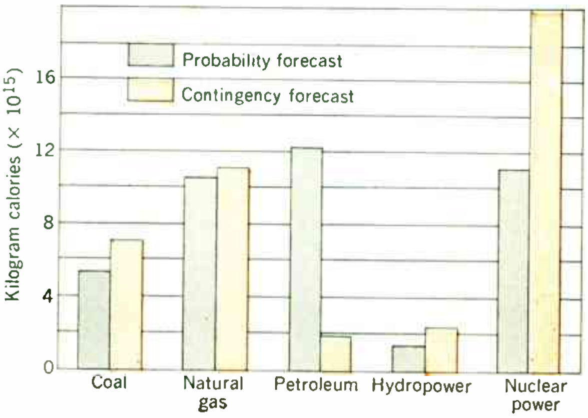


FIGURE 2. Energy sources: probability forecast in terms of contingency forecast within an all-electric transportation sector.



II. U.S. energy balance—future consumption of energy resources for fuel and power as direct fuels and purchased electricity (10^{12} kilogram calories)[1]

Consuming Sectors	Anthracite	Bituminous Coal and Lignite	Natural Gas (dry)	Petroleum and Natural Gas (liquid)	Hydropower	Nuclear Power	Total Gross Energy Inputs	Utility Electricity Generated and Distributed	Total Sector Energy Inputs
Household and commercial[2]									
1966	37	136	1 457	1 284	—	—	2 914	530	3 444
1980	13	69	2 521	1 443	—	—	4 046	1 397	5 443
2000	—	—	4 807	504	—	—	5 311	5 053	10 364
Industrials[3]									
1966	22	1 438	1 981	703	—	—	4 144	447	4 591
1980	7	1 443	2 970	1 046	—	—	5 466	939	6 405
2000	—	504	4 413	867	—	—	5 784	2 721	8 505
Transportation[4]									
1966	—	4	137	3 187	—	—	3 328	6	3 334
1980	—	—	176	5 228	—	—	5 404	12	5 416
2000	—	—	252	10 500	—	—	10 752	25	10 777
Electricity generated, utilities[5]									
1966	14	1 597	678	224	514	13	3 040	983	—
1980	7	3 351	750	217	763	1 028	6 116	2 348	—
2000	—	4 719	1 041	217	1 275	10 973	18 225	7 799	—
Miscellaneous and unaccounted for									
1966	—	—	—	134	—	—	134	—	134
1980	37	—	—	85	—	—	122	—	122
2000	—	—	—	—	—	—	—	—	—
Total gross energy			4 253						
1966	73	3 175	6 417	5 532	514	13	13 560	—	11 503
1980	64	4 863	10 513	8 019	763	1 028	21 154	—	17 386
2000	—	5 223	—	12 088	1 275	10 973	40 072	—	29 646

[1] Excludes energy sources used for raw materials other than fuel and power. Raw material uses in 1966 were equivalent to 612 × 10^{12} kilogram calories. For 1968 these are projected at 1231 × 10^{12} kilogram calories, and for 2000 at 2982 × 10^{2} kilogram calories.

[2] Energy inputs into the household and commercial sector, used as direct fuels in equipment averaging 75 percent efficiency, and as utility electricity at 100 percent thermal efficiency.

[3] Energy inputs into the industrial sector, used as direct fuels in equipment averaging 75 percent efficiency, and as utility electricity at 100 percent thermal efficiency.

[4] Energy inputs into the transportation sector, used as direct fuels in vehicles with a maximum efficiency of 25 percent [design point gasoline engines at 70 miles per hour (113 km/h)].

[5] Energy inputs into the electric utility sector, used as direct fuels at central power stations (conventional thermal, hydropower, and nuclear power) at 32 percent average efficiency in 1966, 38 percent in 1980, and 43 percent in 2000.

III. U.S. energy balance—the conventional energy model for the year 2000 adjusted for an all-electric transportation system. Consumption of energy resources for fuel and power as direct fuels and purchased electricity (10^{12} kilogram calories)[1]

	Anthracite	Bituminous Coal and Lignite	Natural Gas (dry)	Petroleum and Natural Gas (liquid)	Hydropower	Nuclear Power	Total Gross Energy Inputs	Utility Electricity Generated and Distributed	Total Sector Energy Inputs
Household and commercial[1]	—	—	4 805	504	—	—	5 309	5 052	10 361
Industrial[2]	—	504	4 411	867	—	—	5 782	2 720	8 502
Transportation[3]	—		—	—	—	—	—	6 398	6 398
Electricity generated, utilities[4]	—	8 535	1 879	395	2 306	19 837	32 952	14 170	—
Total gross energy	—	9 039	11 095	1 766	2 306	19 837	44 043	—	25 261

[1] Energy inputs into the household and commercial sector, used as direct fuels in equipment averaging 75 percent efficiency, and as utility electricity at 100 percent thermal efficiency.

[2] Energy inputs into the industrial sector, used as direct fuels in equipment averaging 75 percent efficiency, and as utility electricity at 100 percent thermal efficiency.

[3] Energy inputs into the transportation sector, using utility electricity in vehicles with an average efficiency of 42 percent (this is the estimated efficiency for a battery-electric vehicle system based on a battery charging efficiency of 80 percent—conversion of ac power to dc; battery discharge effectiveness, 70 percent; and motor and controls efficiency, 75 percent).

[4] Energy inputs into the electric utility sector, used as direct fuels at central power stations (conventional thermal, hydropower, and nuclear power) at 43 percent average efficiency in 2000.

Source: Division of Economic Analysis, Bureau of Mines, U.S. Department of the Interior.

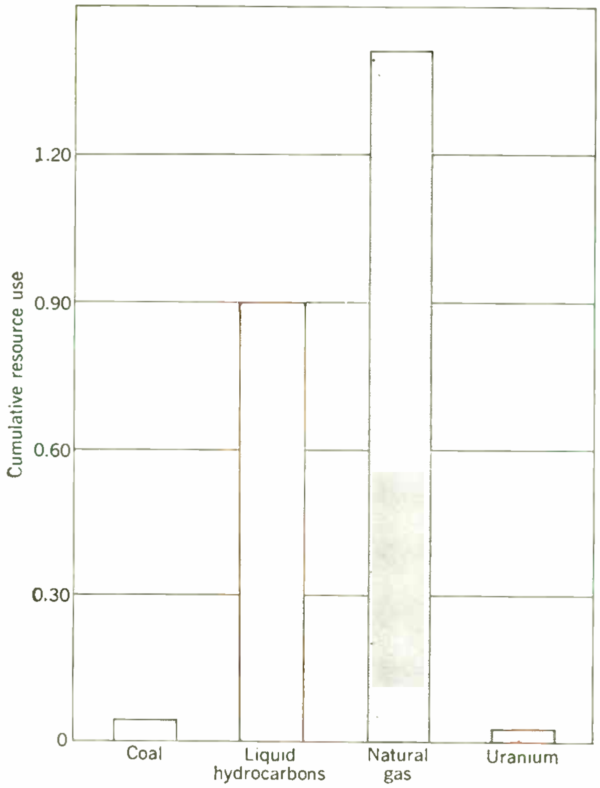


FIGURE 3. Cumulative resource use as compared with total resources (1960–2000) for an all-electric economy.

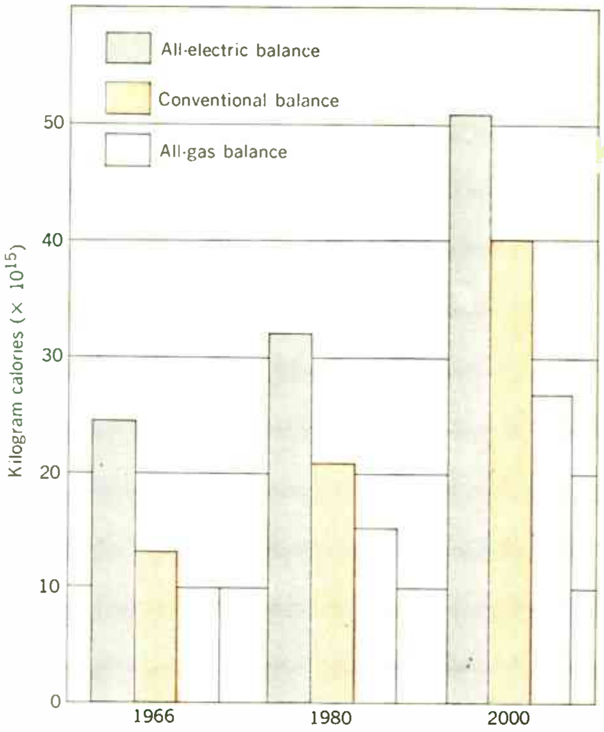


FIGURE 4. Gross energy inputs for 1966, 1980, and 2000.

There is a trend toward increased use of electricity in the U.S. economy. What are resource implications, from the point of view not of a single energy source, but of efficiency of gross energy input utilization in a move toward an all-electric economy? Such a calculation is performed in Table V, in which we have assumed an all electric economy with central generation by a coal at a thermal efficiency that will reach 44 percent in the year 2000. The gross energy requirements of this model are higher than that in Table II.

There are technologies now known, but not economic, that hold out possibilities of much greater efficiency of energy utilization. Table VI gives a model of an all-gas economy using small hydrocarbon fuel cells as a source of all power. This model shows a distinct resource saving of quite substantial proportions. These implications are clearly indicated in Fig. 4.

The picture from the gross energy input side, however, does not express the real savings of efficiency that are possible. Mineral resources are produced from the earth at widely varying efficiencies. For example, the average recovery of liquid petroleum from the ground is about 35 percent; of natural gas, 80 percent; and coal, about 65 percent. A technology that permits the use of natural gas will give greater efficiency in the utilization of the resource in the ground than one that is dependent upon petroleum.

This kind of an analysis has several important applications. With it, for example, one can evaluate the worth or value of an advanced recovery technique for

IV. Comparison of U.S. fuel resources with various requirement estimates

Consuming Sectors	Coal, metric tons	Liquid Hydrocarbons (crude oil and natural gas liquid), metric tons	Natural Gas, cubic meters	Uranium, metric tons
All-electric energy requirements (cumulative, 1960–2000)	112×10^9	76×10^9	94×10^{12}	18.144
Known recoverable reserves	200×10^9	7×10^9	8×10^{12}	293.019 *
Undiscovered recoverable resources	not estimated	29×10^9	34×10^{12}	317.513
Difference	$+88 \times 10^9$	-40×10^9	-52×10^{12}	+592.388
Total resources	$3\,828 \times 10^9$	85×10^9	66×10^{12}	billions
Difference	$+3.716 \times 10^9$	$+9 \times 10^9$	-28×10^{12}	—
RFF cumulative projection (medium)	23×10^9	29×10^9	28×10^{12}	small

* Includes 181.436 metric tons already mined and still available

Source: "Energy R&D and National Progress," June 1964, p. 28. Prepared for the Interdepartmental Energy Study by the Energy Study Group under the direction of Ali Bulent Cambel.

V. U.S. all-electric energy balance—gross consumption of energy resources for fuel and power (10^{12} kilogram calories)[1]

Consuming Sectors	Bituminous Coal	Total Gross Energy Inputs	Total Electricity Generated and Distributed
Household and commercial[2]			
1966	—	—	2 716
1980	—	—	4 431
2000	—	—	9 037
Industrial[2]			
1966	—	—	3 556
1980	—	—	5 038
2000	—	—	7 059
Transportation[3]			
1966	—	—	1 981
1980	—	—	3 216
2000	—	—	6 400
Electricity generation[4]			
1966	24 713	24 713	8 253
1980	31 990	31 990	12 685
2000	50 800	50 800	22 496
Total gross energy			
1966	24 713	24 713	8 253
1980	31 990	31 990	12 685
2000	50 800	50 800	22 496

[1] Total sector requirements are met from central-station conventional utility plants with coal as source fuel.

[2] Energy inputs into the household and commercial, and industrial sectors, using utility electricity at 100 percent thermal efficiency.

[3] Energy inputs in the transportation sector, using utility electricity in vehicles with an average efficiency of 42 percent (this is the estimated efficiency for a battery-electric vehicle system based on a battery charging efficiency of 80 percent—conversion of ac power to dc; battery discharge effectiveness, 70 percent; and motor and controls efficiency, 75 percent).

[4] Energy inputs into the electric utility sector, used as direct fuels in thermal power stations, with coal as source fuel at 33 percent average efficiency in 1966, 40 percent in 1980, and 44 percent in 2000.

VI U.S. all gas energy balance—gross consumption of energy resources for fuel and power (10^{12} kilogram calories)[1]

Consuming Sectors	Natural Gas (dry)	Total Gross Energy Inputs and Total Sector Energy Inputs
Household and commercial[2]		
1966	3 621	3 621
1980	5 907	5 907
2000	12 048	12 048
Industrial[3]		
1966	4 741	4 741
1980	6 716	6 716
2000	9 411	9 411
Transportation[4]		
1966	1 664	1 664
1980	2 708	2 708
2000	5 388	5 388
Total gross energy		
1966	10 026	10 026
1980	15 331	15 331
2000	26 847	26 847

[1] Natural gas assumed a single source fuel and all sector requirements assumed to be met from gas-fired fuel cells with no purchased utility electricity.

[2] Energy inputs into the household and commercial sector, using natural gas as direct fuel in hydrocarbon–air fuel cells with 50 percent heat recovery and a maximum efficiency of 75 percent. No purchased utility electricity.

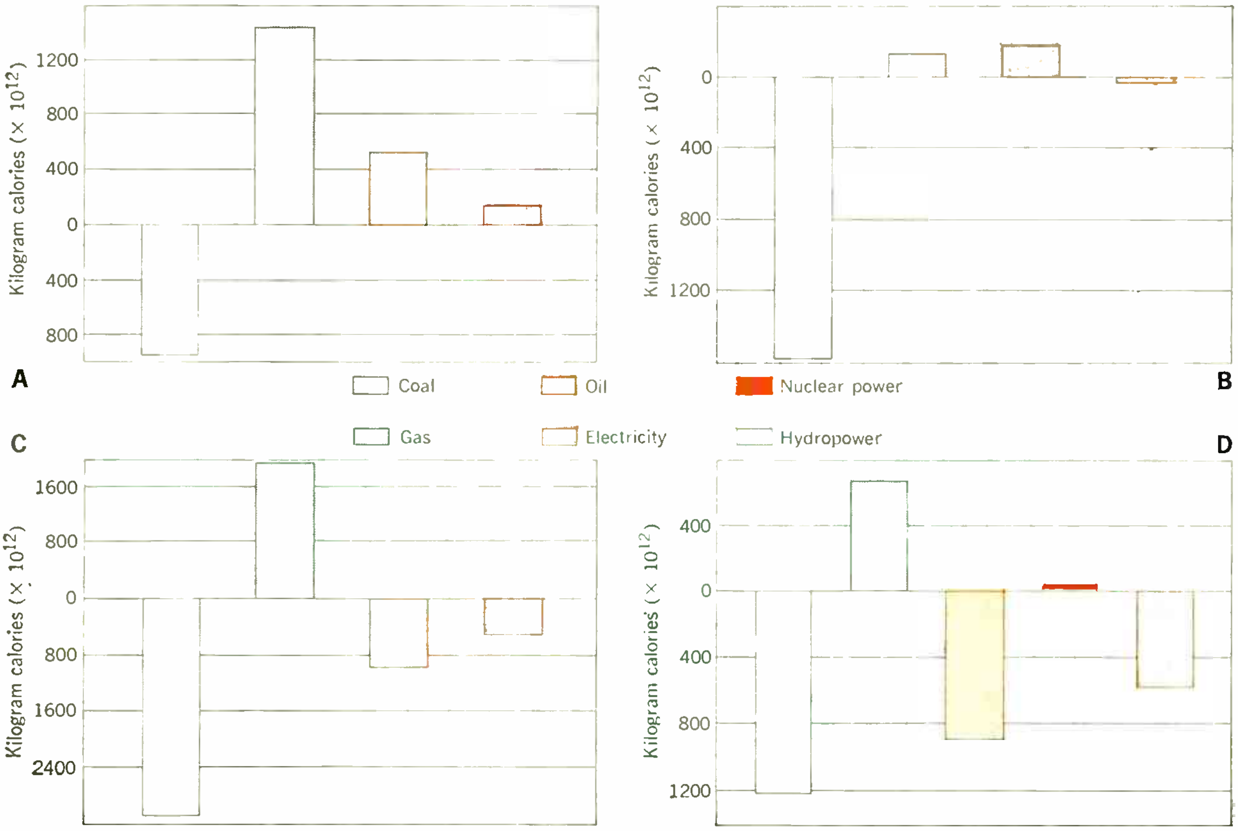
[3] Energy inputs into the industrial sector, using natural gas as direct fuel in hydrocarbon–air fuel cells with 50 percent heat recovery and a maximum efficiency of 75 percent. No purchased utility electricity.

[illegible]nergy inputs into the transportation sector, using natural gas as direct fuel in hydrocarbon–air fuel cell powered vehicles with an estimated maximum [illegible]iciency of 50 percent (no heat recovery). No purchased utility electricity.

VII. U.S. energy consumption shifts—1947–1966 (10^{12} kilogram calories)

Consuming Sectors		1966 Consumption, C_t	1947 Consumption, C_o	Market Growth Shift, Y_s	Substitution Shift, D_s (elasticity = 1)	Technology Shift, T_s (elasticity = 0)	Technology Shift, T_s (elasticity = 1)
Household and commercial							
Coal		145	652	238	201	−745	−946
Gas		1 498	284	104	−344	1 111	1 454
Oil		1 218	567	207	− 98	443	542
Electricity		530	99	35	241	395	154
	Total	3 391	1 602	585	0	1 204	1 204
Transportation							
Coal		5	758	826	5	−1 579	−1 584
Gas		139	neg	neg	0	139	139
Oil		3 222	1 452	1 532	5	188	185
Electricity		6	8	8	−9	−10	−2
	Total	3 372	2 218	2 416	0	−1 262	−1 262
Industrial							
Coal		1 426	1 768	2 437	316	−2 779	−3 095
Gas		1 968	758	1 045	−1 803	165	1 968
Oil		1 328	627	855	805	−165	−971
Electricity		447	116	160	681	172	−509
	Total	5 169	3 269	4 507	0	−2 607	−2 607
Electric generation (utility)							
Coal		1 598	502	1 823	478	−728	−1 206
Gas		678	97	353	−450	228	678
Oil		228	118	428	575	−318	−892
Nuclear		13	none	none	0	13	13
Hydro		514	358	1 334	−603	−1 187	−585
	Total	3 031	1 085	3 938	0	−1 992	−1 992

FIGURE 5. Energy consumption shifts, 1947–1966. A—Household and commercial technologic shifts (elasticity = 1). B—Transportation technologic shifts (elasticity = 1). C—Industrial technologic shifts (elasticity = 1). D—Electricity generation technologic shifts (elasticity = 1).



liquid petroleum from the point of view of conservation of energy. Also, the implication on resources can be examined with respect to the conversion of an energy source to another form, such as the gasification or liquefication of coal.

Consumption shift analysis

A time series of energy balances permits the application of a shift technique to understand the forces at work in the changing patterns of consumption and production of energy. Under this analysis, the behavior of each of the consumption sectors, with respect to a given energy source supply, can be described by a series of equations. These equations take the following form:

$$C_t = C_o + (Y_x, D_x, T_x)$$

where C_t is the consumption in the current period; C_o, the consumption of the base period; Y_x, the market growth shift; D_x, the substitution shift between energy sources attributed to price; and T_x, the technology shift resulting from the changing technology in energy utilization within the sector. The market growth shift Y_x is determined by applying a measure of change of size to the market to the base-period consumption.

The price substitution shift D_x is determined by applying standard econometric techniques to determine the substitution elasticity between the various forms of energy serving this specific sector. In Table VII, the elasticity is assumed to be either one or zero, and the computations are based upon this assumption. This is believed to bracket the true value of T_x. The technology shift T_x is a residual; that is, the unexplained portion of the shift. However, in specific sectors we can delineate quite precisely what has occurred. In the household and commercial sector [Fig. 5(A)], T_x is negative for coal and positive for other fuels; hence, the total is positive. The negative shift in coal and part of the positive shift in gas and oil reflect the decline in coal for space heating. This resulted largely from the geographic spread of gas, and also of oil, through the extension of the pipeline network to all parts of the economy. The large positive shift for the total reflects the increase in energy use per household, because of the development and spread of appliances such as dishwashers, television, garbage disposals, washing machines, dryers, and myriad smaller items, as well as the increase in air-conditioning devices. Note that all these changes are of the nature of new technology either in the transmission or utilization of energy.

In the transportation sector [Fig. 5(B)], the large negative shift in coal results from the change over to diesel power for the railroads. The positive shift in gas reflects the use of gas in gas pipeline transportation. The relatively small positive shift in oil is a measure of the increased efficiency in energy utilization, since it contains both an equivalent positive shift as a result of the coal decline, and negative forces arising from increased efficiency.

In the industrial sector [Fig. 5(C)], similar forces were at work. The negative total indicates increased efficiency, and the positive shift in gas measures the geographic spread in the marketing areas. The identical pattern is evident in the electrical generation sector [Fig. 5(D)].

Shift analyses untangle the various factors that affect the present level of energy consumption over any period of time. They analyze more precisely what is really happening in the energy markets. With an understanding of the sources of shifts in supply, much more realistic projections can be made, and misconceptions that would follow from looking only at historical trends can be avoided.

Summary

Four major analytical uses of energy balances have been described, and appropriate examples furnished. *Forecasting*: A projection of the United States energy balance to A.D. 2000 was given. It indicates a threefold increase in energy inputs by the year 2000, with a substantial shift to electricity and to nuclear power. *Resource adequacy*: Models of all fuel economies were used to test outside limits of resource adequacy. Such an approach is needed because of the high levels of substitution possible between energy sources. *Resource efficiency*: Two approaches to efficiency were developed. The first indicates the cost in resource terms of the trend toward secondary energy use; the second illustrates the impact of resource—saving technology on gross energy inputs. The efficiency of actual production from resources in the ground is also brought into the analysis. *Consumption shift analysis*: A shift model of energy consumption for 1947–1966 is developed, as well as a behavioral equation for each source by sector. The equation permits the historical changes to be distributed among three factors: market growth, price substitution, and a residual technology change. The analysis indicates that growth in the market is the major contributor to change, but that increased efficiency negatively affects the growth in three sectors, while new uses reinforce the change in the household and commercial sector.

Based on a paper presented at the 1968 World Power Conference, Moscow, U.S.S.R., August 20–24. The original paper will appear in the proceedings of the conference, and is copyrighted by the Soviet National Committee.

William A. Vogely is currently the assistant director, mineral resource evaluation, at the Bureau of Mines, U.S. Department of the Interior, and has been engaged in mineral economics analysis for the past 20 years. Receiving the A.B. degree from Kenyon College, Gambier, Ohio, and the Ph.D. degree from Princeton University, he began teaching economics at Kenyon College in 1947, continuing at the U.S. Department of Agriculture Graduate School and George Washington University, where he is presently teaching. Dr. Vogely's career in economic research began in 1948, when he served as the only economist with the Bureau of Research and Statistics in New Jersey. He was consultant to the Department of Highways in Ohio and, for a year and a half, worked on a special research project at the Rand Corporation. For 15 months during 1960 and 1961, he was sent by the Department of the Interior as a delegate to the General Agreement on Tariffs and Trade meetings at Geneva, Switzerland, serving as the Interior member of the Trade Agreements Committee.

Integrated circuits for television receivers

Eizi Sugata, Toshihiko Namekawa *Osaka University*

Integrated circuits are being investigated for use in as many television circuits as feasible. In Japan, considerable work has been done by different groups within the TV industry to develop ICs that will permit cost reduction, increased reliability, and simplification of assembly-line operations. This article is a report on results of the concentrated efforts made by five major television set manufacturers to develop and produce black-and-white and color receivers, in collaboration with four universities, two institutes, and seven components manufacturers.

Since April 1966, five major television set makers, seven parts makers, four universities, and two institutes around Osaka have sponsored a committee to develop jointly a black-and-white and a color receiver utilizing integrated circuits (ICs) for as many functions as possible.

More ICs were employed in these prototype units than in any other TV receivers hitherto introduced [1-9]; these IC receivers are comparable in performance to conventional ones employing discrete components.

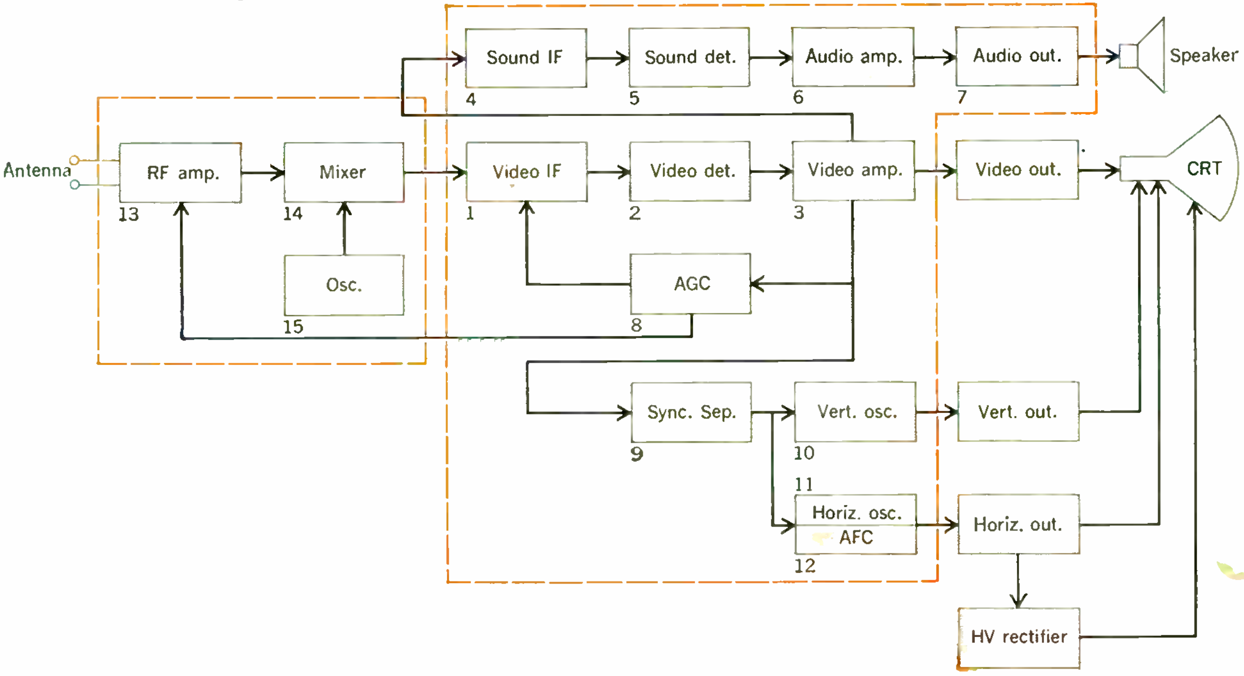
Miniaturization is meaningless for a home TV set, which uses a cathode-ray tube as a display device. Our main line of study has been concerned with increases in reliability and reductions in cost. Our objective has been pursued through the realization of the following research themes:

1. Incorporation of as many ICs as possible.
2. Optimum density of integration.
3. High reliability and low cost.
4. Simplification of assembly processes.

Once we succeed in improving reliability by adopting ICs, then we can advance to the problem of a solid-state display device.

The research and development in the application of ICs for television receivers has been going on for a little over two years. A prototype of a black-and-white receiver was completed in 1967 and of a color receiver in 1968. We have investigated, for IC implementation, all of the necessary circuits with the exception of those subject to high voltages and high power consumption. The circuits not considered were the UHF tuner, video output circuit, horizontal output circuit, vertical output circuit, high-voltage rectifier, and power source. The circuits that have been integrated are shown in Figs. 1 and 2.

FIGURE 1. Block diagram of black-and-white IC television receiver. The ICs are enclosed by the dashed lines.



Several of Japan's leading manufacturers and universities have joined forces in a concerted effort to expedite research and development work on the use of integrated circuits in black-and-white and color television receivers

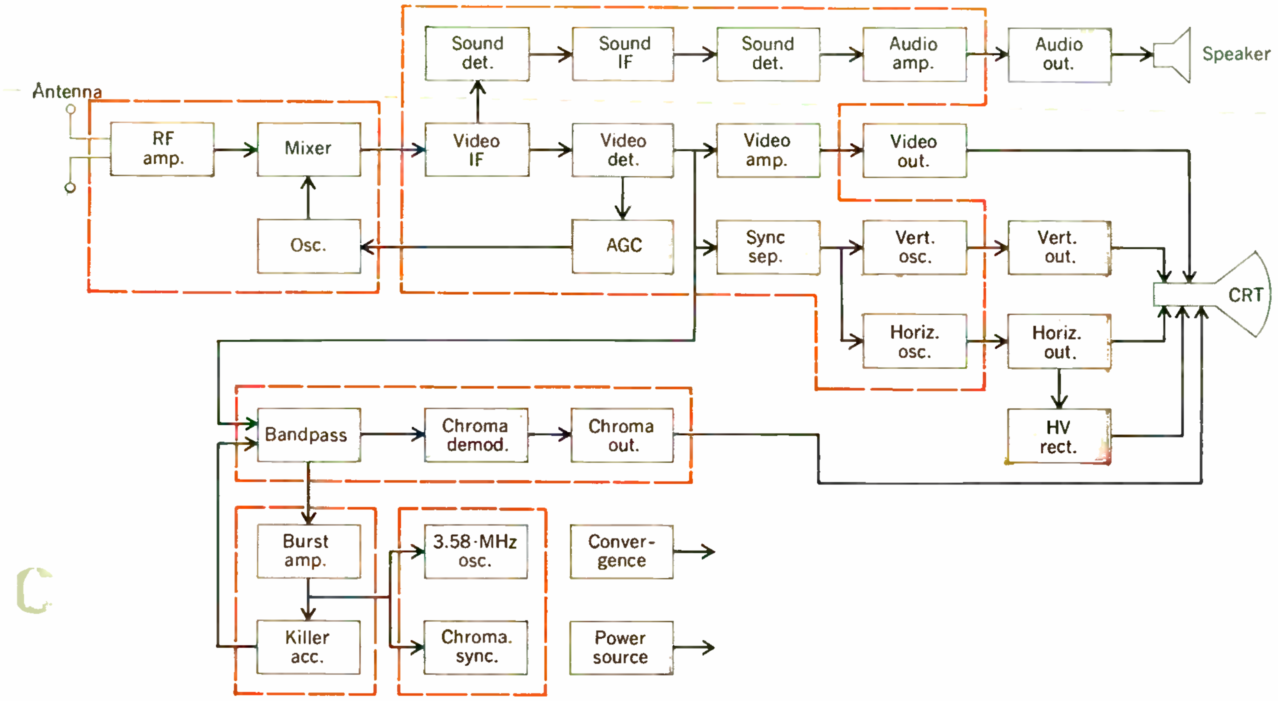
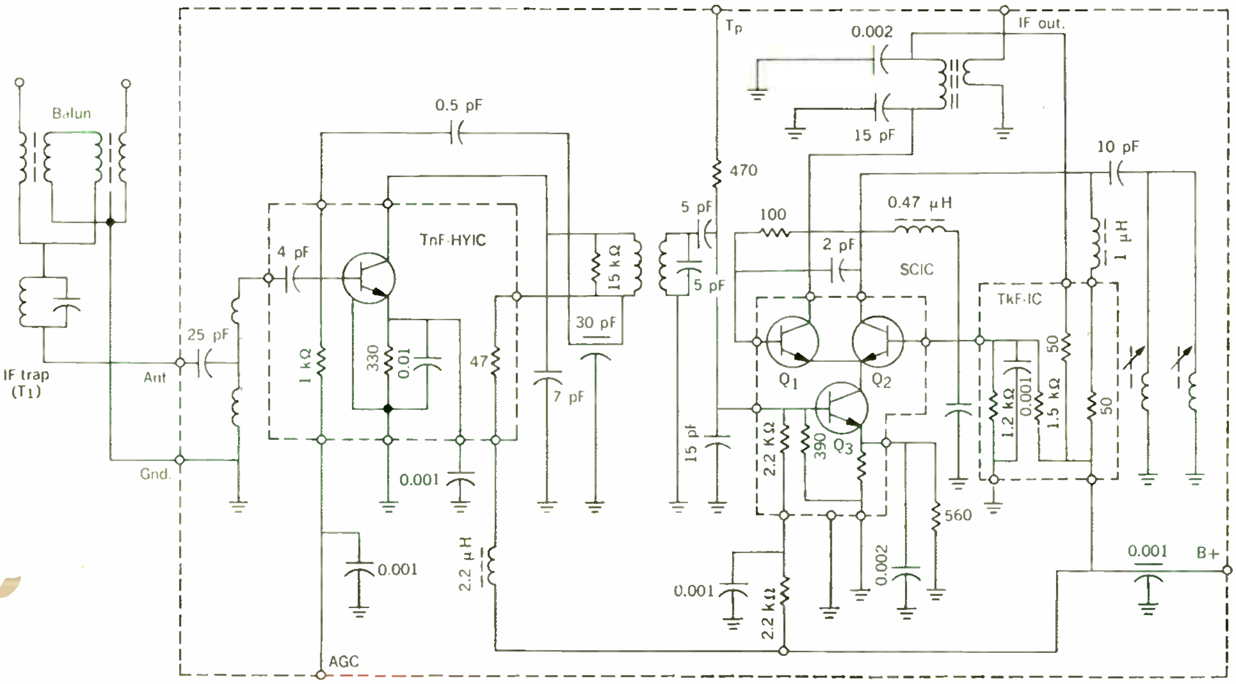


FIGURE 2. Block diagram of color IC television receiver. The ICs are enclosed by the dashed lines.

FIGURE 3. Tuner circuit, containing TnF-HYIC, TkF-IC, and SCIC. Resistance values are in ohms, capacitance values in microfarads, except where otherwise indicated.



VHF tuner

Figure 3 shows the circuit diagram of the tuner. In this figure we use the abbreviations SCIC, TnF-HYIC, and TkF-IC, for semiconductor (monolithic) IC, thin-film hybrid IC, and thick-film IC, respectively.

High-frequency amplifier circuit. In constructing the thin-film hybrid IC of a transistor, thin-film resistors, and ceramic capacitors, the circuit was arranged in such a way as to minimize mutual inductance of terminal lead wires and coupling by strap capacitances. In the input circuit, a thin-film inductor is used as an intermediate-frequency trap (T_1). Forward AGC (automatic gain control) is used in the amplifier.

The overall gain of this IC when incorporated in the high-frequency amplifier circuit is about 15 dB for low channels (90–108 MHz) and about 12 dB for high channels (170–222 MHz).

Mixer and local-oscillator circuit. A local oscillator is composed of a monolithic IC differential amplifier (Q_1 and Q_2). This circuit is a kind of Clapp oscillator. The frequency conversion is performed by the cascode connection of the transistors of the differential amplifiers and the current source of the IC (Q_1 and Q_3). In this circuit, Q_1 functions as a mixer and Q_3 as a high-frequency amplifier. The bias circuit consists of thick-film resistors and ceramic capacitors in addition to built-in diffused resistors.

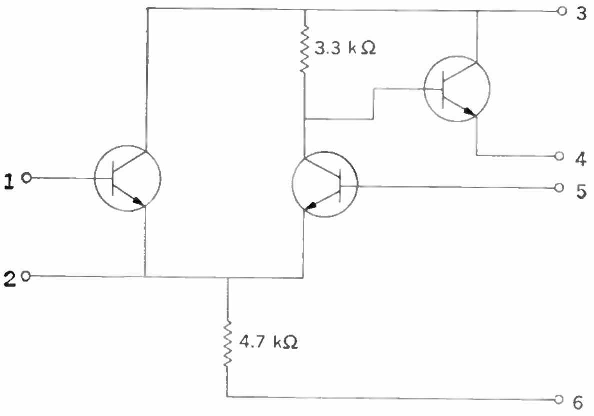
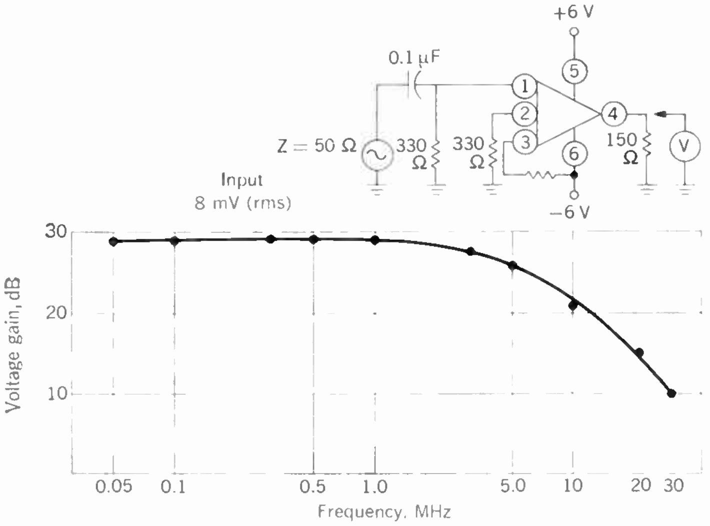


FIGURE 4. Multipurpose monolithic IC differential amplifier circuit diagram.

FIGURE 5. Frequency response of the differential amplifier shown in Fig. 4.



The available power gain of the monolithic IC, when incorporated in the converter circuit, is 22 dB for t low channels and 16 dB for the high channels. This tuner circuit is designed to obtain good performance at high frequencies by using present IC techniques.

Video and sound IF circuits

Multipurpose ICs were developed so that by using just two kinds of linear ICs it was possible to integrate the sound IF, sound detector, audio amplifier, synchronous pulse separator, video-frequency amplifier, horizontal oscillator, vertical oscillator, and AFC (automatic frequency control) circuits. To achieve these functions, either certain operating conditions were imposed or appropriate parts or feedback circuits were added to amplifiers. The simplest IC is shown in Fig. 4, and Fig. 5 graphs its freqency characteristic.

A new self-compensating resistor was developed for the purpose of increasing the yield of the monolithic linear IC. A built-in pinch resistor, it has the same cross section as that of the base region of the transistor in the same chip. Variation in current-amplification factor is compensated by the coherency of the values of bias resistor and h_{fe} of the IC transistor.

Newly developed thin-film hybrid ICs were used for the video amplifier in the interest of high performance and simplified assembly. Figures 6(A) and 6(B) show the design of the concentrated-type filter circuit and the overall frequency response of the video stage. high-pass filter and a low-pass filter were inserted at the front and back of the amplifier, and thus the number of inductors needed for the video IF stage was reduced. At the same time, assembly and adjustment were simplified by this configuration. Figure 6(C) shows the thin-film hybrid amplifiers cascaded in the stage.

The adjustment for the sound IF stage is eliminated by the use of ceramic resonators. In integrating consumer electronic equipment, these simplifications of adjustment are important factors for cost reduction.

The circuit diagram of the sound IF stage is given in Fig. 7. This block is composed of compound components, thick-film components, and monolithic ICs, with the cost kept as low as possible. Transistors Q_2 and Q_3 and their bias circuit were constructed in monolithic form. The conventional discriminator transformer was eliminated by using an especially developed piezoelectric ceramic resonator (X_1).

By applying dc feedback from the second sound IF stage to the first, dc stabilization—which suppresses the bias fluctuation of a transistor under large-signal conditions and results in limiter action—has been obtained. The specifications for the piezoelectric ceramic resonator (X_1) used as a discriminator for this application are as follows:

Resonance frequency	4.35 MHz
Capacitance ratio	7
Resonance impedance	100 ohms
Quality factor	65
Electronic capacitance	40 pF
Size	3 × 3 × 1.2 mm

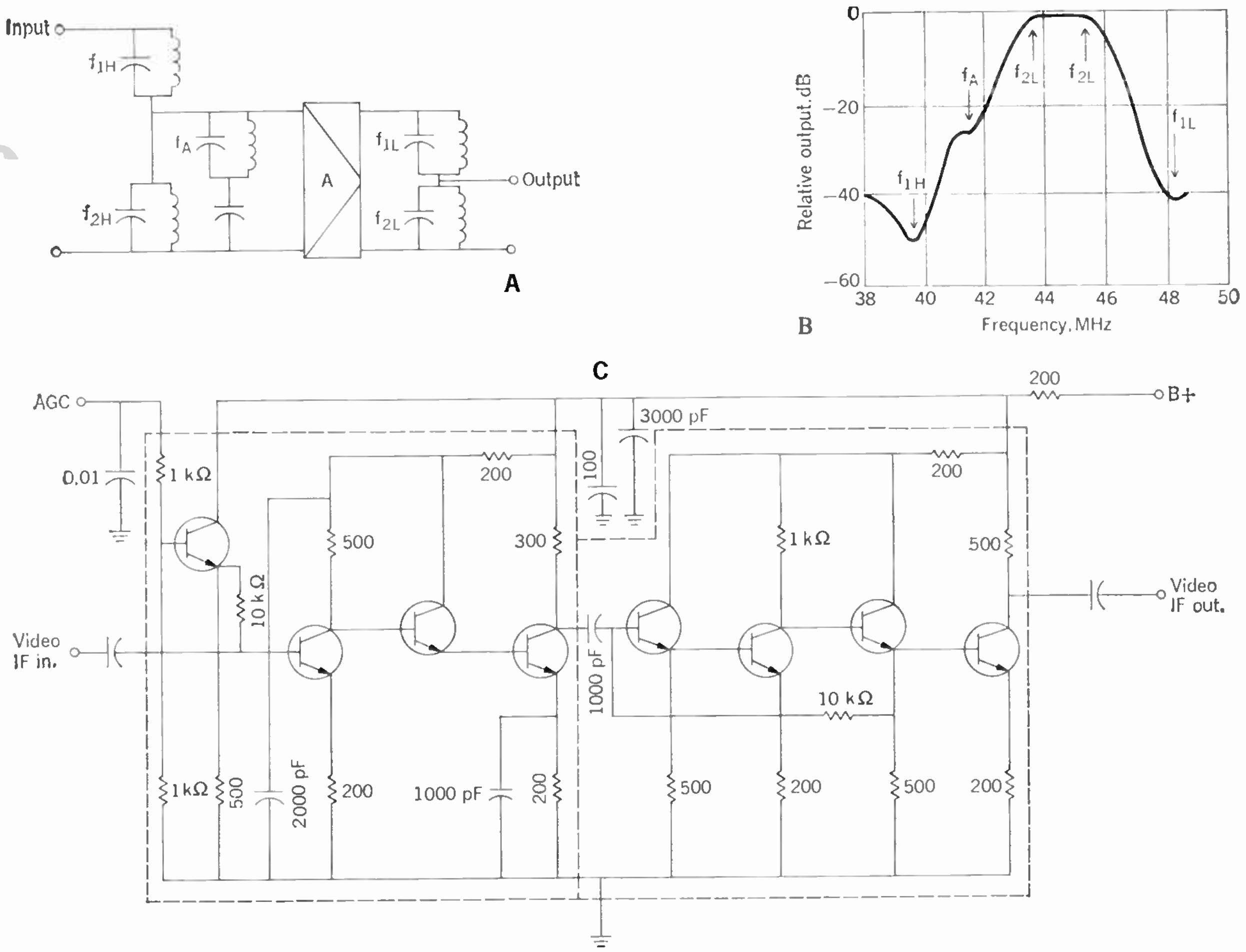
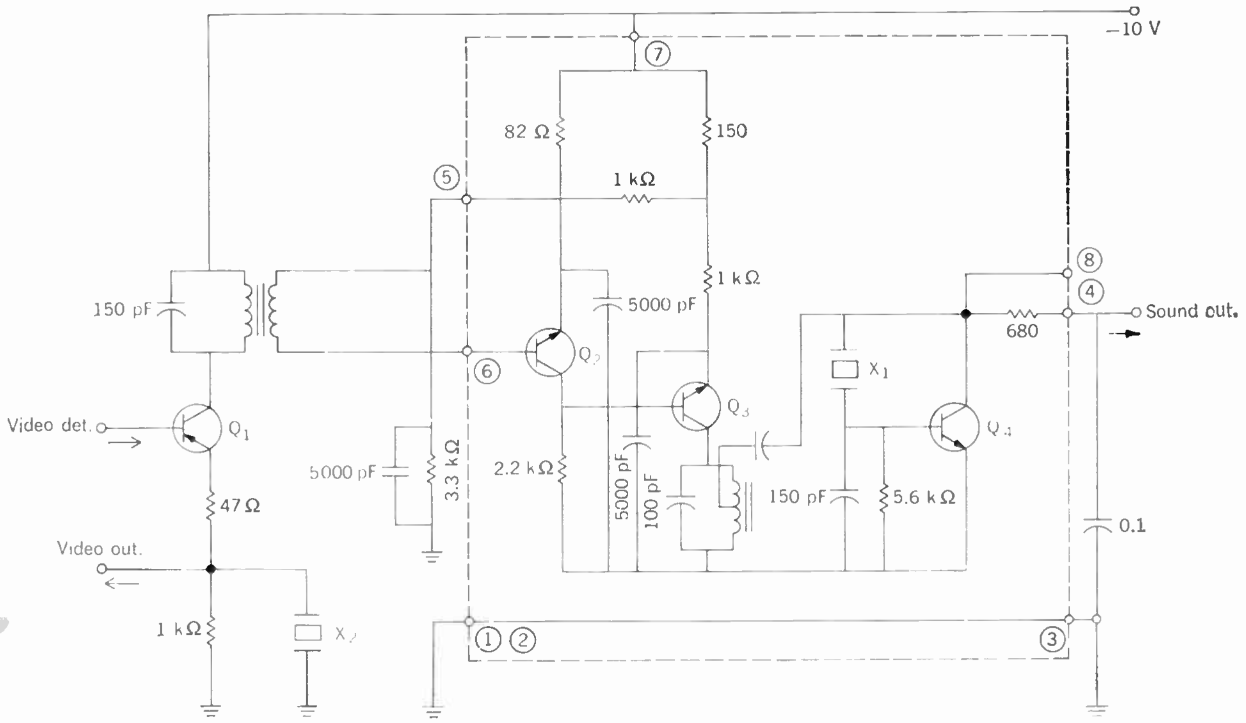


FIGURE 6. Video IF stage suitable for IC form. Desired frequency response can be obtained by inserting high- and low-pass filters at front and back of amplifier. A—Circuit configuration of video IF stage. B—Frequency response of video IF stage. C—Circuit diagram of the amplifiers used in the video IF stage. To get a high gain, two thin-film hybrid amplifiers are used in cascade. Resistance values are in ohms, capacitance values in microfarads, except where otherwise indicated.

FIGURE 7. Thick-film hybrid sound IF circuit. Two ceramic resonators, X_1 (a discriminator) and X_2 (a trap), are used in place of a discriminator transformer.



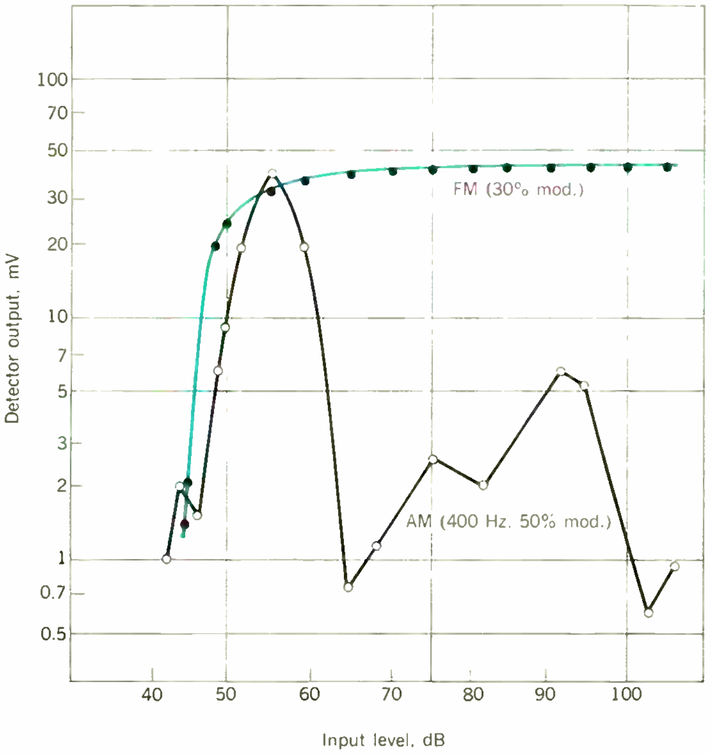
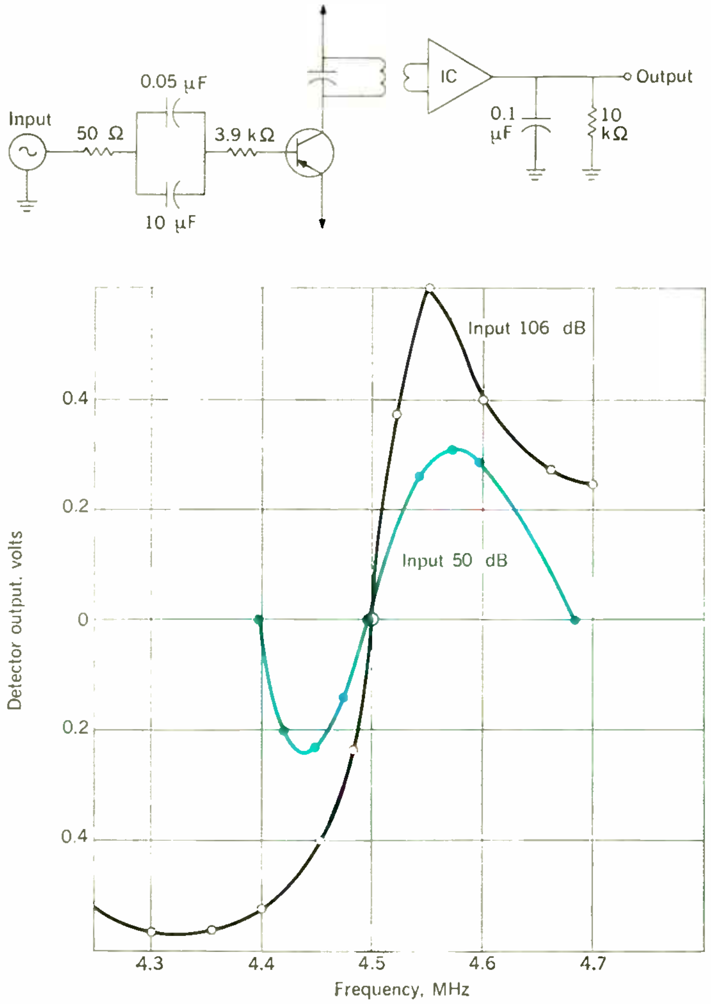


FIGURE 8. Detector output vs. sound input level.

FIGURE 9. Sound S-curve response for circuit shown.



The AM suppression ratio was 40 dB, and input limiting level was about 41 dB; see Figs. 8 and 9.

Circuits relating to synchronous signals

Highly integrated thin-film hybrid ICs perform three functions in one package: video amplification, sy chronous pulse separation, and AFC. The single, serie feedback stage Q_1 in Fig. 10 functions as a video amplifier. When the base is driven by the video detector, the emitter delivers the video driving signal and sound IF signal; the collector provides the positive video signal to be supplied to the synchronous pulse separator.

The synchronous pulse separator delivers separated synchronizing signals of about 11 volts peak-to-peak at the collector after the clamp circuit on the base provides the video signals at an equal amplitude.

The clamp circuit at the base of transistor Q_2 is formed equivalently by a 2-μF capacitor, a resistor load equivalent to a parallel resistor connection of 33 kΩ and 390 kΩ, and an equivalent diode formed between base and emitter of transistor Q_2. The clamped level is set by the built-in voltage of the diode.

The AFC circuit embodies a balanced-type sawtooth-wave AFC configuration. When a sawtooth wave of about 3 volts peak-to-peak is given as a reference waveform, the circuit delivers a horizontal oscillation control signal of about 1 volt. When the circuit is used in conjunction with a horizontal oscillation circuit of conventional discrete components, the horizontal pull-in range is approximately 250 Hz and the variation of the oscillation frequency due to temperature change is 0.015 percent per degree C.

Video and sound IF amplifiers and synchronous pulse separator circuits for color

The block provided in our color TV receiver satisfies the following conditions and has improved characteristics when compared with the previously described black-and-white circuit.

1. To prevent various disturbing signals due to such interference as cross modulation of sound signal and subcarrier wave in a composite signal, attenuation at the IF amplifier stage of a sound signal must have an amplitude of 50 dB. At the same time, a sound signal must keep a sufficient signal level.

2. Since a video detection signal is divided into a luminance signal, a chrominance signal, and a synchronous pulse signal, the delay and phase relations among these signals must be stable.

3. Since a horizontal deflection pulse is used for drawing out a synchronous burst signal, the phase relation between a horizontal synchronous pulse and the deflection pulse must be stabilized.

The features of the video IF amplifier and detector circuit are as follows:

1. The amplifier circuit is composed of a two-stage cascade connection.

2. A frequency-selection circuit is set up concentrically at the input and output of the video IF amplifier stage, and choke loads are used for interstage connections.

3. An inductor integrated by a multilayer boar method is used for the trap at the input side.

4. Since transistors are used in the video detector circuit, the problems of saturation and cross-modulation

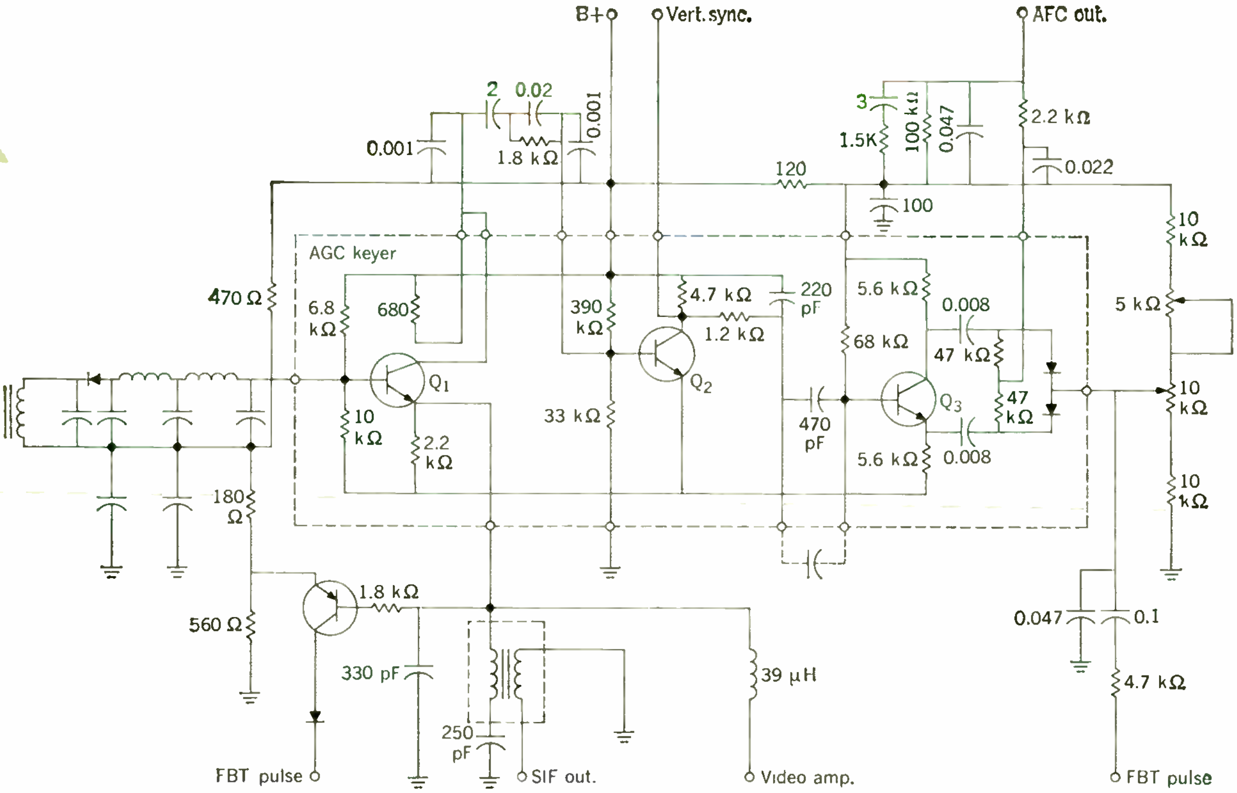


FIGURE 10. Thin-film hybrid IC for video IF stage amplifier, synchronous pulse separator, and AFC.

distortion due to high signal level diminish.

5. To provide video IF resonance inductors with the required *Q* (100 to 120), multilayer flat inductors are constructed using magnetic cores.

The overall characteristic of the video IF stage is shown in Fig. 11, with AGC voltage as a parameter. To obtain a stable output level and, at the same time, reduce noise to a minimum, AGC dc amplifiers follow a keyed AGC circuit. The AGC voltages are supplied separately to the video IF amplifier and to the tuner.

Sound intermediate frequency is amplified in three amplifier stages, and audio frequency is amplified after detection to keep a sufficient sound level. The synchronous pulse-separator circuit consists, basically, of a differential amplifier with a self-bias circuit; it is designed to obtain a normal separated output at a video signal level of 2 to 3 volts peak-to-peak. The AFC circuit is composed of a diode comparator circuit with compound components.

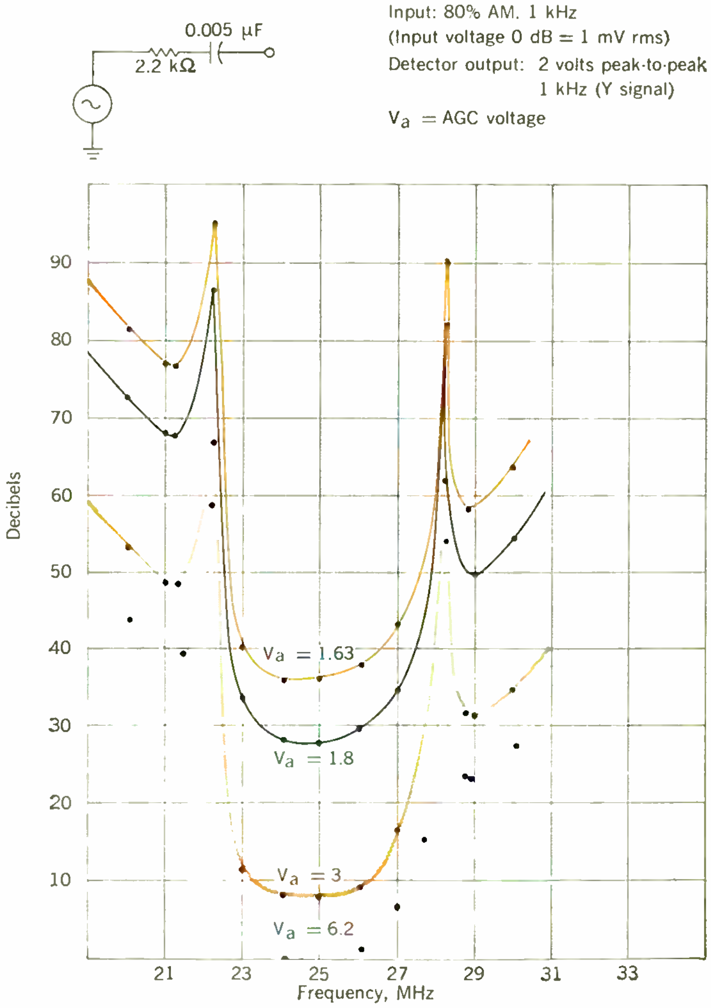
A thick-film hybrid IC blocking oscillator is employed in the vertical deflection circuit and a power transistor at the output. Electron tubes were used in the output and oscillator sections of the horizontal-deflection system.

Chroma bandpass amplifier, chroma demodulator, and matrix circuits

Special attention was paid to the application of ICs to the chromaticity circuits. Design goals included:

1. Little color distortion.
2. Circuit configuration to work ACC (automatic color control) satisfactorily and provide a sufficient output dynamic range.

FIGURE 11. Overall characteristics of video IF stage.



Input: 80% AM. 1 kHz
(Input voltage 0 dB = 1 mV rms)
Detector output: 2 volts peak-to-peak
1 kHz (Y signal)
V_a = AGC voltage

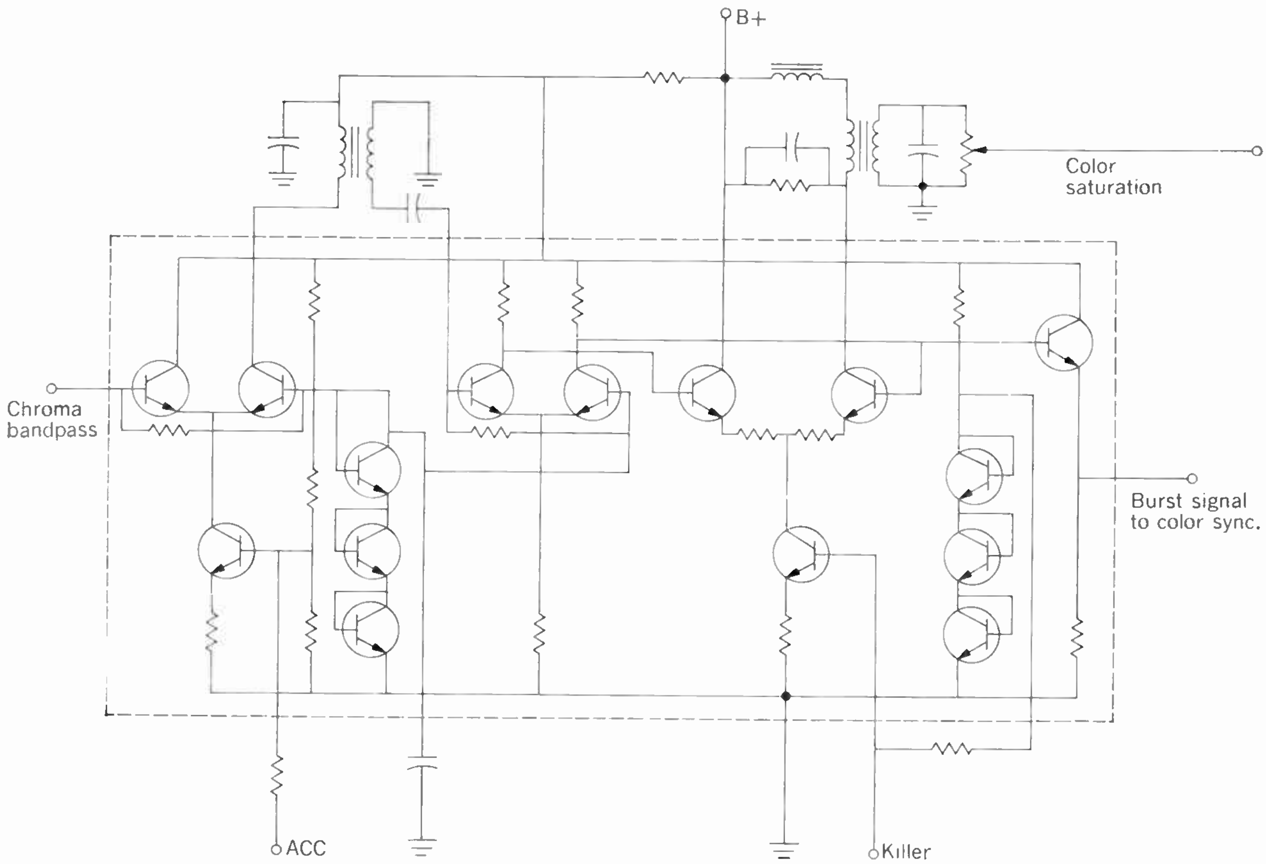


FIGURE 12. Typical monolithic IC for use in chroma bandpass amplifier.

3. Color amplifier output capable of supplying 120 to 150 volts peak-to-peak.

The chroma bandpass amplifier and chroma demodulator were made of monolithic ICs, and the matrix preamplifier and three matrix outputs were made of thick-film ICs.

Chroma bandpass amplifier circuit. In designing the chroma bandpass amplifier circuit, care was taken to include:

1. High-frequency compensation of 4.1 MHz.
2. Stable color-killer operation.
3. Satisfactory ACC operation.
4. Excellent input and output linearity and minimum color distortion.

Figures 12 and 13 show the circuit and its IC pattern, respectively. The circuit is composed of three differential-amplifier stages with ACC operation performed and the high-pass compensation characteristic obtained in the first stage, burst signal transfer to the color synchronous circuit performed in the second stage, and bandpass characteristic determined and color-killer operation performed in the third stage. Its electrical characteristics are as follows:

Range of color-carrier-wave input signal	0–50 mV rms
Input impedance	800 ohms at 3.6 MHz
Range of ACC voltage	20-dB variation, 0.8–1.2 V dc
Killer voltage	0 for black and white, 1.5 V dc for color

Chroma demodulator circuit. In the design of the chroma demodulator circuit, shown in Fig. 14, care was taken to ensure:

1. Satisfactory phase-detector operation at 3.58 MHz.
2. Good linearity to chroma bandpass output signal.
3. Well-balanced X-axis and Z-axis chroma demodulator output.

On designing the pattern inside the monolithic IC, the circuits are made to be bisymmetrical to get the balance of X-axis and Z-axis signals on the chip. The output (gain) difference between the X-axis and Z-axis demodulator was kept to less than ±5 percent.

The electrical characteristics are as follows:

Range of color-carrier-wave input signal	0–50 mV rms
X- and Z-axis subcarrier wave inputs	100 mV rms

Matrix preamplifier circuit. The chroma demodulator output levels are amplified by one stage before proceeding to the last-stage matrix output, so that they have sufficient levels as E_x and E_z signals. This circuit, called the matrix preamplifier, is composed of X-axis and Z-axis demodulated-output amplifier circuits.

The following points were taken into consideration in the design:

1. Frequency characteristics.
2. Gain (voltage gain of 26 dB).
3. Sufficiently balanced X-axis and Z-axis signals.
4. Output voltage of 15–20 volts peak-to-peak.

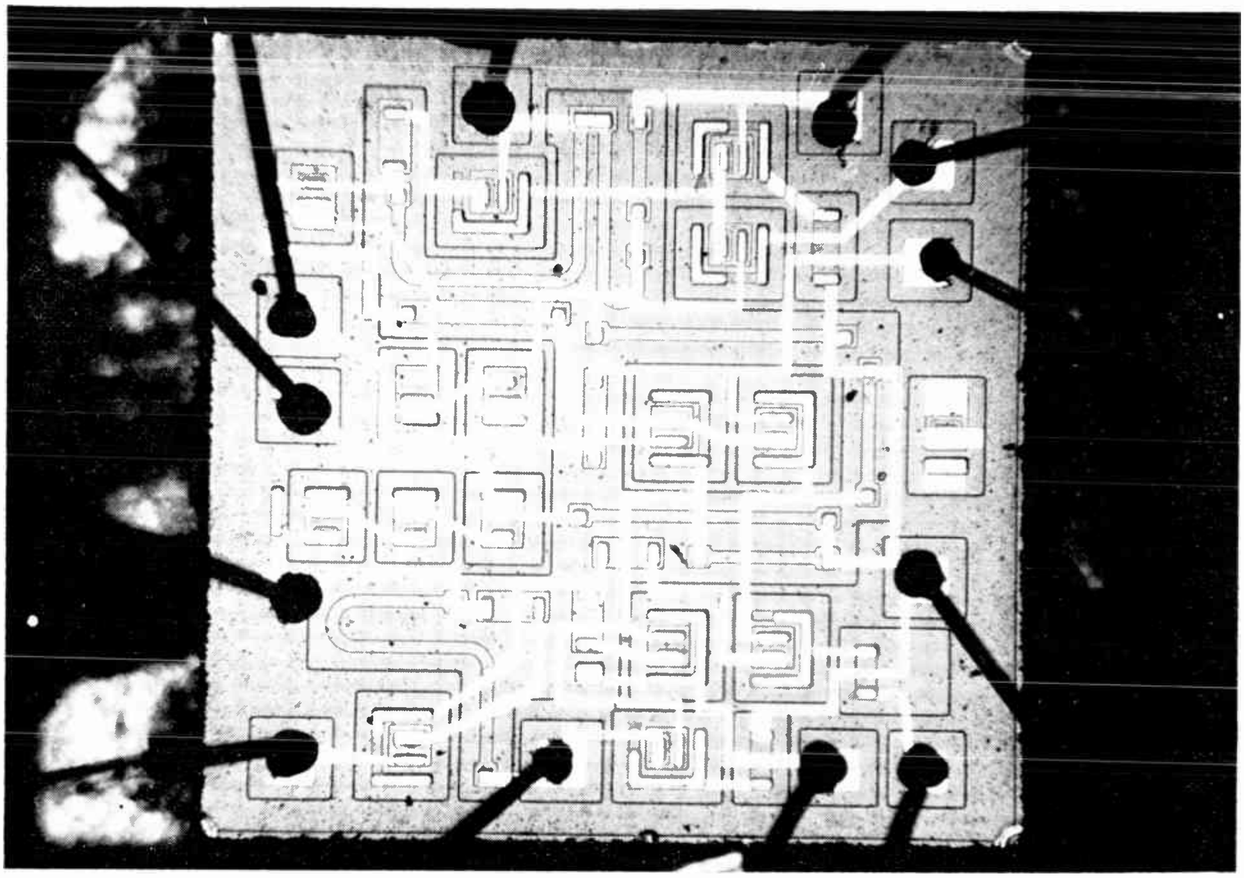


FIGURE 13. IC pattern inside the chroma bandpass amplifier. (Courtesy Hayakawa Electric Co.)

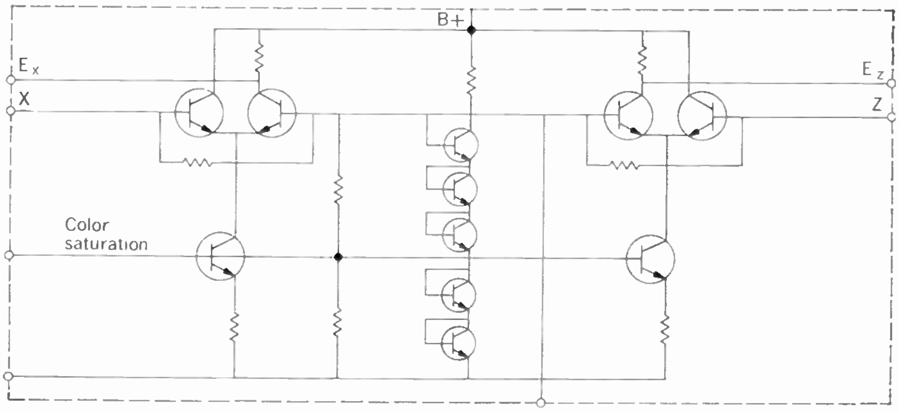


FIGURE 14. Monolithic IC for chroma demodulator.

By selecting transistors properly, the balance between X-axis and Z-axis signals was kept within ±5 percent.

Matrix output circuits. The matrix output circuits were composed of electron tubes and thick-film ICs. It was not possible to integrate this stage, since it treats high-voltage signals (120–150 volts peak-to-peak). The problem encountered in the design of thick-film ICs was that the heat generated from plate-load resistors raised the temperature of the IC to such a degree that thermal stresses changed the electrical characteristics of the built-in capacitors.

Burst amplifiers and the killer circuit

A monolithic burst amplifier extracts a burst signal contained in a composite video signal, and the output of the amplifier is used to control the APC (automatic phase control) circuit, the killer circuit, and the ACC circuit. Therefore, an IC used in the burst amplifier must have a high gain and must be placed under the influence of AGC.

The electrical characteristics are as follows:

Peak-to-peak input and output levels	input, 0.3 V; output, 10 V
Voltage gain	more than 30 dB
Impedance	$Z_{in} = 400\ \Omega$; $Z_{out} = 100\ k\Omega$

A color-killer circuit is composed of a thin-film hybrid IC in which tantalum thin-film resistors, glass thin-film capacitors, chip transistors, and diodes are bonded together. The killer output supplies the chroma bandpass with 0 to −0.2 volt dc for the color signal and −5 volts dc for the black-and-white signal.

Figure 15 shows this circuit in block-diagram form. The input–output characteristics of the burst amplifier with a preamplifier at the input are given in Fig. 16. The ICs used for this circuit block are shown in Figs. 17 and 18.

Color-reference oscillator circuit

The color-reference block consists of the oscillator and phase detector that demodulate and separate two color-difference signals. It is desirable to keep a phase error within ±5° to avoid a shift of hue visible to the eye. The circuit is shown in Fig. 19. The following points were taken into account:

1. Stabilization of the oscillator.
2. Reduced number of inductors and capacitors.
3. Reactance circuit with excellent linearity and high sensitivity in frequency control.
4. Satisfactory relation in levels between the burst amplifier and demodulator circuits.

The phase-detector circuit shown in Fig. 20 is composed of a diode balanced phase-detector circuit, an integrator, and an emitter follower. In this circuit, the transistor and diodes and capacitors are of the chip

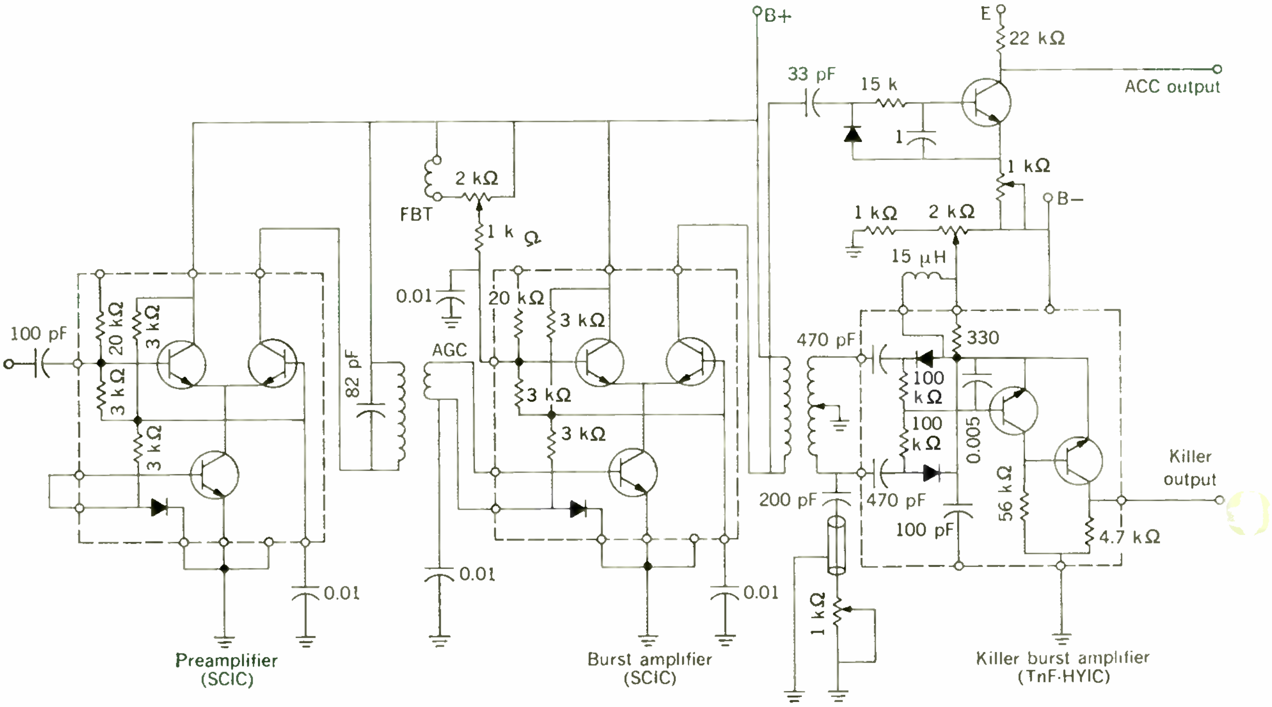


FIGURE 15. Circuit diagram of burst amplifier, killer, and automatic color control.

FIGURE 16. Killer and burst output characteristics of burst amplifier with a preamplifier at the input.

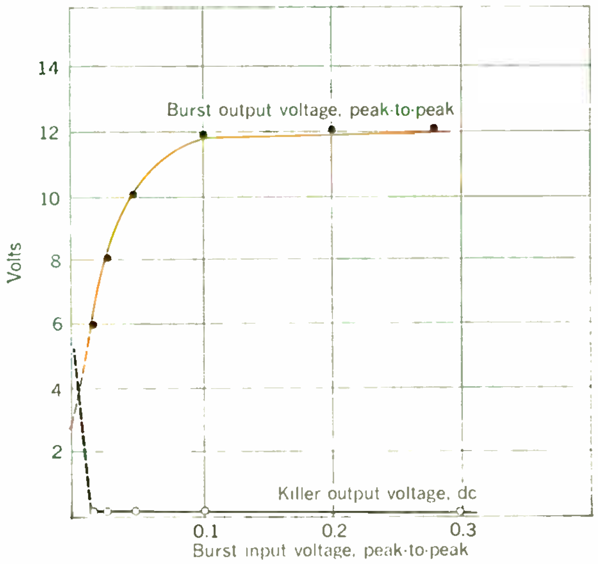
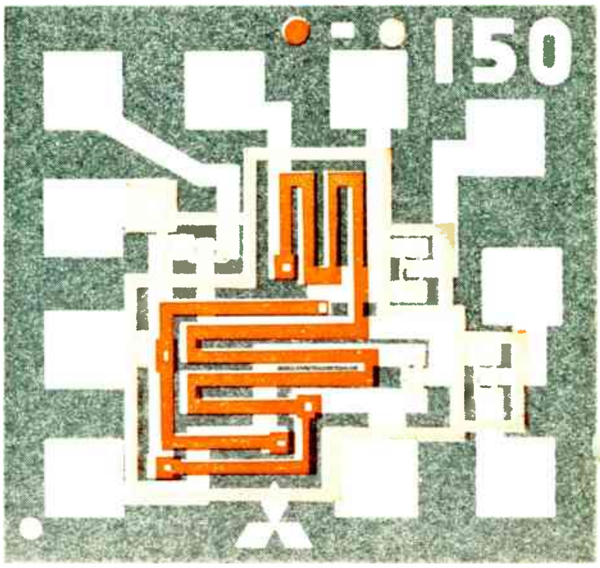


FIGURE 17. Monolithic IC used for burst amplifier and preamplifier. (Courtesy Mitsubishi Electric Co.)



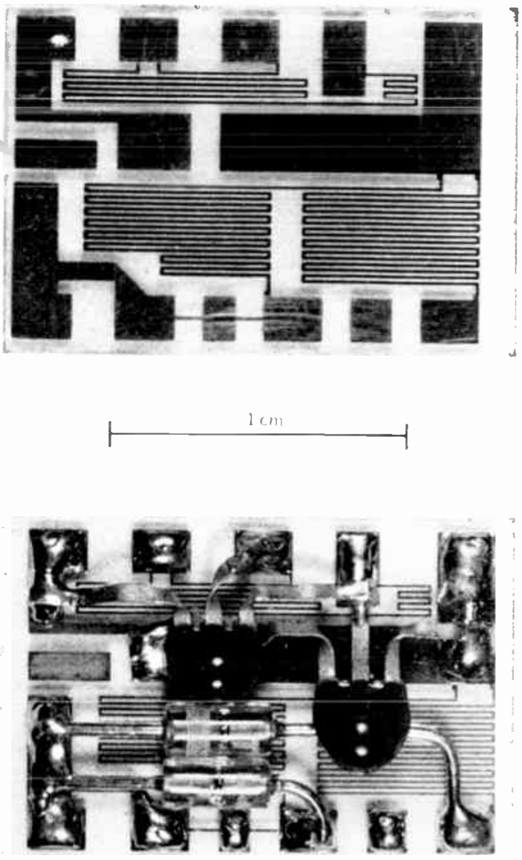


FIGURE 18. Thin-film hybrid IC used for killer burst and amplifier. (Courtesy Mitsubishi Electric Co. and Sodenshya Co.)

type, the two 100-kΩ resistors are of thick-film type (and capable of being adjusted later for accuracy), and all other resistors are of the thin-film type.

A monolithic IC is used to provide the active elements of a 3.58-MHz oscillator and an amplifier. A thick-film hybrid IC is used in the reactance transistor circuit and to provide some resistors and capacitors that could not be included in the monolithic IC. Reactance transistor Tr_2 uses a hyperabrupt variable-capacitance diode D_2 to obtain high sensitivity of frequency control. The oscillation frequency control characteristic is shown in Fig. 21.

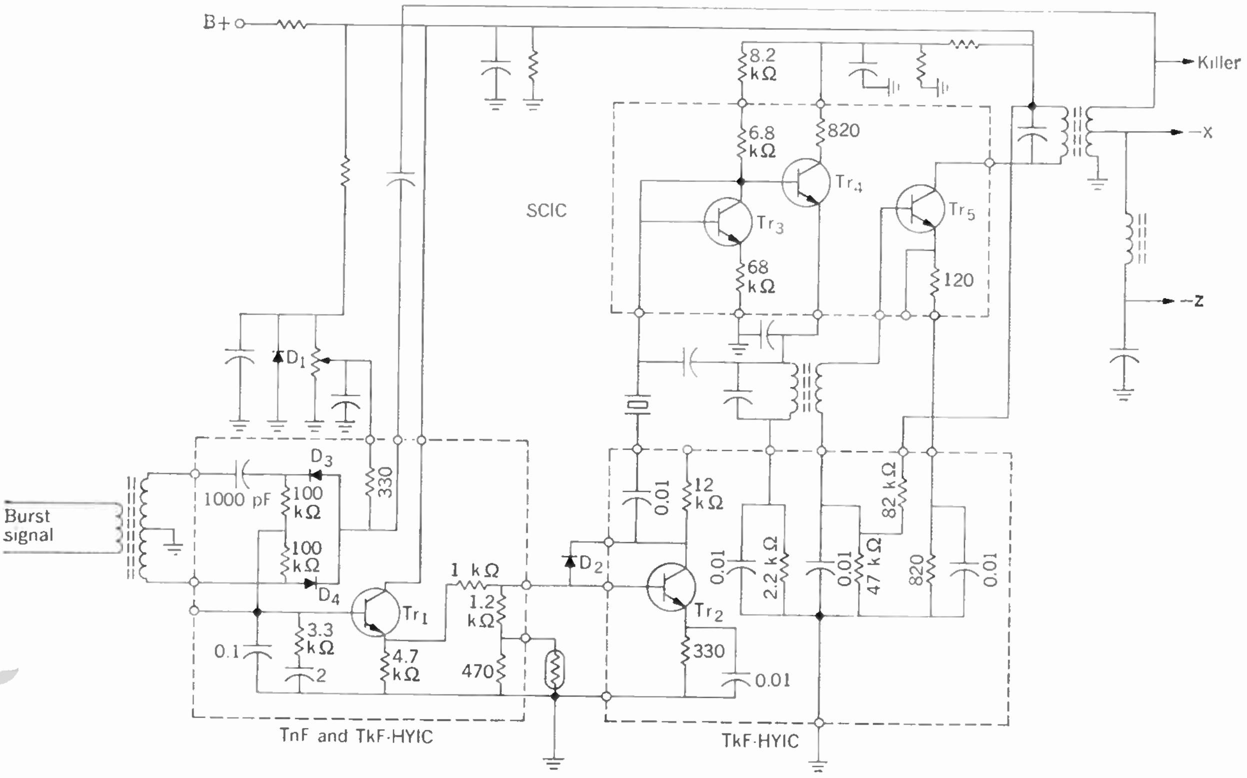
The output of the 3.58-MHz oscillator is applied to an X and Z chroma demodulation system; the phase-shift circuits for giving subcarrier waves of $-X$° and $-Z$° are composed of discrete components. By this means, the number of inductors was reduced to three, one each for the X- and Z-axes and one for the oscillator.

Conclusion

At present, black-and-white and color TV receivers incorporating integrated circuits have appeared on the market as commercial products both in Japan and in the United States. The prototype receiver described herein indicates the possibility of linear ICs being applied to almost all except high-power circuits—the performance being equal to that of conventional receivers incorporating only discrete components.

Under development now is the all-IC color TV receiver, which includes the circuits that are subject to high voltages and high power consumption—for example, video-output, horizontal-output, and vertical-output circuits, high-voltage rectifiers, and power sources. It appears that such a design can be realized

FIGURE 19. Circuit diagram of color reference oscillator.



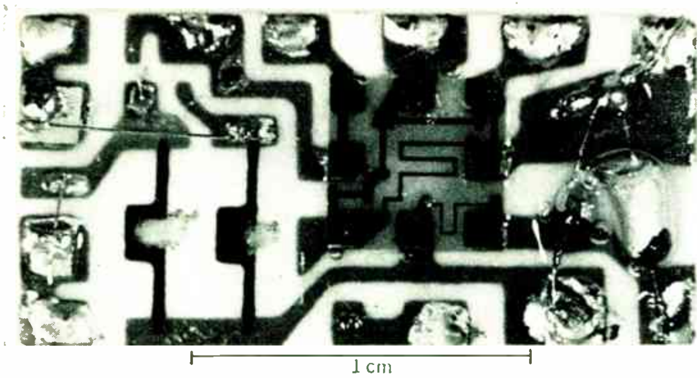
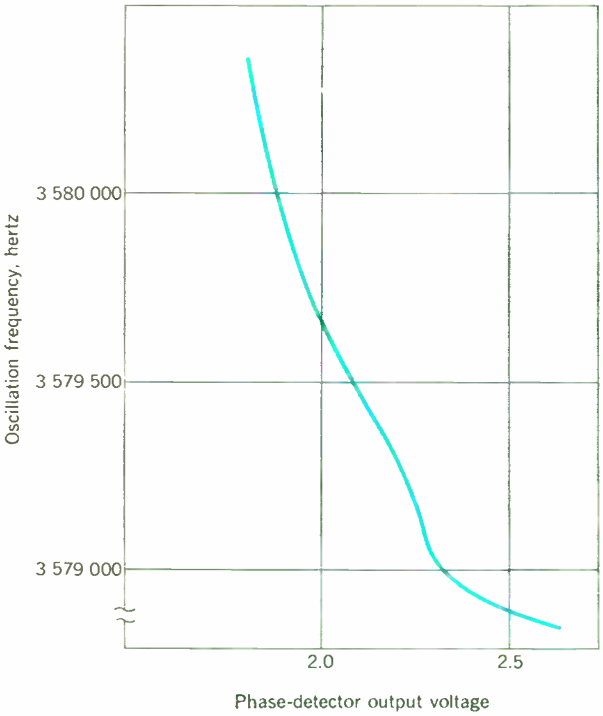


FIGURE 20. Thin-film and thick-film hybrid integrated phase-detector circuit. (Courtesy Nichicon Capacitor Co.)

FIGURE 21. Oscillation frequency control characteristic of color reference oscillator circuit.



at low cost. As the next development in TV receivers, we have in view a solid-state display device.

This article is an expanded version of a paper presented by the same authors at the International Solid State Circuits Conference, held in Philadelphia, Pa., Feb. 14–16, 1968. The editor apologizes for the abnormal delay in publication.

The cooperative research and development work described in this article was carried out by a joint group consisting of the Kansai Electronic Industry Development Center, acting as the main body, along with manufacturers, universities, and institutes in the Kansai district, with the financial aid of the Ministry of the International Trade and Industry of Japan. The working members of this joint research group are: Osaka Onkyo Co., Ltd.; Hayakawa Electric Co., Ltd.; Matsushita Electric Ind. Co., Ltd.; Mitsubishi Electric Co., Ltd.; Sanyo Electric Co., Ltd.; Elna-Fox Electronics Co., Ltd.; Hoshi Electric Co., Ltd.; Kobe Ind. Corp.; Matsuo Electric Co., Ltd.; Murata Mfg. Co., Ltd.; Nichicon Capacitor Co., Ltd.; Sodensha Co.; and Osaka University.

The authors thank Profs. K. Miyawaki and J. Nakai of Osaka University, who have served as leaders of the working groups, and acknowledge the assistance of N. Okabe in the preparation of the manuscript.

REFERENCES

1. Avins, J., "It's a television first—receivers with IC," *Electronics*, vol. 39, pp. 137–142, Mar. 21, 1966.

2. Avins, J., "Integrated circuits in television receivers," *IEEE Trans. Broadcast and Television Receivers*, vol. BTR-12, pp. 70–74, July 1966.

3. Whiteneir, P. J., Jr., "The analysis and design of semiconductor and integrated circuit IF amplifiers for television," *IEEE Trans. Broadcast and Television Receivers*, vol. BTR-12, pp. 75–86, July 1966.

4. Bray, D., "Semiconductor video amplifiers for monochrome and color TV receivers," *IEEE Trans. Broadcast and Television Receivers*, vol. BTR-12, pp. 20–27, Nov. 1966.

5. Blaser, L., and Bray, D., "Color TV processing using integrated circuits," *IEEE Trans. Broadcast and Television Receivers*, vol. BTR-12, pp. 54–60, Nov. 1966.

6. Welling, B., "An IC color TV video IF facilitates alignment and improves AGC," *IEEE Trans. Broadcast and Television Receivers*, vol. BTR-13, pp. 24–33, July 1967.

7. Weber, R. L., and Prabhakar, J. C., "A thick-film television video IF amplifier using compatible components," *IEEE Trans. Broadcast and Television Receivers*, vol. BTR-13, pp. 7–12, Nov. 1967.

8. Rongen, J. J., "An HF monolithic integrated circuit for TV-sound IF amplification and its stability criterion," *Dig. 1968 Solid-State Circuits Conf.*, pp. 140–141.

9. Kossakowsky, L. L., "TV sound IF amplifier monolithic IC," *Dig. 1968 Solid-State Circuits Conf.*, pp. 142–143.

Eizi Sugata (SM) received the B.S. degree in electrical engineering from Osaka Institute of Technology in 1932 and D.Sc. degree in engineering from Osaka Imperial University (later renamed Osaka University) in 1944. He began his professional career in 1932 as a lecturer at Osaka Institute of Technology, where he became assistant professor in 1937. Since 1944 he has been a professor at Osaka University. He designed and fabricated the first electron microscope in Japan in 1939. Since that time he has worked continuously in the fields of electron optics and electron-beam applications. Using the electron-beam evaporation method, he designed an developed a continuous-production plant for thin-film integrated circuits. At present he is pursuing his research in the IC industry and also in surface physics, employing the field-emission microscope, field-ion microscope, and low-energy electron diffraction. Prof. Sugata is a member of the International Society of Hybrid Microelectronics, as well as the Institute of Electronics and Communications Engineers and many other societies in Japan.

Toshihiko Namekawa received the B Sc. degree in electrical communication engineering in 1945 and the D.Sc. degree in 1962, both from Osaka University. From 1950 to 1955 he was a member of the staff of the Mainichi Newspaper Company, Osaka, where he was engaged in the development of a communication system. From 1955 to 1960, as associate professor in the Electrical Department at Kobe University, he worked on problems concerning noise in vacuum tubes. He was appointed associate professor in 1961, and professor in 1968, of electrical communication engineering at Osaka University, where he teaches courses in electronic circuits, communication theory, noise theory, and active circuits. He is now involved in active-circuit analysis and synthesis, applications of integrated circuits, and signal-processin systems. Prof. Namekawa i member of the Institute of Electronics and Communications Engineers.

Magnetic traveling fields for metallurgical processes

Since the invention of the induction stirrer in Sweden in the early 1930s, significant progress has been made in the application of traveling magnetic fields to processes in the metals industry

Yngve Sundberg ASEA

Alternating magnetic fields lead to electrodynamic forces, which can be used industrially for stirring melts. These magnetic traveling fields, produced by straight or cylindrical stirrers, have become a well-established tool for the metallurgist. Since the size of melts is constantly growing, the stirrers are becoming increasingly valuable for achieving homogenization, temperature equalization, rapid and complete chemical reactions, and effective degassing.

Maxwell's second law states that alternating magnetic fields can produce electric currents through molten or solid metal. These currents lead to heating, which is utilized in induction furnaces, and to electrodynamic forces, which can be used for stirring a melt. This article will be devoted exclusively to the various aspects of stirring.

It is usual to distinguish between repulsive and motory stirring. Repulsive stirring is obtained with a stationary magnetic alternating field, where the stirring forces act *perpendicularly* to the surface of the furnace refractory, as in the case of a normal induction melting furnace, for example.

Since the flow must run *along* the surface of the refractory, it is of greater value if the stirring forces act tangentially. This is the basis of motory stirring, which is achieved with the aid of a magnetic traveling field.[1] The motory-stirring principle is used in induction stirrers, invented in the early 1930s by Dr. Ludwig Dreyfus, who was employed at ASEA in Sweden. He discovered that it was possible to develop sufficient electrodynamic forces in the melt by means of a magnetic traveling field to achieve satisfactory stirring. He also found that it would be possible to generate the traveling field with a coil located outside the steel shell of the furnace or ladle within which the melt is located, if the shell were made of nonmagnetic steel and a sufficiently low frequency were selected.

The first experiments, performed in 1933, were being carried out on a cylindrical device. At an early stage, Dr. Dreyfus linked together the stirring with induction heating of melts in dual-frequency furnaces, where in a common coil he could feed both medium-frequency alternating current for heating and low-frequency current for stirring. The first plant with induction stirring to be delivered—to Sandvikens Jernverk, Sweden, in 1934—employed this combination.

During these first investigations it was found, as expected, that the metallurgical processes took place much more rapidly than they would have without stirring. The shape of the melt in a medium-frequency furnace, however, is not particularly favorable for slag reactions, since the surface is small in relation to the volume. A modification was introduced in the form of a conical furnace, but this did not prove to be very satisfactory either.

The best electrically heated furnaces for steel refining are arc furnaces, in which the bath surface is large and the depth small. It was therefore natural that induction stirring would be of greatest value for such furnaces. The first experiments were performed at Surahammar in 1939. In this case the coil was located inside the bottom shell of the furnace. With this location the temperature was too high, however, and the coil was destroyed.

In 1947 the first stirrers of the type now used for arc furnaces were supplied. Since then, over 100 induction-stirrer installations have been supplied to many different countries.

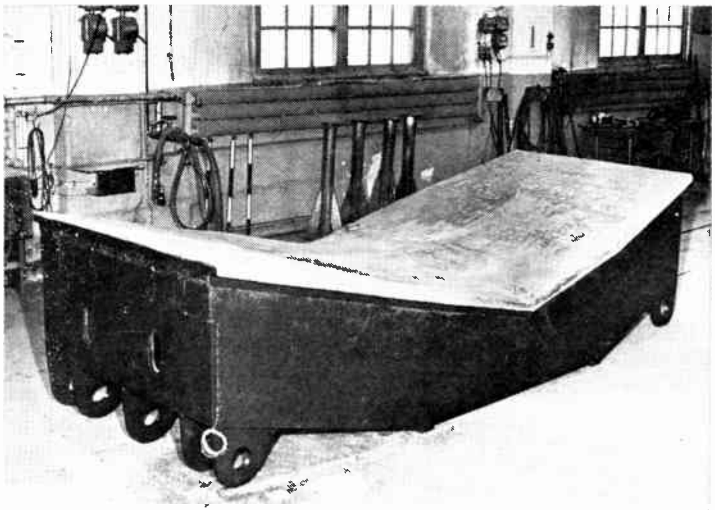
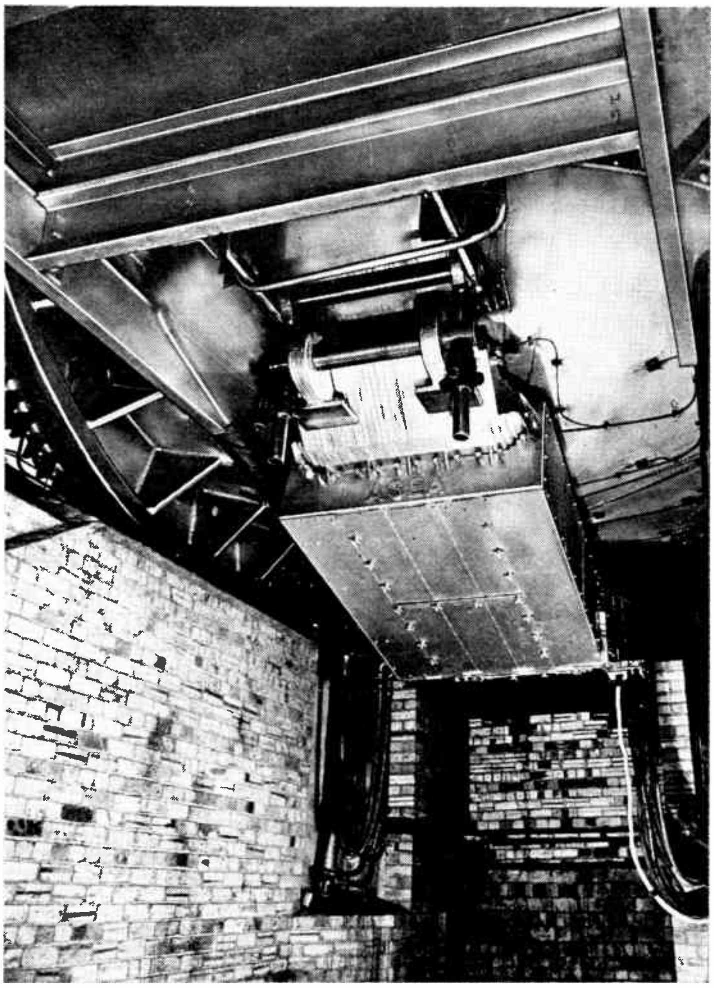


FIGURE 1. Induction stirrer for arc furnace having an inner shell diameter of 500 cm; 350-kVA, two-phase, 220-V/phase, 0.75-Hz feed.

FIGURE 2. Induction stirrer mounted beneath a 60-tonne arc furnace with an inner shell diameter of 580 cm; 350-kVA, two-phase, 130-V/phase, 0.6-Hz feed.



The cylindrical stirrer lay more or less dormant until the end of the 1950s, when the vacuum degassing of melts started to attract attention. It proved to be the solution to the problem of exposing all parts of the melt to vacuum.

Straight stirrers for arc furnaces

Design and location. Figure 1 shows the design of the stirrer; its location beneath the bottom shell of the furnace can be seen in Fig. 2. It consists of a laminated core with a two-phase winding, resembling most closely a section of the stator of an induction motor. Since the frequency used is low (0.5–2 Hz), the sheet used for the core can be considerably thicker than is normally the case. The winding must carry a large current; for this reason it consists of copper tubes, through which cooling water flows. In order for the magnetic field of the stirrer to penetrate through the bottom shell of the furnace, the shell must be constructed of nonmagnetic steel.

The length of the stirrer should be approximately equal to the diameter of the furnace. The distance from the upper surface of the stirrer to the bottom surface of the melt should be as small as possible, and roughly constant over the entire length of the stirrer, so as to ensure that there will be an effective stirring of the entire melt.

Function. The normal connection of the stirrer windings is shown at the bottom of Fig. 3. The windings are connected to two phases, separated from one another by 90°. The way in which the traveling field is produced is illustrated at the top of the figure. The distribution of the field strength over the surface of the stirrer is illustrated at four times (A, B, C, and D). At time A phase 1 has a current maximum, while the current in phase 2 is zero. Note the rectangle pointing upward to the left in the diagram. At time B the current phasors have been turned through 90° in the counterclockwise direction. Phase 2 now has a current maximum, and the phase 1 current is zero. The rectangle to the left in the diagram has now moved one step to the right.

At time C, when the current phasors have been turned through a further 90°, phase 1 once again has a current maximum, but in the negative direction, and the current in phase 2 is zero. The rectangle pointing upward has now moved to the middle field of the stirrer. Finally, at time D, phase 2 has a current maximum in the negative direction and the current in phase 1 is zero. By this time the rectangle pointing upward in the diagram has moved still another step to the right.

After yet another turning of 90° of the current phasors, the rectangle is situated farthest to the right. One cycle has now been completed, and the traveling field has moved over a distance equal to the length of the stirrer. If the frequency is denoted ν and the length of the stirrer 2τ (where τ = pole pitch), the time for one cycle is $1/\nu$ and the speed of the traveling field (the "synchronous speed") is therefore

$$v_s = \nu \cdot 2\tau \qquad (1)$$

The traveling field pulls along with it the melt in the same way that the rotating field of an induction motor causes the rotor to rotate. The speed of the melt can never be as high as the synchronous speed, since the pull will then be zero.

The magnitude of the pull is a very complicated function of the field strength, frequency, angle between the two phase currents, speed of the melt, resistivity of the melt, pole pitch, depth of the melt, and vertical distance between the upper surface of the stirrer and the bottom surface of the melt. Figure 4 shows the horizontal volume force f at a certain instant and at a certain point in the melt. It is obtained from the relation

$$f = s_1 B_{2v} \qquad (2)$$

where s_1 is the density of the induced current generated by phase and acting perpendicular to the plane of the paper, and B_{2v} is the vertical component of the flux density caused by the current in phase 2.

The volume force is at its largest at the bottom of the melt and decreases in the upward direction. It can be integrated across the depth of the melt, and a force $\bar{f}$ related to the surface unit of the bottom surface of the melt is then obtained. This "specific stirring force" is of the order of 500 N/m^2 for a field strength on the upper surface of the stirrer of 100 kA/m and a suitably selected frequency. It is assumed here that the thickness of the refractory in the bottom of the furnace is normal. The force varies with the vertical distance δ between the stirrer and the melt approximately according to

$$\bar{f} = \bar{f}_0 \, e^{-2\pi\delta/t} \qquad (3)$$

The stirring force therefore rapidly decreases with increasing refractory thickness. In addition, the stirring force is proportional to the sine of the angle between the current vectors of the two phases. It is therefore important that this angle be maintained equal to 90°. (The traveling field can be made to change direction if one phase is reversed.)

Bath motion. Figure 5 shows the appearance of the bath motion. Since the stirrer extends over the diameter of the furnace and, furthermore, is shaped so

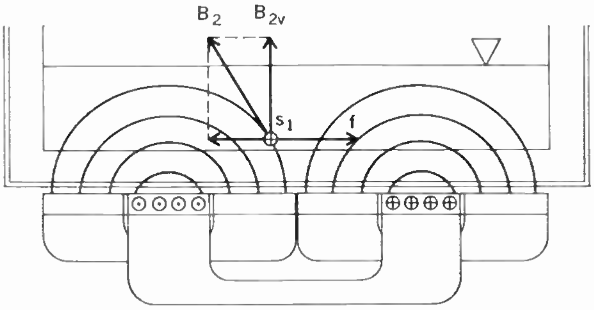


FIGURE 4. Distribution of the field strength and direction of the force with a straight stirrer.

FIGURE 3. Distribution of the field strength at different times with a straight stirrer.

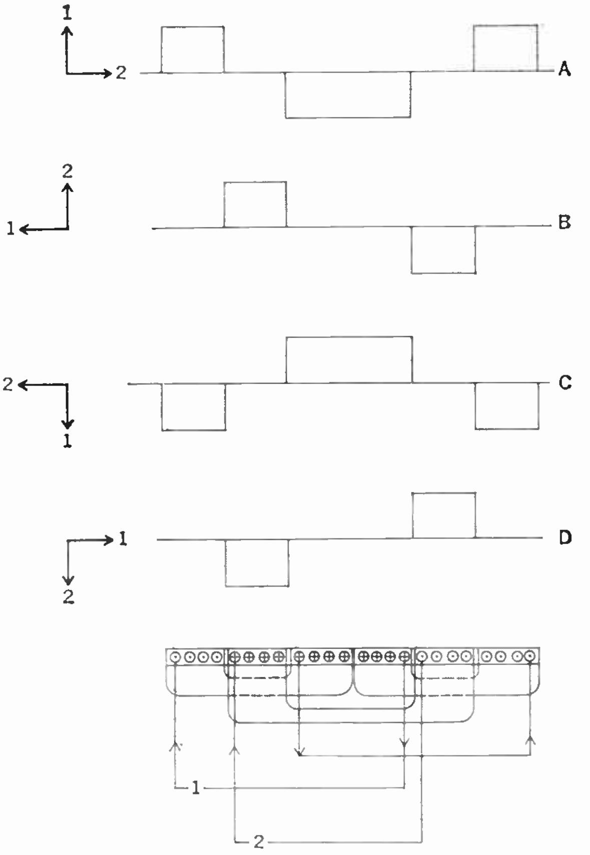
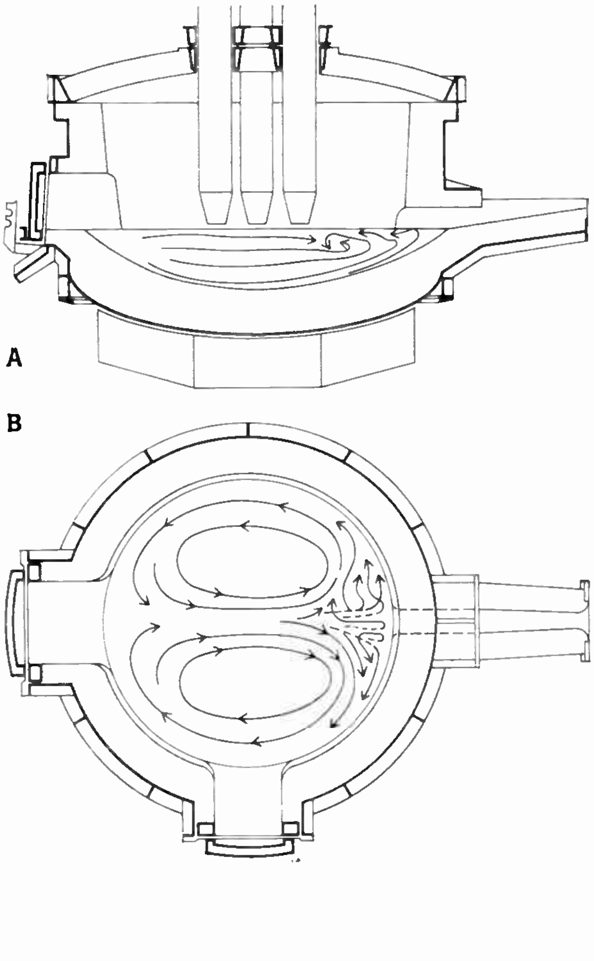


FIGURE 5. Arc furnace with an induction stirrer. A—Bath movement. B—Flow pattern on the surface.



that it follows the bottom profile of the melt, stirring forces are obtained across the entire furnace diameter and all the melt is forced to take part in the movement.

If the vertical distance between melt and stirrer increases at the ends of the stirrer, there is a risk that only local stirring will be obtained in the middle of the melt. This has in fact occurred in practice. When one phase is reversed, the flow will take place in the opposite direction.

Flow rate. It is difficult to determine a theoretical relationship between flow rate and stirring force, since the flow is turbulent. A calculation based on the assumption of laminar flow yields too favorable results. A pessimistic way of looking at the problem is to assume that the speed of the melt is zero at one end of the stirrer and is accelerated with the stirring force at the other end. This gives the acceleration

$$a = \frac{\bar{f}}{\rho\Delta} \qquad (4)$$

where ρ is the density of the melt and Δ the depth of the melt, and the flow rate after the passage of the length, 2τ, of the stirrer is

$$v = \sqrt{a \cdot 4\tau} = \sqrt{\frac{f \cdot 4\tau}{\rho\Delta}}$$

If, for example, $f = 500$ N/m², $\tau = 2.5$ meters, $\rho = 6900$ kg/m³, and $\Delta = 0.8$ meter, then the following will be obtained:

$$v = \sqrt{\frac{500 \times 4 \times 2.5}{6900 \times 0.8}} = 0.95 \text{ m/s} \qquad (5)$$

As already mentioned, the flow rate can never be as large as the synchronous speed, which is determined according to Eq. (1). If, for example, $\nu = 0.8$ Hz and $\tau = 2.5$ meters, then

$$v_s = 0.8 \times 2 \times 2.5 = 4.0 \text{ m/s}$$

A flow rate on the surface of the melt of 0.5–1 m/s exactly above the stirrer is normally obtained in practice. This is fully adequate for the desired metallurgical effects to be obtained.

Electric equipment. The low-frequency current can be produced either with rotating machines, as shown in Fig. 6, or by means of thyristors, as shown in Fig. 7. In both cases only the equipment for one phase is illustrated. A similar set of equipment is required for the other phase.

Rotary converters with cyclically excited dc generators driven by induction motors are used for the first method. The excitation is accomplished by means of a low-frequency device.

A double-converter connection is most suitable for the thyristor method. A single-converter connection, which will be considerably cheaper, will cause three times the coil losses for the same stirring force in the melt.

The electric equipment must satisfy the following requirements:

1. The phase angle between the phase currents must be capable of being adjusted accurately and permanently to 90° to obtain the best stirring force.
2. It must be possible to adjust the frequency within certain limits to obtain the best stirring force.
3. It must be possible to control the voltage on the two phases from the maximum value down to about 50 percent or lower to obtain suitable stirring for different types of operation.

FIGURE 6. Motor–generator set with control equipment for feeding a coil section. 1: Low-frequency device. 2: Transductor amplifier. 3: Exciter with two field windings. 4: Low-frequency generator. 5: Motor. 6: Coil section.

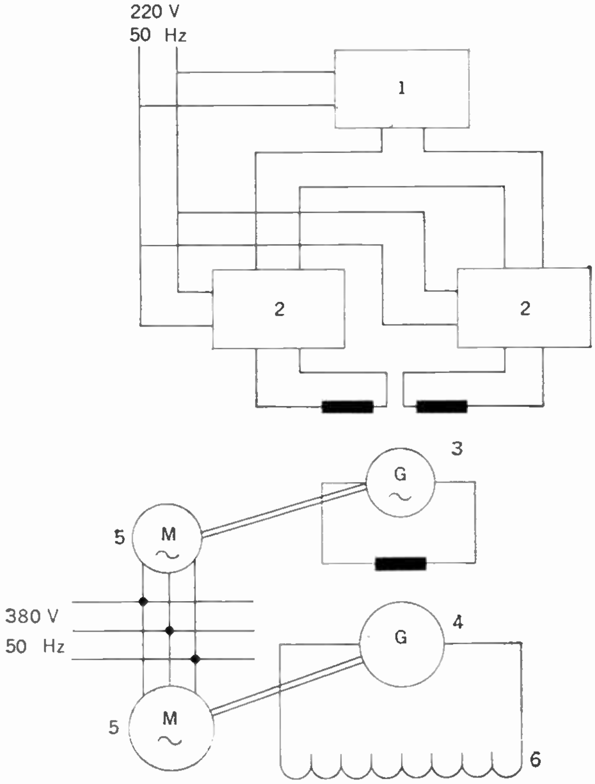
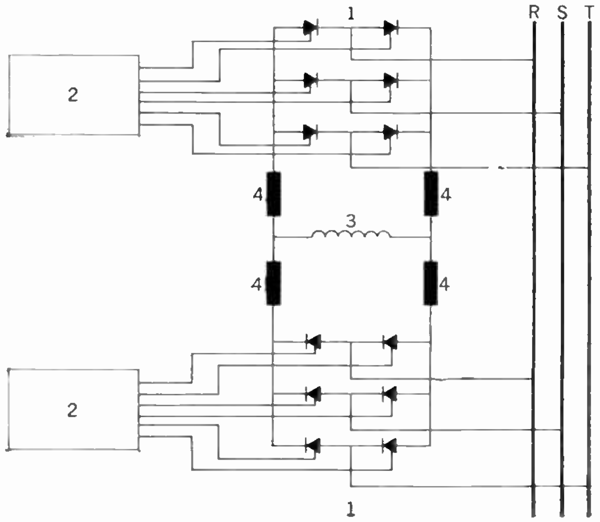


FIGURE 7. Thyristor converter in double-converter connection for feeding a coil section. 1: Thyristor assembly. 2: Trigger equipment. 3: Coil section. 4: Circulating-current reactor.



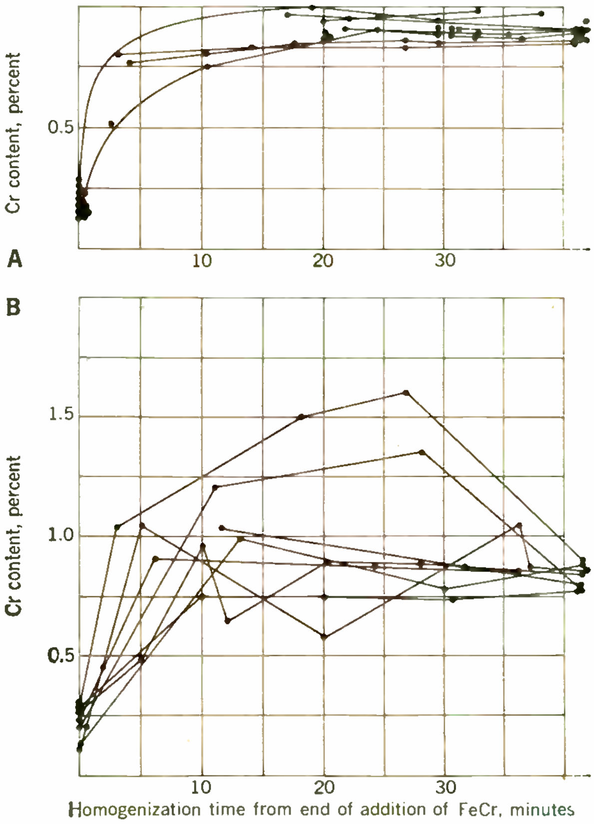
Metallurgical considerations[2]

Homogenization. The induction stirrer causes the melt to move along the entire bottom of the furnace, and yields very effective mixing when the molten metal is forced to reverse its flow near the wall and return in the opposite direction; see Fig. 5. The result is improved homogeneity, which permits a better overall control of the bath composition, a more accurate calculation of the alloy additions, and a more precise determination of the temperature, as will be illustrated by a few examples.

In a French steelworks it has been found that a chromium addition of about 0.9 percent to a 30-tonne melt is well distributed within 20 minutes. Figure 8 shows the chromium content of samples from about 15 melts taken at regular intervals over a period of 40 minutes after the addition. As a comparison, corresponding curves without the stirrer are given. The difference is striking.

Figure 9 shows the homogenization of manganese in a 40-tonne melt in a Swedish steelworks. Samples were taken at the same spot at intervals of two minutes, and the manganese content is plotted against the time elapsed after the addition. Curve 1, which represents an average of three melts, shows that a time of 11 minutes is required for representative samples with the stirrer. With curve 2 the stirrer was started 13.5 minutes after the addition, and then the melt became homogeneous after another 6.5 minutes.

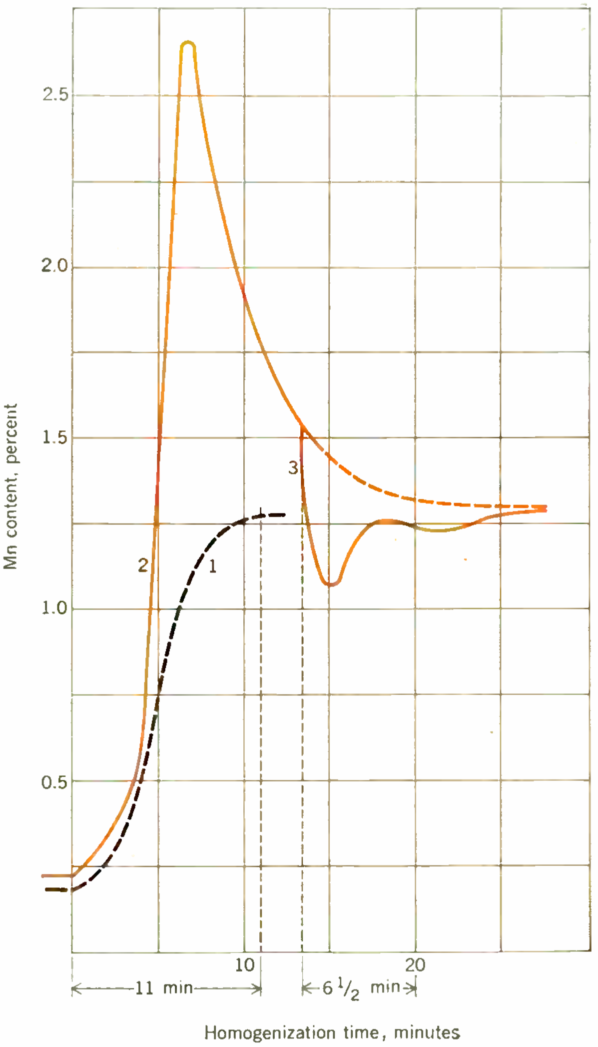
FIGURE 8. Distribution of chromium in samples from a French steelworks. A—Stirred heats. B—Nonstirred heats.



The fact that the samples are representative of the analysis of the entire bath means that there will be a saving in alloy elements, since the additions can be calculated with smaller margins than normally. This is, of course, particularly important when high-alloy steels are produced in large furnaces, but it is also an economic factor of importance to very small furnaces working with expensive alloy elements. This means that steelworks having an induction stirrer can offer more accurate analysis than steelworks in which induction stirring is not used. This very interesting feature is demonstrated by Figs. 10(A) and (B), which represent all molybdenum-alloyed stainless-steel melts (approximately 18 percent chromium, 10 percent nickel) produced in a Swedish steelworks during a certain period. The number of melts, in percent, is plotted against the final analysis, which has been calculated here as a deviation from the lower limit given to the furnace operator.

Chemical reactions. Several of the more important metallurgical reactions in a steelmaking furnace are characterized by an exchange of materials over the

FIGURE 9. Homogenization of manganese in a 40-tonne melt at a Swedish steelworks. 1: Average value of three stirred melts. 2: Stirrer switched off. 3: Stirrer on.



phase boundaries, mainly between the metal and the slag, but also between the furnace refractory and the melt. The reacting elements must be moved in some way to and from these zones to enable the reactions to attain, as far as possible, a state of chemical equilibrium. Since the average distance of the material from these reaction zones is comparatively large, the overall reaction rate is very low in those cases in which the diffusion of the elements in the different phases is the only means of moving the material. Fortunately, in most cases, other factors give rise to a certain degree of external convection or stirring, which compensates for the low rate of diffusion.

The most important exception to this favorable condition is the reduction period for a two-slag process in an arc furnace. The bath is, in itself, almost completely inert during that period. The induction stirrer accelerates the chemical reactions and makes them more complete. An example of this action is illustrated in Fig. 11, which shows desulfurization in a French steelworks, both with and without induction stirring. Figure 12 shows the decrease in the oxygen concentration in the steel as a function of time in a German steelworks.

Slagging. Slagging work is normally very heavy, and involves exposure to intense heat near arc furnaces. When the induction stirrer is used, the slag can be accumulated near the slag door if a suitable stirring direction is selected (see Fig. 13) and it is then a simple matter for the furnace operator to remove the slag. It is, therefore, natural that the stirrer has been accepted with considerable enthusiasm by steelworkers.

It is also important for the slag to be efficiently removable so that no residue of oxidizing slag remains in the furnace when the reducing slag is charged. This results in an increased alkalinity of the reducing slag, which, in combination with a more efficient dissolution of the lime addition, leads to a lower phosphorus content in the melt.

Time reduction. The use of a stirrer permits tap-to-tap time to be reduced

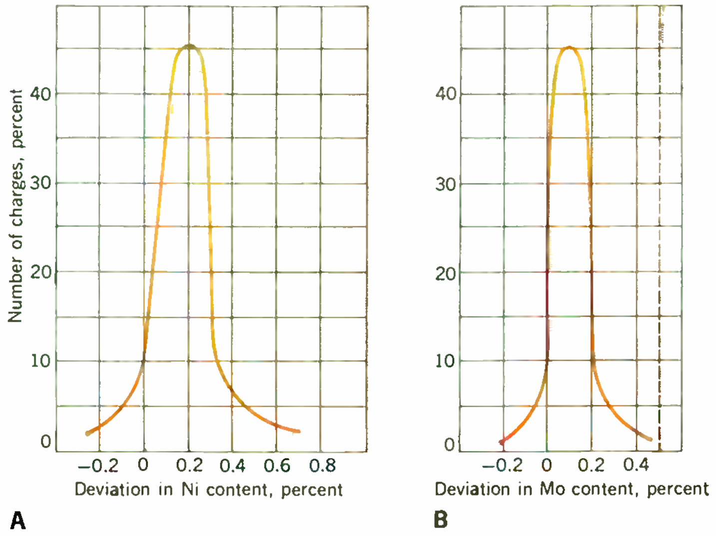
1. During slagging, particularly when complete removal of the slag is desired.

2. During the reduction period, through faster reactions and shorter waiting time for the dissolution of the additions.

3. When it is desired to raise the temperature of the melt immediately prior to tapping, since a higher power can be fed in without large temperature variations developing in the melt.

The time saved depends largely on the grade of steel being produced. For example, in one Swedish steelworks a saving of 30 minutes has been achieved in the production of stainless steel compared with the reladling technique.

FIGURE 10. Distribution of analyses of molybdenum-alloyed stainless steel melts in which a stirrer was employed. Shaded areas represent ranges of acceptabiilty. A—Deviation from nominal content of nickel (42 melts). B—Deviation from nominal content of molybdenum (81 melts).



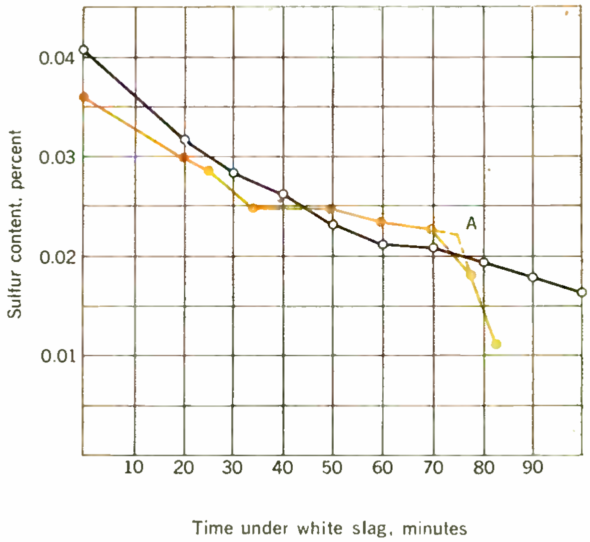
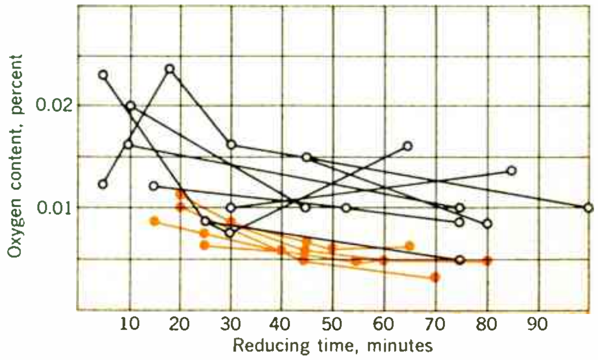


FIGURE 11. Example showing acceleration of desulfurization by stirrer action. Curve in black is an average curve for several melts without use of a stirrer. Curve in color is for an individual melt in which stirrer is switched on after a period of time (point A).

FIGURE 12. Oxygen content during the reducing period at a German steelworks. Curves in black: without stirrer. Curves in color: with stirrer.



Cylindrical stirrers[3]

Design and location. The design of a stirrer for the melt in a ladle is shown in Fig. 14. The coil consists of a number of sections located over one another, fed with alternating current in different phase positions. The coil is generally water-cooled; however, amply designed solid copper conductors are sometimes preferred. These are then allowed to be heated up during the operation, which is often intermittent. Although they are sometimes omitted, sheet-steel yokes are usually included, arranged vertically outside the coil and uniformly distributed around its periphery. Nonmagnetic steel must be used for the ladle shell to enable the magnetic field of the stirrer to penetrate through it.

The height of the stirrer should be at least the same as that of the melt to ensure effective stirring over the entire height of the melt. The radial distance between the stirrer and the melt should be as small as possible to ensure that the greater part of the stirrer field will reach the melt.

As mentioned earlier, there is also a special variant involving combination with a medium-frequency melting furnace in which the induction coil is fed with both medium-frequency, single-phase current for heating and low-frequency, polyphase current for stirring. The design of such a dual-frequency furnace does not differ substantially from that of a normal medium-frequency furnace.

Function. A two-, three-, four-, or six-phase connection may be selected for the stirrer. The higher the number of phases, the greater will be the winding factor—that is, the more powerful will be the stirring for a given field strength. As a rule, however, the three-phase or four-phase connection is selected.

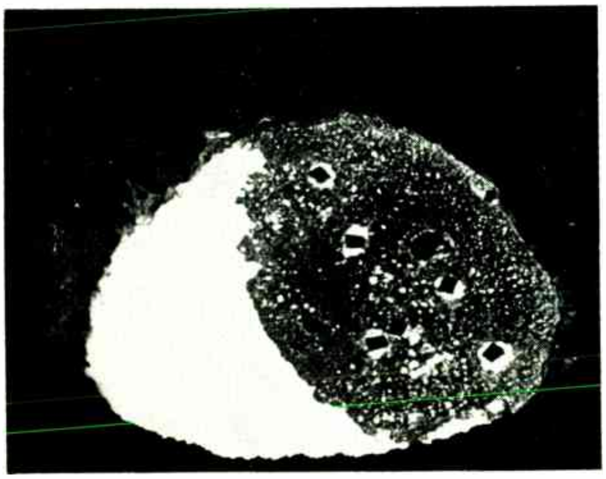
Figure 15 shows the design of a four-phase connection. The way in which the traveling field occurs is also illustrated here, in the same way as for the straight stirrer. Again, the synchronous speed of the traveling field will be

$$v_s = v \cdot 2\tau$$

where τ is the pole pitch—that is, half the height of the stirrer. The traveling field exerts an axial pull on the melt. The force per volume unit is at its largest at the cylindrical surface and decreases inward toward the center axis. The calculations of this gradient are complicated. With a field strength of 100 kA/m at the inner surface of the coil, a force related to the cylindrical surface of the melt of 200–500 N/m^2 will be obtained with a suitably selected frequency for normal radial distances between coil and melt. The direction of the traveling field can be reversed by reversing of one of the phases.

Bath motion. Figure 16 shows the appearance of the bath motion. Since the melt is to flow backward in a direction opposite to that of the traveling field in the center, the dragging force should have decreased to almost zero within this zone, which means that the frequency must not be too low.

FIGURE 13. View looking down into the melt in an arc furnace having a shell diameter of 5 meters during stirring. Slag is accumulated in front of slag door.



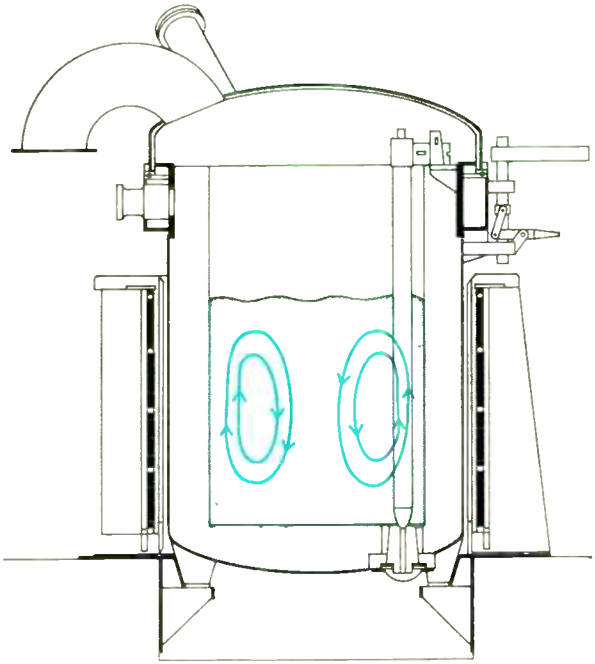
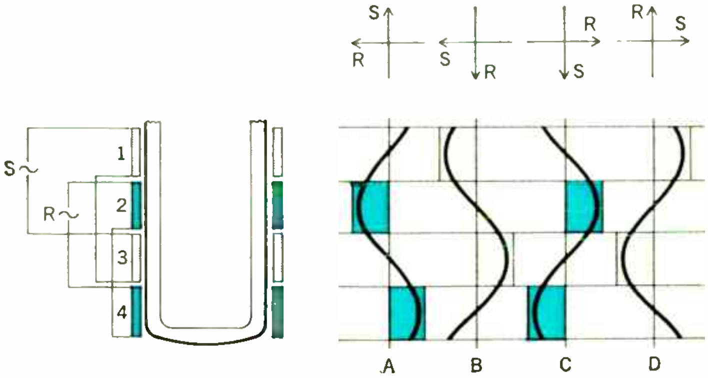


FIGURE 14. Ladle with molten steel inserted in a cylindrical stirrer.

FIGURE 15. Distribution of field strength in a cylindrical stirrer at different times.



The bath motion is clearly very suitable when it is desired to move all parts of the melt up to the surface, as during slagging or vacuum degassing. Homogeneity and temperature equalization will also be good with this form of stirring. When one of the phases is reversed, the direction of flow will also be reversed.

Flow rate. As is the case with the straight stirrer, it is difficult to determine a theoretical relationship between flow rate and stirring force. Turbulence occurs even at very low flow rates. A calculation of the flow rate on the assumption that the flow is laminar would yield quite unrealistic results for the forces normally produced by the stirrer.

In a practical investigation conducted in the United States by the Republic Steel Corporation,[4] a piece of a manganese isotope was added to the surface of the melt in a 90-tonne ladle with a depth of about 2.7 meters during steady-state stirring with a specific stirring force of about 175 N/m^2. Geiger–Müller counters were placed beneath the ladle, and the time at which these started to indicate that isotopes had reached the bottom was measured; the result was 27 seconds. It was also observed that after 55 seconds the sputtering had become regular, which showed that the isotopes were then uniformly distributed in the melt; that is, the melt was homogeneous.

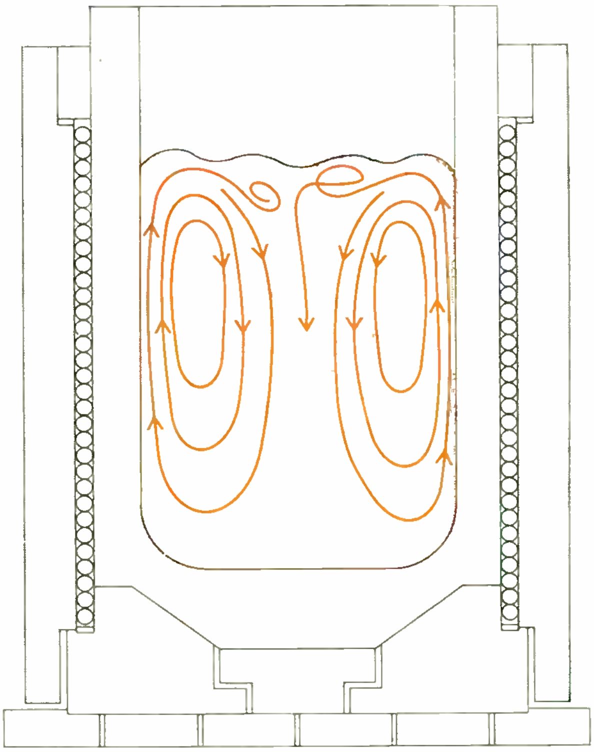
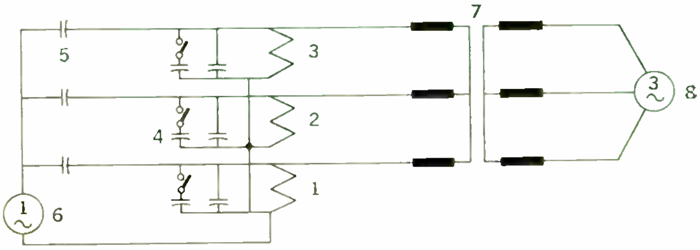


FIGURE 16. Bath motion produced by cylindrical stirrer.

FIGURE 17. Connection diagram for dual-frequency furnace. 1, 2, 3: Induction coils. 4: Parallel capacitor for medium frequency. 5: Series capacitor for medium frequency. 6: Medium-frequency generator. 7: Transformer. 8: Low-frequency generator for stirring.



A mean flow rate down in the center of the ladle can be calculated from these investigations as follows:

$$v = \frac{2.7}{27} = 0.1 \text{ m/s}$$

It should be noted that the stirrer in this case only covered about 80 percent of the height of the melt, which probably meant that the average rate was lower than normal.

The results from a Swedish plant can also be mentioned. In a ladle with 25 tonnes of steel the melt was stirred with an estimated specific stirring force of 230 N/m^2. An average rate of 0.5 m/s was then observed on the surface of the melt.

Electric equipment. When the frequency is low (for example, below 6 Hz), rotary converters with cyclically excited dc generators can be selected in the same way as for straight stirrers. Static converters of the type mentioned in the section dealing with straight stirrers can be designed for almost any frequency range, but they are more expensive than the rotary type. Salient-pole synchronous generators have been used for $16\frac{2}{3}$ and 25 Hz.

Line frequency (50–60 Hz) is also feasible for smaller melts—for example, up to one tonne. The limit is not sharply defined, however, and the reduction in stirring force in relation to that obtained with the optimum frequency and unaltered field strength is not discouraging, even for 10-tonne melts. The electric equipment will then naturally be simple, particularly if means for controlling the stirring force are not required.

Figure 17 shows a connection used for a dual-frequency furnace. Three-phase stirrer equipment is adopted here.

It is important that the medium-frequency current reaching the low-frequency generator be kept small; moreover, the low-frequency current reaching the medium-frequency generator should also be limited. On the medium-frequency side, the series capacitor is divided and connected in such a way that it will create a barrier for the low-frequency current. On the low-frequency side, no medium-frequency current will be obtained with the connection selected. All the points at which the low-frequency circuit is connected to the induction coil have in fact the same medium-frequency potential.

Since a low value is often selected for the medium-frequency voltage, the low-frequency voltage will also be low. This means that a low-frequency transformer is generally required to enable practical data to be obtained for the low-frequency generator.

Metallurgical considerations

Vacuum degassing of steel. Although the usefulness of vacuum degassing has been known for a long time, the technical and economic prerequisites were lacking until the 1950s.[5] Several methods were developed, of which ladle degassing techniques proved to be the most economical. The simplest method was to lower the ladle, with steel, into a vacuum chamber, which was then evacuated. However, only the upper parts of the melt were degassed. Further down in the melt, the hydrostatic pressure was so high that the entrapped gases could not be desorbed. It was therefore necessary for the melt to be moved up to the surface so as to achieve a good degassing of the entire melt. It was found that the induction stirrer is an effective aid in achieving this purpose. The following advantages can be mentioned:

1. The stirrer can be easily combined with vacuum equipment—for example, in such a way that the ladle is provided with a vacuum-tight cover connected to a vacuum pump. It is then necessary to evacuate only the small volume of air inside the ladle itself.
2. The stirrer yields a bath motion that is highly suited for the degassing (see Fig. 16).
3. The induction stirrer is kept quite separate from the chemical reactions in the melt and from the vacuum degassing itself. This technique makes for the greatest possible flexibility concerning the choice of process.
4. The stirrer can be easily combined with heating equipment. This combination is effective since it is necessary both to compensate for the heat losses during the gassing and to add heat for melting the alloy elements added in conjunction with the degassing. In the ASEA-SKF process,[6–8] heating is carried out with the aid of arcs in the same manner as in an arc furnace. The electrodes are located in a roof structure, which is used alternately with the vacuum-tight cover.
5. As a result of the efficient stirring, the alloy additions are thoroughly distributed in the melt, and the temperature is equal.
6. There is no limitation in the size of melts that can be inductively stirred. At present, as far as is known, stirrers for melts up to some 200 tonnes are in use.

Vacuum melting. Induction stirring in combination with medium-frequency heating in the dual-frequency furnace was originally developed to accelerate the metallurgical processes.[3] As mentioned earlier, however, this furnace type was not widely used during the first several decades after its introduction.

In the 1950s nickel-based high-temperature alloys were developed for use in turbine blades for jet engines.[9] The production of these high-temperature alloys necessitated vacuum degassing, both to raise the contents of aluminum and titanium to the desired level and to ensure that the concentration of entrapped gases would be as low as possible. Induction melting furnaces proved to be very suitable for this purpose.

As the demand for such alloys increased, so did the need for larger furnaces. In this connection, it also became necessary to stir the melt, in order to achieve a homogeneous analysis and to ensure that all parts of the melt would be exposed to the vacuum above the surface. The dual-frequency furnace was the natural answer to this problem.

Plants in operation. Figure 18 shows equipment for a 30-tonne ASEA-SKF vacuum degassing plant as an

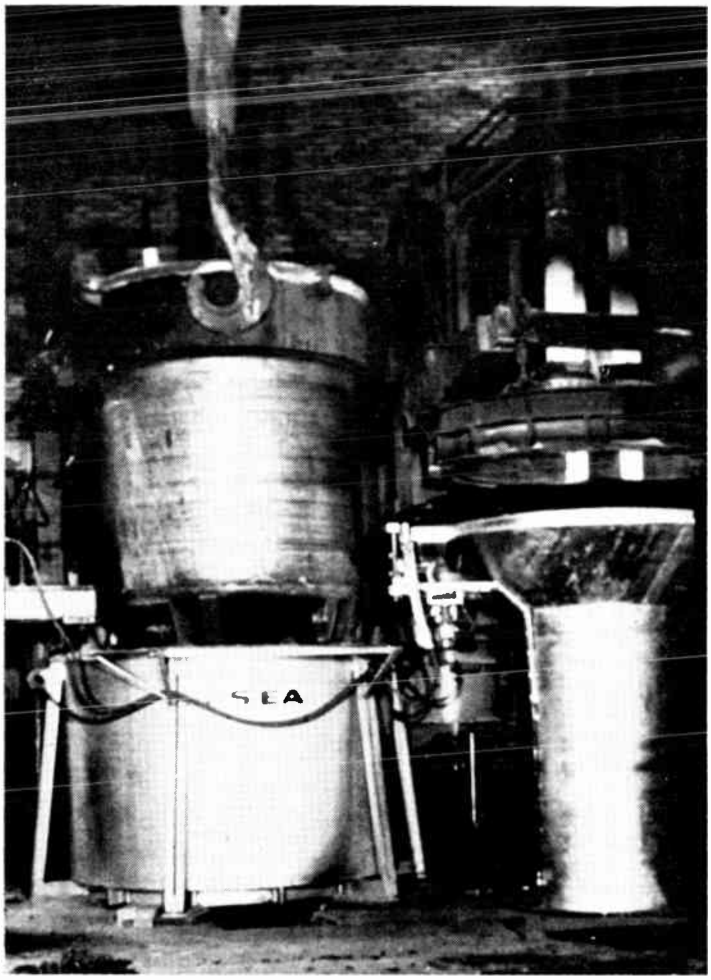
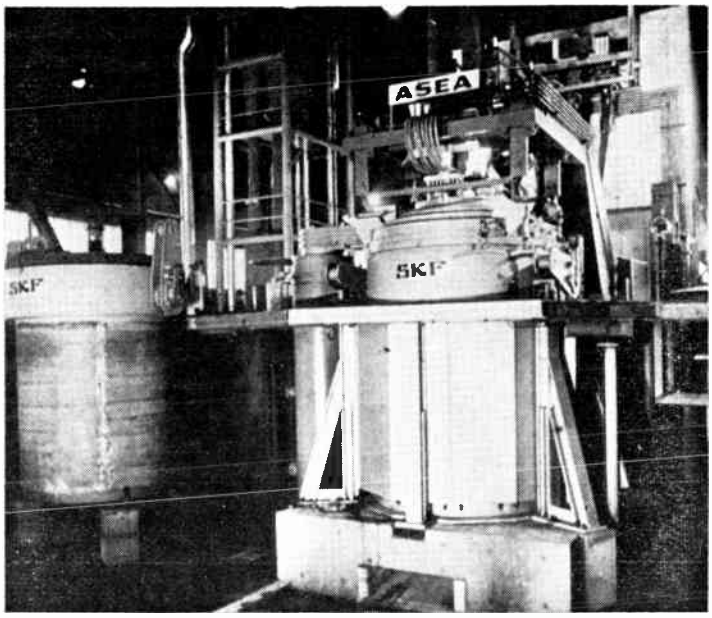


FIGURE 18. ASEA-SKF vacuum degassing plant for 30 tonnes. Ladle, at left, is shown being lifted out of the stirrer coil. The electrode roof is shown on the right.

FIGURE 19. ASEA-SKF vacuum degassing plant for 60 tonnes. A ladle is inserted in the stirrer coil, and the electrode roof is in position. To the left of the stirrer can be seen part of the steam ejector.



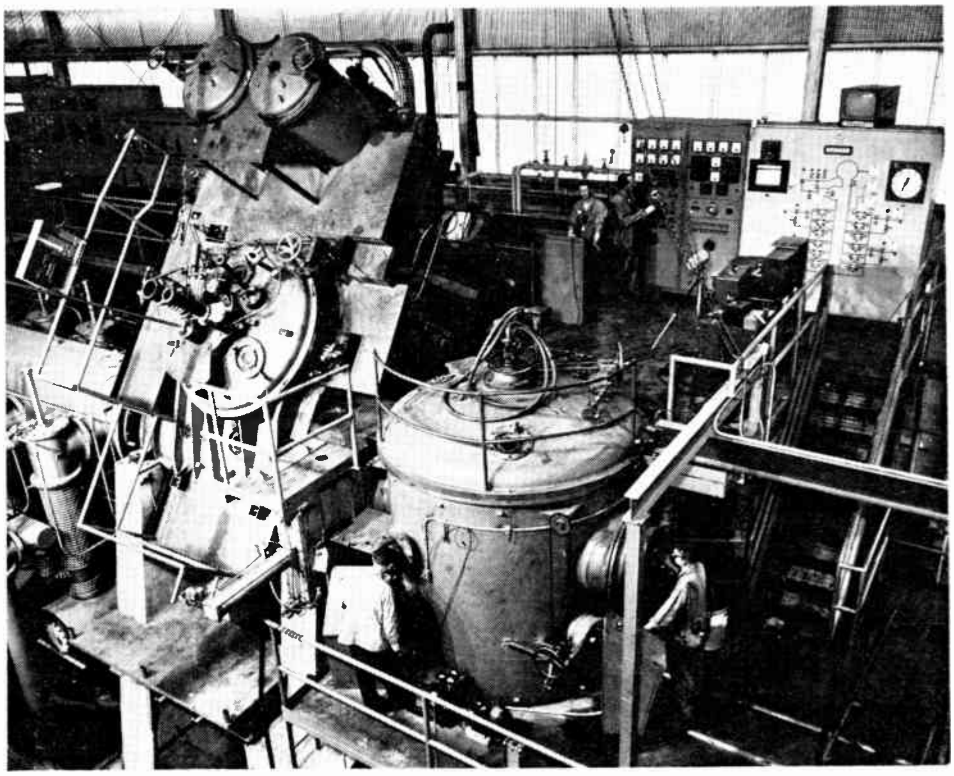


FIGURE 20. Dual-frequency high-vacuum furnace with a charge capacity of 2 tonnes. For melting: 1100-kW, 800-V, single-phase 890-Hz feed. For stirring: 2 × 400-kVA, 85-V, two-phase, 30-Hz feed. The furnace tank is in a tilted position. The mold tanks are at the right and the diffusion pumps at the left.

example of the use of the induction stirrer for ladle degassing. At the bottom, to the left, can be seen the stirrer itself, which is fed with 500-kVA, two-phase, 315-volt/phase, 1.75-Hz power. The ladle, which has a nonmagnetic steel shell, is being lifted out of the stirrer. At the top, to the right, can be seen the roof, with electrodes for arc heating, which has been swung aside. The heating power input is 3000 kW.

A 60-tonne ASEA-SKF plant is illustrated in Fig. 19. The ladle is shown here inserted in the stirrer and with the electrode roof in position. The stirrer is fed with an input of 2 × 350 kVA, 260 volts, 1.4 Hz. The arc-heating portion of the system is designed for 7200 kW.

Figure 20 shows a 2-tonne vacuum melting furnace as an example of a dual-frequency furnace. The furnace itself is located in a vacuum tank, which is illustrated in the tilted position here. To the right can be seen a vacuum tank with molds into which the vacuum-degassed melt is teemed. To the left is the vacuum-pumping installation with oil-vapor (booster) diffusion pumps. The furnace is fed with an 1100-kW, single-phase, 800-volt, 890-Hz source for heating and an 800-kVA, two-phase, 85-volt/phase 30-Hz source for stirring.

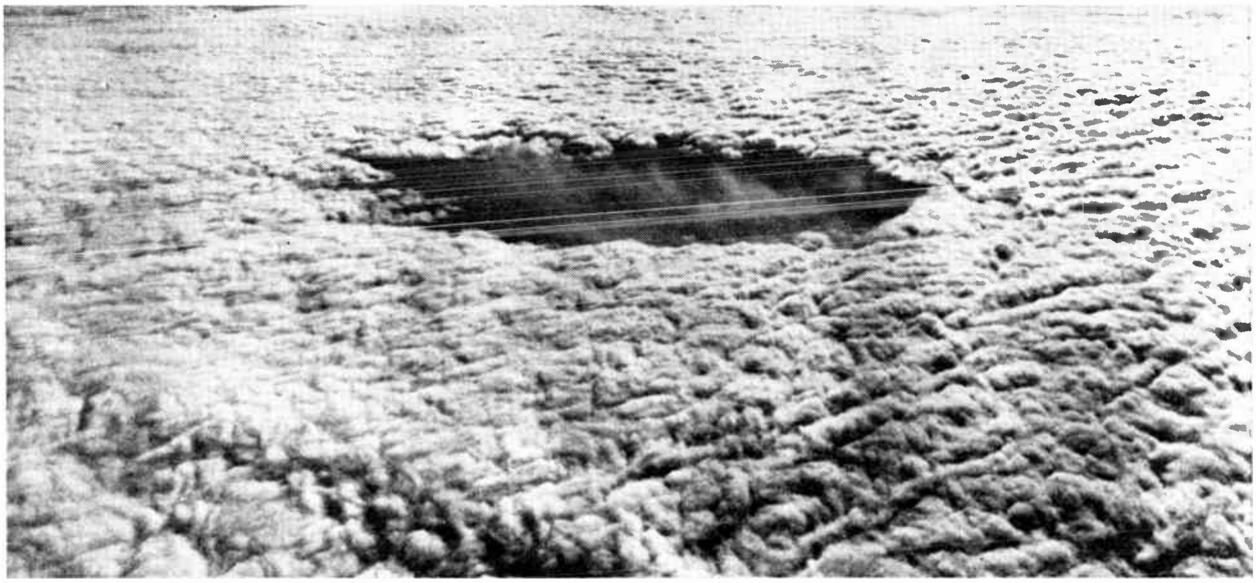
Conclusion

It should be apparent from the foregoing discussion that the magnetic traveling fields produced by straight or cylindrical stirrers provide a useful tool for the metallurgist. Now that the size of melts is growing, the stirrers will prove to be an increasingly valuable aid for producing homogenization, temperature equalization, rapid and complete chemical reactions, and effective degassing.

REFERENCES

1. Dreyfus, L., and Nilsson, F., "Induktiv omröring i ljusbagsugnar (Induction stirring in arc furnaces)," *Jernkontorets Ann.*, vol. 133, pp. 371–468, 1949 (in Swedish).
2. Hammarlund, P. E., and Hanas, B., "Results obtained with inductive stirring in arc furnaces," *ASEA J.*, vol. 32, pp. 147–156, 1959.
3. Dreyfus, L., "Virvelströmsugnens senaste utveckling (Latest trends in the development of eddy-current furnaces). *Värmländska Bergsmannaföreningens Ann.*, pp. 101–124, 1934 (in Swedish).
4. Smith, R. D., "Ladle vacuum degassed steel, production and quality," *J. Metals*, vol. 18, pp. 59–61, Jan. 1966.
5. Eketorp, S., "Vakuumbehandlingens möjligheter (The possibilities of vacuum degassing)." *Jernkontorets Ann.*, vol. 150, pp. 585–685, 1966 (in Swedish).
6. Tiberg, M., Buhre, T., and Herlitz, H., "The ASEA-SKF steel refining process," *ASEA J.* vol. 39, pp. 87–89, 1966.
7. Stickler, H. "Electrical equipment for the ASEA-SKF steel refining process," *ASEA J.*, vol. 39, pp. 90–92, 1966.
8. Östlund A., "Vacuum equipment for the ASEA-SKF steel refining process," *ASEA J.*, vol. 39, pp. 93–95, 1966.
9. Darmara, F. N., Huntington, J. S., and Machlin, E. S., "Vacuum induction melting," *J. Iron Steel Inst.*, vol. 191, pp. 266–275, Mar. 1959.

Yngve Sundberg graduated in electrical engineering from the Royal Institute of Technology in Stockholm, Sweden, in 1945. In 1949 he joined the technical staff of the Furnace Division of ASEA, in Västeras, where he has served as head of the Research and Development Section since 1961. He received the degree of doctor of technology from the Royal Institute of Technology in 1966. His doctoral dissertation dealt with induction heating, with special reference to bodies inside metallic shells.



IT TOOK JUST eleven pounds of dry ice pellets and one hour to open the 5-kilometer hole you see in this otherwise impenetrable deck of supercooled stratus clouds. Pellets were made from liquid carbon dioxide right on the Air Force plane that dropped them into the clouds. This technique is being used to good economic advantage to clear supercooled fog from airport runways. Other weather modification goals, ranging from rainmaking to hurricane control and lightning and hail suppression, are still technologically and economically unfeasible.

Electricity and weather modification

II. Plans and programs, prospects and purpose

International research programs may yield mathematical models for assessing the possibility of climate modification—both intentional and inadvertent—by 1980. Meanwhile it's worth assessing where we stand in domestic research programs, how we got here, and where we are going and why

Seymour Tilson *Staff Writer*

The international and domestic research programs described in this article may close the gap between dream and reality in weather control, especially if atmospheric and electrical scientists work together more closely than in the past. Can such cooperation also close the gap between research directed at realizing the dream and research needed to prevent a nightmare, that pollution may place the weather beyond anyone's control and create climatic instabilities that could lead to social chaos?

In last month's installment we surveyed electrical weather phenomena on scales ranging from the submicroscopic to the extraterrestrial. We explored a wide variety of processes in order to emphasize that the search for a comprehensive theory of cloud electricity would have to extend far beyond the clouds that—like those shown in the introductory figure above—are the immediate targets of most weather modification programs. Such a theory could be critical to the success of these programs according to reports issued by the National Science Foundation [1] and the National Academy of Sciences–National Research Council.[2] The search for it will have to extend to the surface of the windswept sea. When Duncan Blanchard of the State University of New York at Albany evaluated thoroughly,[3] for the first time, the quantity of electric charge entering the atmosphere from the world ocean, he found that droplets breaking free from jets of seawater that rise at high speeds from the bottom of collapsing air bubbles (Fig. 1) could carry positive charge into the air at a rate of about 160 amperes. This may not sound like much of a charge flow, coming as it may from three quarters of the earth's surface, but it is about 17 percent of the air–earth current over the world ocean that the classical spherical-condenser theory of atmospheric electricity attributes to thunderstorm activity. This is not a large entry in the atmosphere's electrical budget, but it is one of the most carefully documented, and when viewed in the perspective provided by the work of Imyanitov and Chubarina[4] (which we discussed earlier), enough to cast further doubt on the spherical condenser theory.

In an interesting followup to Blanchard's work, however, E. T. Pierce and A. L. Whitson of the Stanford Research Institute [5] suggest that negative charge might be released into the air over bodies of fresh water—rivers, lakes, and the like—which are sufficiently agitated by the wind, rapids, or waterfalls, 100 times as intensively as over seawater. They contend therefore that, since the areal ratio of fresh water to seawater is the reverse of this, the global flow of

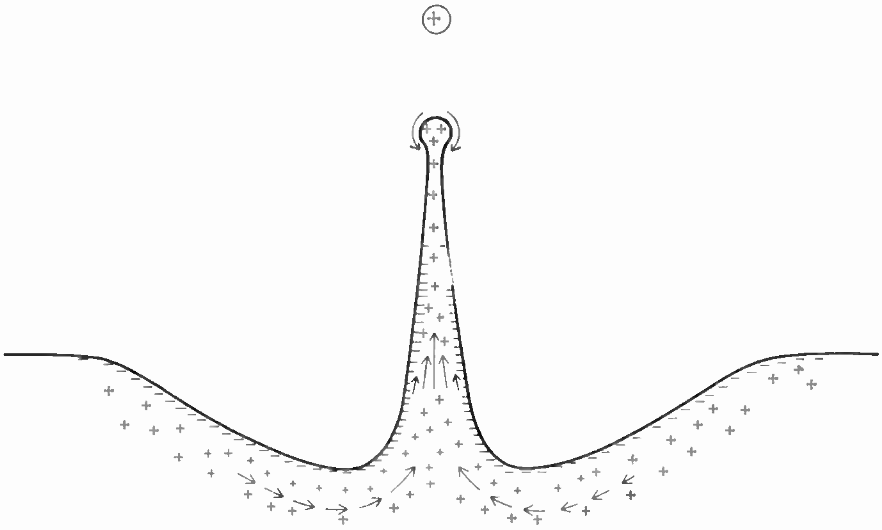


FIGURE 1. Using an argument akin to that used to explain the streaming potential, Duncan Blanchard (of the State University of New York at Albany) proposed the mechanism illustrated here to explain how jets of seawater, rising at high speeds from the bottom of collapsing air bubbles, could eject positively charged droplets into the air at an estimated oceanwide rate of 160 A. The charge on the airborne drop at the top is positive, although the assumed electric double layer at the water surface has its negative section nearest the surface, because laminar velocity shear in the bubble jet (suggested by different lengths of arrows representing velocity vectors) favors upward transport of deeper water over that of surface water through the core of the jet. The flow of charge across the air–sea boundary is only one element of the more comprehensive mathematical model of atmospheric processes that's needed to put weather modification programs on a sounder scientific basis and a more prudent ecological one.

negative charge over fresh water might nullify the 160-ampere positive current over the sea.

The need is critical for a fresh, searching look at every existing entry in the earth's electrical budget. Resolving differences of opinion about the relative magnitude of the entries that comprise the global electric budget will help to resolve differences of opinion about the origin of the feeble electric fields and currents that prevail in the cloud-free parts of the atmosphere. Theories of cloud and precipitation electricity are numerous and many invoke the fair-weather field as freely as the ancient playwrights invoked the *deus ex machina*. Depending on the requirements of any particular theory, the fair-weather field is called upon to serve as the ultimate source or the ultimate product of those physical–chemical processes through which clouds grow, precipitation develops, and electric charge is generated and separated. Some decision about the relations of the fair-weather field to electrified clouds and the precipitation they yield is long overdue. And ecologic prudence, another name for systems thinking on an unparalleled scale, is an even more compelling reason for trying to understand the atmospheric electric system as the entity it is before tinkering too extensively with its parts.

A chief barrier to understanding this system is that we have not yet created satisfactory institutional procedures for mounting a sustained assault on environmental problems. By their very nature, these transcend the boundaries of any nation and any single discipline. Indeed, such problems exceed the capabilities any nation or discipline can bring to bear upon them. The striking advances in understanding of our terrestrial environment that were achieved during the International Geophysical Year and the International Years of the Quiet Sun point the way to what may be needed in the realms of atmospheric and cloud electricity, and in weather modification as well. The sponsors of a worldwide Ten-Year Program in Atmospheric Electricity that is just beginning, and that has as its chief objective the critical testing of the spherical-condenser theory, are in fact basing their efforts on these earlier models of successful international cooperation. This program will be carried on in the context of, and in close cooperation with, a still broader Global Atmospheric Research Program (GARP) that the international meteorological fraternity is now preparing to pursue through the decade of the 1970s. Both programs are necessary if we ever hope to understand the atmosphere well enough to control it and use it without abusing it.

The Global Atmospheric Research Program

The primary impetus for GARP came from the recognition among meteorologists that a shortage of data was handicapping the development of better computer-based numerical models of the atmospheric circulation. Improvements in the accuracy and period of weather predictions on all geographic scales—obviously a critical baseline for attempts to modify the weather—can be attained only through an understanding of the circulation on all scales. Numerical models capable of simulating these circulations realistically are considered to be the existence proofs of this understanding. Six global atmospheric circulation models have now been designed in the United States and the Soviet Union. Using data and equations already available, and present computer capabilities, these models show temperature, pressure, and gross motions in the entire troposphere and lower stratosphere. Other models show the circulatory structure of more limited atmospheric features, such as hurricanes or frontal disturbances. Scientists in Japan, Australia, East Germany, and other countries are also designing circulation models on various scales. But none of these models can be considered definitive. Atmospheric scientists believe that the development of definitive models depends not only on getting more data with which to test present models and the predictions they provide, but on getting new computers large enough to handle these additional data and capable of handling more elaborate modeling equations.

Even the most elaborate present computers will not be able to cope with the flood of data that GARP and a closely related observational, communications, and data analysis program—the World Weather Watch—are expected to obtain. New supercomputers, several hundred times faster and more voracious of data than present models, are now being designed to meet this need. Such computers also will be able to handle the more complex equations needed to represent various important processes and parameters omitted from most present mathematical models. These include the humidity parameter, for example, as well as most of the microphysical and electrical processes critical in cloud and precipitation growth.

Atmospheric data—meteorological and electrical—are particularly inadequate over tropical, oceanic, and sparsely populated areas. Together these make up more than 80 percent of the earth's surface. As GARP and related experiments move into higher gear they will rely increasingly on satellites to obtain data from these areas. Such data will include not only daytime cloud-cover photographs, but nighttime coverage by infrared sensors and by extra-sensitive TV cameras that can photograph the earth by moonlight. Satellites may also be used eventually for measuring ozone, aerosols, and certain other atmospheric constituents. They also may be valuable in attempts to monitor worldwide thunderstorm activity, one of the core goals of the Ten-Year Program in Atmospheric Electricity, which we shall later discuss in greater detail. In addition, satellites will serve to gather and relay information from the new types of automatic weather stations—constant-density horizontally drifting balloons, stationary and drifting ocean buoys, and automatic ground stations in remote land areas—that many IEEE SPECTRUM readers have no doubt heard about, and perhaps have even worked on, by now.

Thus, the data available for the supercomputers will eventually be finer in detail—both horizontally, as areas on the globe are filled in, and vertically, as data are obtained from more atmospheric levels. Moreover, the new data will include not only such conventional meteorological parameters as cloud cover, temperature, pressure, and humidity, but also such unconventional parameters as solar (and perhaps urban) heat inputs to the atmosphere, heat losses to space and the sea, energy transformations and momentum transfers over land and sea, and the transport of water (and perhaps other substances, both natural and those introduced by man) through the hydrologic cycle.

GARP is designed to respond primarily to the needs of basic atmospheric research. These, though intimately connected with the needs of existing weather prediction services, are not identical with them. GARP will assemble data for research purposes, not only from existing networks but from temporary stations and networks operated for short periods of time to produce data on a more than usually intense basis. Results of GARP research, however, will help to determine what the future patterns of observing systems should be in order to provide the nations of the world directly with more efficient operational weather prediction systems.

Congressional support for this large-scale, long-range international effort seems assured—at least in moral terms. How firm this commitment will prove to be when appropriations must be made remains to be seen. Last year, the Senate resolved, with the House concurring: "That it is the sense of Congress that the United States should participate in and give full support to the world weather program which includes (1) a world weather watch—the development and operation of an international system for the observation of the global atmosphere and the rapid and efficient communication, processing, and analysis of worldwide weather data, and (2) the conduct of a comprehensive program of research for the development of a capability in long-range weather prediction and for the theoretical study and evaluation of inadvertent climate modification and the feasibility of intentional climate modification. . ." The second item refers to GARP.

No evaluation of inadvertent climate modification, or of the feasibility of intentional climate modification, can be complete unless it includes the electrical dimensions of the earth's atmosphere.

The Atmospheric Electricity decade

The Atmospheric Electricity Ten-Year Program was drafted by the Joint Committee on Atmospheric Electricity of two associations of the International Union of Geodesy and Geophysics: the International Association of Geomagnetism and Aeronomy, and the International Association of Meteorology and Atmospheric Physics. In addition to its major goal—a critical check of the spherical condenser theory—the sponsors of the program hope to compare the different methods now used to measure such atmospheric electric parameters as field intensity, space charge, current density, air conductivity, and so on. They hope to accomplish this by applying diverse measurement techniques to the same, presumably worldwide sources of change in these parameters, as synchronously as may prove possible. To this end, they, like the meteorologists planning GARP, have provided for "intensification" periods, short intervals of enhanced and coordinated measuring activity.

The first such period was planned for the first few months of this year.

Measurements suggested and planned by the committee for future intensification periods include remote satellite and ground-level monitoring of several kinds of lightning-induced electromagnetic activity, measurement of the earth–ionosphere potential difference, and measurement of air-earth current densities at many stations at once. Relations between these electric parameters and meteorological parameters also will be investigated simultaneously, once the earth–ionosphere potential difference—a necessary baseline for such correlation studies—is known. During the first and most subsequent intensification periods, data will be taken only during predetermined and agreed-upon hours, and results will be delivered in the form of hourly averages to a Data Collection Center in Leningrad.

Interest in the Atmospheric Electricity Ten-Year

Program is running quite high. Dr. Hans Dolezalek (Secretary of the Committee, 1812 Drury Lane, Alexandria, Va. 22307) informs me that his supply of a manuscript describing the program,[6] originally prepared in several hundred copies, was completely used up by early last fall because demand was much greater than had been anticipated. Preparation of a second edition was in progress at that time, and I suggest that interested readers write directly to Dr. Dolezalek for detailed information the problems posed by the Program and the possibilities it offers.

Getting a grip on the ionosphere

Some proposals for achieving the goals of this ambitious program are already in the pipeline of course. To measure the electric potential of the ionosphere directly, for example, Bernard Vonnegut of the State University of New York at Albany—who heads the Units, Methods, and Instruments Working Group of the IUGG Joint Committee on Atmospheric Electricity—hopes to use the null-measurement system illustrated in Fig. 2. The measuring instruments are hung from a tethered balloon flown at an altitude in excess of 15 000 meters, where the electric potential can be expected to be very nearly that of the ionosphere. Attached beneath the balloon is a conducting electrode carried on an insulating line. At the electrode is a radiosonde field meter that measures the electric field produced at the electrode. Connected to the electrode and suspended from it is a tapered steel wire attached to a winch on a boat beneath. This winch is mounted on insulators capable of withstanding about 500 kV—the maximum expectable voltage—and is connected to a variable dc high-voltage supply used to maintain the winch and the steel wire at any desired potential relative to the earth. The apparatus is operated by increasing the potential of the winch and wire in a positive direction until the radiosonde on the electrode at its upper end gives a signal to a radio receiver on the ship, indicating no electric field. At this point the applied voltage is a direct measure of ionospheric potential. The apparatus may be operated manually or automatically by adjusting the voltage to give a zero reading and recording this voltage. Alternatively, it can be operated at a fixed, high voltage approximating that estimated to exist in the ionosphere, and the difference between this voltage and that of the ionosphere can be determined by the signal from the radiosonde.

An important requirement of the high-voltage apparatus is that it be capable of supplying the corona current that will be lost from the lower part of the wire connecting the instruments to the ship. Vonnegut anticipates that this current will be somewhat less than 10 mA.

In principle the apparatus described should work as well over the land as at sea. However, if it is flown from a boat, it may be possible to minimize stresses on the wire by following along under the balloon as it moves with the wind, which at the altitude of the balloon would be relatively constant in direction.

There is good evidence that the ionosphere is at a high positive potential with respect to earth. Estimates based on the integration of soundings of the fair-weather electric field, made from balloons and airplanes, indicate that the total potential difference is of the order of 300 kV. Other estimates based on the measured air–earth current and on estimates of the columnar resistance of the atmosphere between ionosphere and earth also range about this value. Though such estimates seem reasonable, we will not know how reasonable they are until Vonnegut's proposed

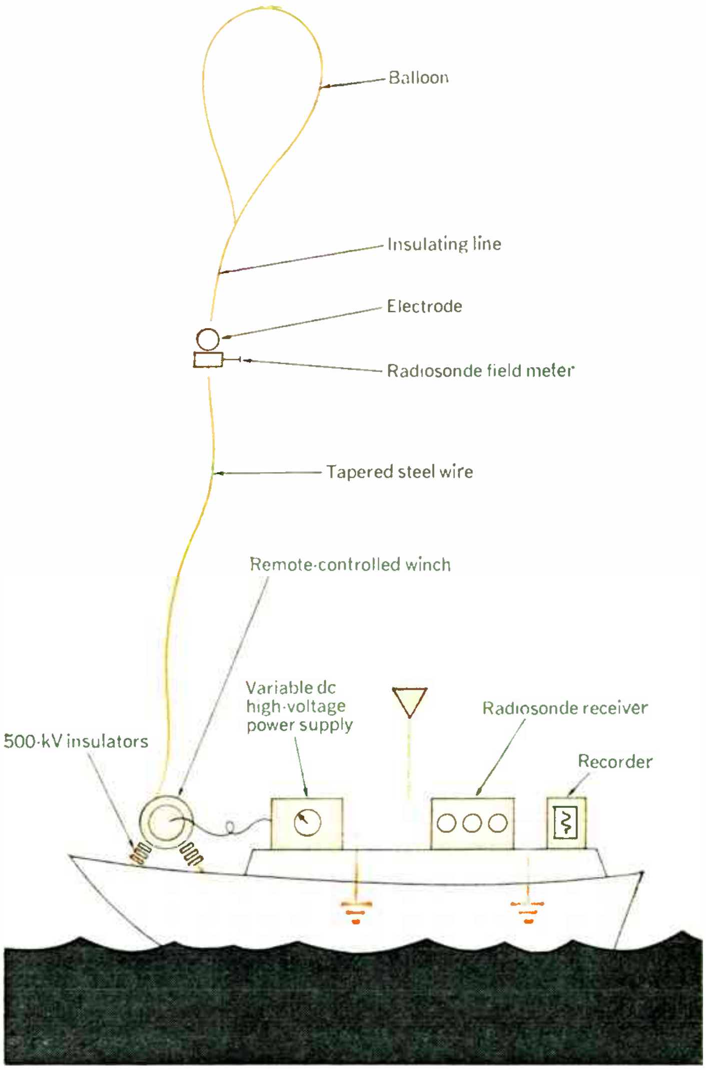


FIGURE 2. One aim of the global Ten-Year Program in Atmospheric Electricity is to determine the electric potential of the ionosphere. This is a necessary baseline for studies that seek to correlate the atmosphere's electric parameters with its meteorological parameters. Bernard Vonnegut (of the State University of New York at Albany) hopes to measure this potential directly and continuously with some version of the system illustrated here. The system, fully described in text, is essentially a null-measurement technique for determining the potential at the altitude of the balloon-borne instruments. The plan is to fly these at altitudes in excess of 50 000 feet (about 15 000 meters) where the potential approximates that of the ionosphere. Since winds at these altitudes are relatively constant, the boat may be able to follow the balloon to minimize stresses in wire.

technique or some other one is used successfully. Even then it will be some time before we know how this potential changes over various time scales—minutes, hours, days, seasons, years—and over various geographic scales, and how these changes are influenced by solar flares, volcanic eruptions, high-altitude jet and rocket exhausts, and perhaps other factors. In the spherical-condenser theory, fluctuations in ionospheric potential are assumed to result chiefly from fluctuations in worldwide thunderstorm activity. A thorough check of the spherical-condenser theory therefore requires not only direct and continuous measurements of ionospheric potential but equally continuous, if not equally direct, measurements of worldwide lightning.

Monitoring thunderstorms around the world

Dr. E. T. Pierce of the Stanford Research Institute surveyed the problem of monitoring thunderstorms around the globe with great care,[7] at the Fourth International Conference on Universal Aspects of Atmospheric Electricity held in Tokyo a year ago.

Comprehensive monitoring of global thunderstorm activity would be more than satisfactorily achieved, he observed, if equipment capable of registering the position and time of occurrence of every lightning flash were functioning continuously. The successful routine operation of ground-based sferics networks has demonstrated that it has been technically feasible for at least 20 years to monitor lightning on a worldwide scale, as it occurs, by using a series of interlinked networks. Sferics are the familiar radio-frequency signals generated by lightning. Lightning generates these signals at all radio frequencies, and, for certain ranges of frequency, these propagate over the surface of the earth with only slight attenuation. This is true, for example, over portions of the VLF (3–30-kHz) band, and consequently VLF atmospherics, even after propagating several thousand kilometers, are still easily identifiable above the local background noise.

Sferics work up until now has tended to concentrate upon the VLF band for several reasons, none directly connected with the problem of monitoring thunderstorms. Communication needs have provided the major stimulus. A vast literature on VLF has grown in the last ten years as a consequence of these studies. However, other radio frequencies, notably those in the ELF ($<$ 3-kHz) range, seem better fitted for global thunderstorm monitoring. Comparing the relative utility of 10-kHz and 100-Hz frequencies for this purpose, Dr. Pierce notes that the lower frequency has several definite advantages: The propagational attenuation is smaller and this attenuation depends less than it does at higher frequencies on variations in ground conductivity, path orientation with respect to the geomagnetic field, and temporal changes in the ionosphere. Moreover, interpretation difficulties introduced by differences in the source signals—as between intracloud lightning flashes and discharges to earth, for instance—tend to be less marked at ELF.

An interconnected system of sferics networks can undoubtedly locate all lightning flashes. However, using conventional techniques under usual atmospheric conditions, stations could not operate effectively at separations much greater than perhaps a thousand kilometers. For worldwide coverage this would require an inconveniently large number of stations. Many attempts have therefore been made over the past few years to develop methods whereby the origin of a sferic can be established from a single station[8]; such a station could then replace an entire network of conventional stations. Single-station techniques usually combine a bearing observation with the measurement of some characteristic of the incoming atmospheric from which distance can be deduced. A system consisting of a few advanced-type sferics stations could effectively monitor worldwide thunderstorms if the stations were located where the background man-made noise level is low and in areas that, though not far removed from the main equatorial thundery areas, do not experience many local storms. Such locations are most likely to be found at latitudes approximating those of the Tropics of Cancer or Capricorn.

The barriers to extending ground-based sferics networks have been primarily political and monetary.

How to reduce both the costs and the number of countries concerned?

With satellites the geopolitical factors of multination cooperation in implementing suitable ground-level networks would not exist. But serious research problems related to the nature of the signals and sensors do exist and Dr. Pierce outlined many factors, some not immediately obvious, that could severely limit the capabilities of satellite-borne sensor systems.

Limitation of satellite sensors

The characteristics of thundery activity themselves impose the most severe restrictions upon the effectiveness of thunderstorm surveillance from satellites. Storms typically occur within complexes some 30–40 km in diameter and in most weather situations the storm complexes are separated by perhaps a hundred kilometers. In any storm complex there are usually only one or two electrically active cells; these are only about 6 km in diameter and they produce, on the average, only some three lightning flashes per minute. These geographic and time considerations, when measured against the constraints imposed by various possible combinations of satellite orbits, scan-antenna characteristics, and signal-to-noise criteria, suggest that a satellite system should attempt to identify thundery regions rather than individual thunderstorms or lightning strokes. The monitoring problem is further complicated because signals arriving at a satellite originate from several kinds of lightning strokes—intracloud, intercloud, and cloud to ground. Each variety of stroke is different and none is a point source, so the original signals are complex. These signals are then modified during propagation through the ionosphere; the propagation involves both attenuation of the signal and refraction of the signal path. Generally speaking, as the radio frequency at which the satellite sensor operates increases, the source signal strength decreases, the ionospheric complications diminish in importance, the unwanted background noise at the satellite becomes greater, and the practical problems of incorporating an efficient sensor in a satellite are reduced. Obviously these various effects do not all operate in the same direction. Performance cannot be improved simply by increasing frequency; tradeoffs must be made, and the choice is not an easy one.

A satellite sensor operating in the VLF band (3–30 kHz) has the advantage that the source signal is large. Although the attenuation during propagation through the ionosphere is appreciable, VLF atmospherics received above the ionosphere are much stronger than the background noise, even with the small inefficient antennas that must be used at VLF because of space and weight limitations in a satellite.

The greatest objection to employing a satellite sensor at VLF is that the data acquired are not straightforwardly related to the monitoring of thundery areas, because the propagation is by the whistler mode. In this mode, the atmospherics propagate in a guided fashion along a geomagnetic field line.

The geomagnetic control of the path followed by VLF whistler signals imposes severe limits on the detection of thunderstorms by satellite VLF sensors. A VLF atmospheric generated by lightning travels predominantly in the quasi-waveguide formed by the earth and lower ionosphere. However, there is a continual leakage into the whistler mode along the upper boundary of the guide, and since the waveguide propagation involves attenuation, the whistler signals of greatest strength usually traverse the ionosphere near the originating lightning discharge. This means that a VLF signal received in a satellite will most likely have originated in one of the two areas centered around the points where the geomagnetic field line through the satellite intersects the surface of the earth. In practice this means that VLF satellite sensors are likely to be poor detectors of the thunderstorms that abound in the earth's equatorial regions.

In the HF band (3–30 MHz) the source signal from a lightning flash is still of appreciable size, and its attenuation during passage through the ionosphere —at least for frequencies in the upper part of the band—is not severe. Consequently, HF signals from lightning are easily detected above the ionosphere, but the characteristics of the ionosphere limit the ground area that can be monitored by a HF sensor. If a signal at high frequency is to penetrate through the ionosphere, its frequency must be greater than f_c, the ionospheric critical frequency at vertical incidence, which varies with latitude and time of day. A satellite receiving signals at a frequency greater than f_c will be able to scan increasingly large areas on the earth's surface as the satellite altitude and/or signal frequency increase. Although the maximum altitude of a satellite cannot be varied conveniently, a set of satellite receivers tuned to different frequencies in the HF band can be used to sample different-sized areas.

The most obviously attractive frequencies at which to monitor radio signals from global thunderstorms by satellite are 100 MHz and above. These signals are only slightly modified during passage through the ionosphere, and the area that can be viewed is limited only by the tangential cone applicable to the satellite–earth geometry. At the small wavelengths corresponding to frequencies exceeding 100 MHz, moreover, efficient antennas are quite compact and can therefore be easily carried in satellites. The major drawback to using a frequency in these VHF and UHF bands is that the ratio of lightning signal to unwanted noise becomes inconveniently low, especially as the frequency increases. This is illustrated by the data in Table I, after Dr. Pierce, which show that the lightning signal/total noise ratio (the total noise being regarded as the rms sum of the various noise sources) deteriorates over a distance of 1000 km from about 10/1 at 100 MHz to some 2/1 at 600 MHz.

Optical sensors also could of course be carried on satellites. A sensor operating on the optical emission from lightning is perhaps—almost by definition—the most direct indicator of thunderstorm activity. Only line-of-sight paths are involved, and with established photographic techniques it is not difficult to obtain records of high areal resolution covering extensive regions. The greatest problem in using an optical sensor is again that of obtaining an adequate ratio of wanted signal to undesired background, a difficulty that is very pronounced during daylight hours. This problem may be overcome by employing a narrow-bandpass interference filter centered at the hydrogen-alpha (6563 Å) line. This line is a strong emission feature in the lightning spectrum; in sunlight, on the other hand, it is a pronounced feature of the Fraunhöfer absorption spectrum. However, before it can be definitely decided that global lightning activity can be monitored effectively using the Hα emission, several uncertainties must be resolved. It must be demonstrated, for example, that the emission is not excessively absorbed, reflected, or scattered during transmission upward through the thundercloud. Also, in daylight use, the recording film used should be capable of remaining exposed without appreciable fogging for a time of perhaps a minute, if the probability of registering a flash from an active thunderstorm is to be high; and this relatively long exposure time must be shown to be feasible at satellite altitudes. Finally, variations in the natural Hα background must be considered, when a solar flare occurs, for example, Hα becomes a prominent feature in the solar emission spectrum.

Back to earth

Global thunderstorm activity can, of course, be estimated by many other techniques that do not depend on satellites. A common method combines records from a very large number of localities to establish a global thunderstorm census. The classic input for this technique is the statistic, furnished by most meteorological stations, of the occurrence or not of thunder and/or lightning during the observational day. This statistic has obvious deficiencies. It does not indicate, for example, the severity of the activity on a "thunderstorm day"; nor is the area involved well-defined, since the intensity of the thunder generated and of the light emitted vary considerably between individual flashes, whereas the range of audibility of thunder and the atmospheric transmission of light change with local weather conditions. However, thunderstorm-day statistics provided by human observers are still the best indicators we have of worldwide thunder activity.

A direct count of lightning flashes is being increasingly advocated as a replacement for the rather coarse-grained thunderstorm-day statistic. Many designs of instruments intended to count lightning flashes automatically have been produced, and some have been widely used. Most lightning-flash counters are triggered when the electric field variation (or electromagnetic radiation) at the site of the counter exceeds a preset threshold, and they have a "dead time" after triggering comparable to the duration (approaching a second) of a typical lightning discharge. Counters differ principally in the frequency passband to which they respond; various designs operate in the ELF (< 3-kHz) range, at VLF (3–30 kHz) and LF (30–300 kHz), in the HF band (3–30 MHz), and even at the wavelengths of visible light (notably Hα).

Counters operating at these different frequencies differ markedly in their relative sensitivity to intracloud and cloud-to-ground discharges. This can introduce some difficulties when attempting to compile thunderstorm statistics from counter network records.

Another major problem with all lightning flash counters is that their areal coverage is difficult to establish because the signals generated at the source of the lightning flash vary widely in amplitude; nor does the reduction in amplitude as the signals propagate necessarily depend upon distance in a simple manner. The difficulties involved in specifying the area being monitored imply that caution is always necessary before regarding counters as absolute instruments. However, comparative results from lightning-flash counters are undoubtedly of great value, and they certainly have proved their worth as indicators of the relative severity of thunderstorm days. As more counter data are obtained, moreover, there is increasing evidence of a consistent and confidence-inspiring relationship between thunderstorm-day statistics and counter results.

There are many other ground-based techniques that purport to measure, locally, a parameter associated with the thundery activity over a very wide area, an area that may even include the entire earth.

If the spherical-condenser hypothesis is valid, for example, it follows that variations in the fair-weather electric field and current should be related to changes in global thundery activity. And much effort has in fact been directed toward identifying this worldwide effect in ground-level records of temporal changes in the electric field and air–earth current. The results have been ambiguous, however, because such records are usually dominated by local influences. However, with synoptic methods like those Dolezalek discusses in his report on the Atmospheric Electricity Ten-Year Program,[6] or that Vonnegut proposes to use, as discussed earlier, it may be quite feasible to measure reliably and continuously the potential difference between the ground and an altitude of perhaps 20 km. At heights of this order, variations in the potential difference should reflect fairly closely the variations in world thunderstorm intensity. Again it remains to be demonstrated, however, using realistic profiles of atmospheric conductivity, how rapidly these variations in the potential difference will respond to the development of a distant thunderstorm, to what extent the response time depends on distance, and to what degree the results are influenced by changes—as from day to night—in the atmospheric conductivity profiles. Furthermore, it has never been convincingly shown that alterations in the earth–ionosphere potential difference due to an individual storm are directly related to the electrical intensity—as measured by lightning occurrence—of that particular storm.

I. Comparison of lightning and noise field strengths (μV/m for bandwidth of 10 kHz) at 1000-km distance

	Frequency, MHz		
Type of Signal	100	300	600
Lightning	6	2	1
Noise			
Cosmic	0.4	0.3	0.2
Man-made	0.2	0.15	0.1
Thermal (earth)	0.03	0.1	0.2
Solar	0.06	0.06	0.06
Receiver	0.4	0.4	0.4

Schumann-resonance research may help

The so-called Schumann resonances observed in the intensity of the natural electric field in the earth–ionosphere cavity are a subject of considerable current research; Dr. Pierce thinks these could be turned to good use in thunderstorm surveillance. Since the excitation of these resonances depends directly upon the occurrence of lightning, variations in their intensity can in theory be used to estimate global thunderstorm activity. However, here too there are substantial complications. If the earth and the ionosphere were both perfect conductors and the intervening atmosphere a perfect insulator, the resonant frequencies f_m would be defined by

$$\frac{c}{2\pi a}\sqrt{m(m+1)}$$

where c is the velocity of light, a the radius of the earth, and m the order of the resonance. In this ideal situation, the intensity distribution of the resonant fields over the earth's surface would depend rather straightforwardly upon the Legendre polynomials $P_m \cos \Theta$, where Θ is the angular separation between the observing site and the lightning source. In actuality the earth and, especially, the ionosphere are not perfect conductors, and the intervening atmosphere is not a perfect insulator; the resonator characteristics consequently depart from the idealized case. Other difficulties arise because the intensity distribution of resonant fields has to be integrated over many sources (i.e., over many different values of Θ); the departure of the cavity characteristics from those of a perfect resonator could be better understood; the effects of nonuniformity within the cavity need further investigation; and the relative importance of cloud-to-ground and intracloud flashes varies somewhat with the order of the resonance. Perhaps the chief problem, however, is psychological. Up to now most Schumann-resonance investigators in trying to explain their measurements have tended to accept the global thunderstorm pattern as established, and they have concentrated their efforts on deducing the resonator characteristics from this pattern. There have been few attempts to work in the opposite direction. Such attempts would be most worthwhile.

In concluding his excellent survey, upon which I have drawn so freely (with his kind permission), Dr. Pierce stressed that most of the suggested methods involve a chain of significant factors—source effects, propagation influences, technical and cost considerations—and he urged that every link in the chain of factors be investigated in comparable depth. He especially emphasized that not every link in the chain was a technical one. "I believe," he said, "that we may legitimately examine questions in this session which are not strictly scientific. Besides comparing for instance the scientific effectiveness of various monitoring methods, we should also consider their relative costs and chances of realizability; this latter factor is political in the broadest sense."

Many factors that connect weather modification programs to the professional interests of the electrical community, and to the social interests of the human community, are also political. The pollution problem (Fig. 3), which we shall discuss in detail in the next installment, is a prime example.

Gathering momentum and maybe muscle

The studies envisioned by the international committee of scientists involved in planning the Ten-Year Program in Atmospheric Electricity should eventually contribute significantly to the construction of a more complete and tenable theory for the total cloud electrification process. But this elusive electrical theory is not all that's needed to move weather modification programs ahead. The theoretical base for man's increasingly widespread attempts to mold the weather to his will is still far from adequate in its meteorological aspects. But this base is sufficiently large and credible and it has been growing with such vigor in recent years that the American Meteorological Society, after having maintained a hands-off attitude for many years, officially sponsored a First National Conference on Weather Modification, which ran for three days last spring at the State University of New York at Albany. The apparent coming-of-age of weather modification, and especially of attempts to influence precipitation processes, was again demonstrated at an International Conference on Cloud Physics,[9] which ran for a week late last summer at the University of Toronto. No less than 15 papers in the session on "Electrification of Clouds and Precipitation," 22 papers in the session on "Weather Modification," and scores of additional papers dealing with closely related topics were offered at this meeting, which was sponsored jointly by the International Association of Meteorology and Atmospheric Physics (of the International Union of Geodesy and Geophysics), the World Meteorological Organization (of the United Nations), the American and Canadian Meteorological Societies, and the National Research Council of Canada. Weather modification, once a dream of madmen or a scheme of charlatans, has become both big-time and respectable. How did this happen?

The work reported at these conferences is an outgrowth of significant commitments to efforts to modify the weather that have been building slowly, but surely, in the United States and in several other countries, for a dozen years or more. Some of these efforts, especially those in the United States and the Soviet Union, may now be near the fabled "takeoff" point. But research and operational teams that have been built are still mostly subcritical in size; testable but still tentative theories that have been constructed cry out for more definitive data; too much needed experimental equipment is still built on a homemade basis and assembled on an *ad hoc* basis; critical decisions on funding, management, and responsibility—at least in the United States—have yet to be made.

The extent of any future commitment made to weather modification in the United States will depend in part, of course, on how sanguine the public and its Congressional representatives feel about the prospects for eventual payoff. Attempts to assess the public, much less the Congressional, mood in the early months of a new national administration beset by competing priorities are risky in the extreme. Nevertheless, the prospects for continued solid, maybe even spectacular, growth in this area seem good. Strangely enough, perhaps, this is because the knowledgeable scientists and politicians (it does take both, you know) are less wildly optimistic, and therefore more appreciative, of both the problems and the real possibilities than they were 23 years ago, when the whole business began.

Whence weather modification?

Most efforts to change the weather, both in the United States and in other countries, seek to exploit variations of the cloud-seeding techniques pioneered back in 1946 by Irving Langmuir, Vincent Schaefer, Bernard Vonnegut, and their colleagues, who were all then at the General Electric Company in Schenectady. This group, as some readers may recall, seeded a subfreezing stratus cloud in upstate New York with dry-ice (carbon dioxide) crystals dropped from an airplane, and thereby intentionally produced the first man-made snowfall on record. Shortly afterward they repeated the trick with silver iodide crystals. The snow that fell each time was a mere dusting, not nearly enough to encourage a weekend skier to start packing. But it was enough to set off a worldwide wave of optimism about rainmaking, snowmaking, and other still more highly spectacular—and speculative—weather control possibilities. (Remember all those deserts that were going to bloom?) This wave of optimism receded rather quickly, of course, as such speculation encountered the real scientific difficulties.

This period of disillusionment persisted well into the late 1950s. A stock phrase overheard at scientific meetings during this period was, "What ever happened to weather control?" During this period of public disillusionment and scientific disenchantment, weather modification efforts were carried on almost exclusively by a small coterie of meteorologists who had turned their attention to commercial cloud seeding—the "rainmakers." Very few scientists at established institutions chose to identify themselves as being involved in work on weather modification; indeed, support would have been hard to find for work so identified. But the basic research and observations on weather processes that many of them carried forward of course proved to have been both relevant and necessary when weather modification found its way back into favor.

This turnabout in opinion came about gradually in the late 1950s and early 1960s. How it happened would make a fascinating story, one that might be called "The Promotional Psychopolitics of Big Science." We will pursue this story no further here, except to note that counterparts to it could be drawn

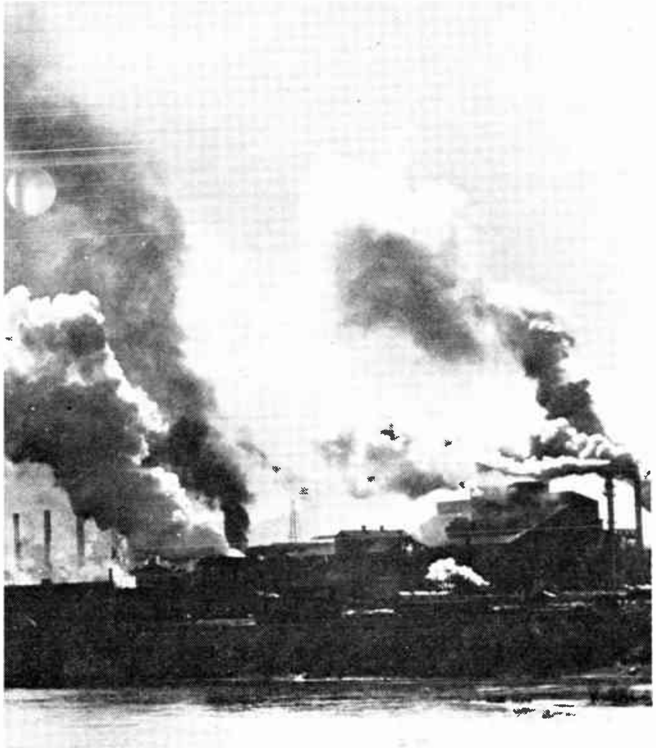


FIGURE 3. Pollution may vitiate weather control work by causing large-scale and irreversible changes in climate.

from a dozen fields—controlled thermonuclear fusion and the desalting of seawater among them. Needless to say, the resurrection of hope for weather modification was also based on some significant technical developments.

The advent of larger-memory, higher-speed digital computers was probably the most important of these. These computers held out the hope upon which GARP is based—that with still bigger computers it might become possible someday to mathematically model systems as complex as clouds (as complex, indeed, as the entire atmosphere) with reasonable fidelity and speed. This day is fast approaching, according to some of the more enthusiastic advocates of weather modification programs, and the proponents of such programs as GARP. Their enthusiasm is commendable, and perhaps justified, but I am reminded of a comment Dr. Pierce made in his Tokyo paper last year. Though the context was different the words seem apt: ". . . we should be suspicious when we hear the word 'model.' Many problems in geophysics are too complicated to be solved mathematically even by modern computer methods. The fashion has therefore developed of representing such a problem by a model which is mathematically tractable and yields a computer solution. However, particularly if the computer result contains an enormous number of apparently significant figures, we should be extremely cautious before deciding that our solution is physically realistic."

Hope that the search for physical realism could be prosecuted as intensively as the search for computer solutions, however, also rose sharply among meteorologists as the decade of the '60s opened, because along with the computer came the new or much improved observational platforms upon which GARP plans depend so heavily—satellites, sounding rockets, balloons, high-altitude aircraft, and plans for ocean-wide buoy systems. These platforms, carrying elaborate new sounding instruments and remote sensors of various kinds, and linked by suitable telemetry and data-processing ties to the supercomputers now being designed, indeed promise to provide the huge amounts of data needed to improve computer-based mathematical models of complex atmospheric processes.

In addition to these developments, the meteorologists engaged in commercial cloud seeding were able to point to a surprisingly large backlog of rainmaking attempts they had carried out during the dark, discredited days of the 1950s. Some of these commercial cloud-seeding efforts apparently did produce rain that would not have fallen naturally; at least they withstood an intense barrage of critical statistical opinion that rejected the results obtained in most commercial cloud-seeding programs as unlikely or unprovable.

These advances, together with some adroit maneuvering by the then-limited numbers of weather modification proponents in the climate favorable to bold technological initiatives provided by the Kennedy administration, have made weather modification both respectable and at least passably supported since.

Current status of weather modification

Thomas Malone, ex-head of the Committee on Atmospheric Sciences of the National Academy of Sciences, "at the risk of oversimplification and with the usual caveats about fallibility," recently summarized the present position and offered some judgmental evaluations concerning the outlook for weather modification[10]:

1. Field results have shown unequivocally that several cubic kilometers of supercooled clouds (comprised of liquid droplets at temperatures below freezing) can be transformed into ice-crystal clouds by seeding with appropriate chemicals.

2. Dissipation of supercooled fogs and clouds over an airfield runway is feasible (see introductory figure) and has been used to good economic advantage by airlines in the United States and in other countries.

3. Recent experimentation in clearing certain types of warm fog (droplets at temperatures above freezing) over airports is beginning to produce modestly encouraging results. A breakthrough in this type of highly localized weather control may well be imminent and is likely to be realized in a matter of years.

4. Relatively little serious attention has been given to conscious interference in the processes at the interface between the atmosphere and the underlying surface beyond demonstrating that it is possible to inhibit evaporation from water surfaces and vegetation. If, as now appears likely, these interface processes turn out to be important to large-scale modification of the climate, there is a high probability that the technology could go through explosive development during the period 1975 to 1995.

5. Persuasive although by no means conclusive evidence suggests that rainfall can be increased through cloud seeding by around 10 percent, depending upon the meteorological conditions. There is a

high probability that residual ambiguity will be resolved by 1975, and a further high probability that by 1980 naturally occurring rainfall can be either augmented or diminished locally by proven techniques. The probability also is high that by 1990 rainfall several hundred kilometers downwind from the site of seeding operations can be increased or decreased at will.

6. There are indications that Soviet scientists have succeeded in reducing hail damage by a factor of three to five by introducing silver iodide directly into susceptible parts of hail-producing clouds. This form of weather modification will probably always remain local, but will develop rapidly over the next decade and there is a high probability that it will begin to be adopted for widespread use by the early 1980s.

7. Physically reasonable approaches to the suppression of lightning (see Fig. 4 for example) have been tried with mixed but, on balance, promising results. Progress on this highly localized modification measure awaits a more satisfactory explanation for the lightning discharge process.

8. Cloud-seeding techniques that are of sufficient merit to warrant field experimentation have been advanced for the modification of hurricanes. The limited tests have not as yet yielded even preliminary conclusions. This approach should be pursued vigorously, Malone feels, but the probability of success is not high—something like 50 percent. New concepts are needed and they are likely to emerge from computer simulation of hurricanes.

9. No technique exists for consciously influencing the large-scale weather patterns, which we describe as climate, and not much progress can be expected until the scientific results of the recently approved Global Atmospheric Research Program begin to become available. Malone thinks there is a better than even chance that we will be able to modify large-scale weather patterns within 50 years.

10. There also is a distinct probability that large-scale climate modification will be effected inadvertently before the power of conscious modification is achieved. Calculations with the relatively crude atmospheric models at hand suggest that the 10 to 15 percent increase since 1900 in the minute amount of carbon dioxide in the atmosphere has caused surface temperatures to rise 0.2°C, whereas temperatures in the stratosphere may have decreased ten times as much. It should be possible by 1980 to predict with precision the effects on the atmosphere of the continued consumption of carboniferous fuel. There is a small probability that these effects will not be tolerable. Air pollution may have already extended its influence beyond the urban domain. Contamination of the upper atmosphere by rocket exhaust is a problem that may be of practical importance sooner than we realize. Finally, agricultural cultivation and urbanization are transforming the nature of the surface underlying the atmosphere on a large-scale basis, with possibly important climatic consequences that we should also be able to assess during the 1980s.

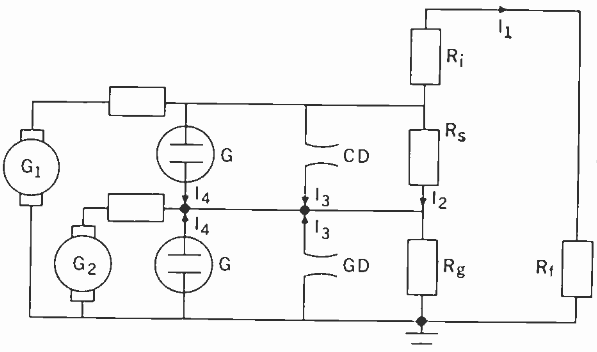


FIGURE 4. Heinz W. Kasemir of the ESSA Research Laboratories has used this simple model of a thunderstorm as the basis for his experimental effort to suppress lightning by seeding thunderclouds with chaff fibers. G and G_2 = generators of positive and negative charge in storm; unmarked resistors at hot terminals of G_1 and G_2 indicate that charge generators are current generators—they produce charge at a constant rate whether it can dissipate or not. Charge can dissipate in three ways: through glow lamps G representing corona discharge of chaff fibers; through spark gaps CD and GD, which represent, respectively, cloud-level and cloud-to-ground lightning discharges; or through ohmic resistors R_s and R_g in parallel to R_i and R_f. The ohmic resistors represent the conduction currents of a thunderstorm. R_s = resistance of air between positive and negative charge centers in the storm; R_g = resistance of air between cloud base and ground; R_i = resistance of air between top of storm and the ionosphere; R_f = resistance of air between ionosphere and earth in fair-weather areas. With exception of R_f, these resistors have comparatively high values (of the order of hundreds of megohms) so that charge produced by the generators cannot leak away very fast. As a consequence, voltage on terminals builds up to high values until a "lightning" spark is ignited through gaps CD and/or GD, unless the ignition voltage of the glow lamps (threshold field for corona discharge on the introduced chaff fibers) is lower than the breakdown voltage of the spark gaps.

Who's paying the bills?

If we exclude the simulation studies mentioned by Malone, it is fair to say, by way of summary and review, that modern weather modification efforts are based almost exclusively on cloud-seeding techniques. They include not only attempts to wring rain or snow from clouds that are reluctant to part with the water they contain, but other efforts aimed at suppressing lightning strikes to aircraft and cloud-to-ground lightning strokes of the long-continuing-current variety that cause 90 percent of forest fires in the western United States. Other weather modification efforts seek, with almost routine success when the air temperature is below the freezing point of water, to clear fog from airport runways. Still other cloud-seeding efforts seek—also with quite successful results reported from the Soviet Union—to diminish the barrage of large hailstones often released by severe thunderstorms. This is a seemingly attractive weather modification target because in the United States alone hail damage to crops amounts to $200 million per year. On even more ambitious and long-range scales, other weather modification programs being carried on under federal sponsorship purport to be seeking ways, ultimately, to curb, deflect, or dissipate harmlessly the awesome destructive power of hurricanes and tornadoes (Fig. 5).

The location of seeding-based weather modification projects carried out in the continental, contiguous United States during Fiscal Year 1967, the latest year for which such information has been assembled, is shown in Fig. 6.

Support for such a broad array of weather modification goals has come from diverse sources. Members of the meteorological profession have been in the lead, of course. Through the relatively bleak decade of the 1950s their precipitation-enhancement efforts received continued support and encouragement from congressmen, ranchers, and farmers from the western states, where chronic aridity was an unwelcome way of life. To this base there was added a surge of support from easterners who suddenly grew more appreciative of the fickleness of nature during the unprecedented drought that afflicted the northeastern states through the mid-1960s. This developing stream gained further strength in the halls of Congress and among scientific advisory bodies to the President when federal agencies whose defined missions were weather-sensitive to a greater or lesser degree also pressed for the development of at least modest exploratory programs related to the execution of their missions.

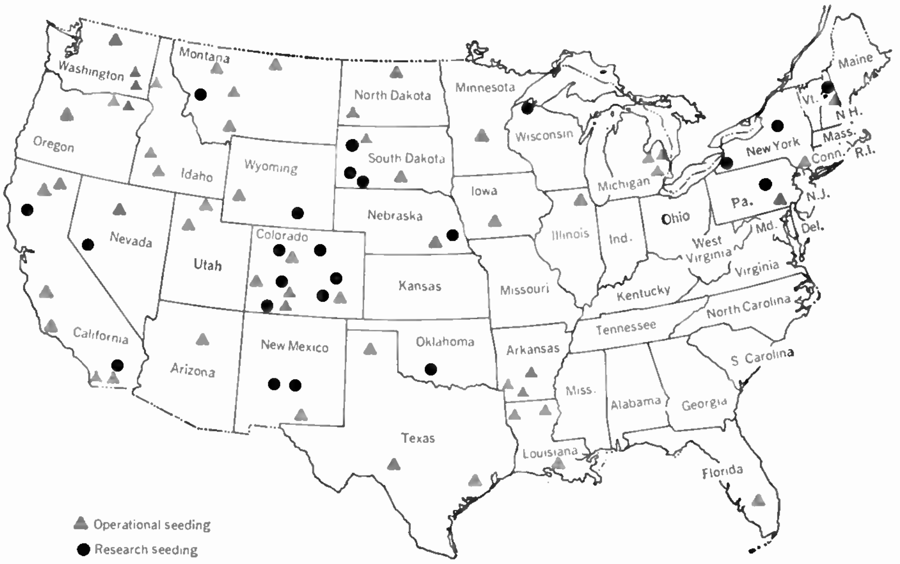
Thus the Federal Aviation Agency and the U.S. Air Force in recent years have supported research in airport fog control; the Forest Service of the Department of Agriculture has led the way in lightning suppression efforts in forested country, and the U.S. Army Electronics Command and other units of the Defense Department have concentrated on research nominally aimed at minimizing lightning strikes to aircraft and other military installations. In more recent years a relatively new agency housed in the Commerce Department—the Environmental Science Services Administration, known to the initiated (thankfully) by the abbreviation ESSA—has cooperated with the Army in these lightning suppression studies. ESSA, which includes the Weather Bureau, also is pursuing weather modification efforts over a much broader front, which includes basic research, studies of severe thunderstorms and tornadoes, and hurricane control studies.

Until last fall when the nature of its overall mandate was changed by Congress, the National Science Foundation was the agency responsible by statute for developing a national program in weather modification and for reporting on activities in the field. Even with its new responsibilities, which entail a greater concentration on applied research and the social sciences than heretofore, NSF of course will continue to be a chief source of support for university-based research and education in fields that can contribute to weather modification.

Early in the 1960s the Bureau of Reclamation in the Department of the Interior picked up the rainmaking ball that nobody else then seemed to want. It has since developed strong plans and begun operational programs in the West—with the considerable encouragement of high-seniority western congressmen—that are aimed at increasing rain and winter snowpack and thereby augmenting the runoff of water into the dam-and-reservoir systems built and operated by the Bureau. This particular element of the United States' not yet truly national weather modification

FIGURE 5. ESSA research programs aimed at predicting the occurrence of tornadoes like this one and, ultimately, at perhaps controlling them, are beginning to pay attention to the electrical parameters as well as the meteorological attributes of these intense vortices. A good deal of work is going into recording and analysis of the radio-frequency signals—so-called atmospherics, usually abbreviated as sferics—coming from lightning in and around the storm. The meteorological conditions that favor their development are also becoming better known.

FIGURE 6. This map shows the location of seedling-based weather modification projects carried out in the continental, contiguous United States during Fiscal Year 1967.



program could prove to be especially important to the hydroelectric power industry.

Rainmaking for hydropower looks rosy . . . but

Dr. James Crutchfield, an economics professor at the University of Washington, Seattle, and long-time consultant to the federal government, in an otherwise rather bearish recent assessment of the economics of weather modification,[11] thought that the improvements in efficiency of the hydroelectric portions of power systems that might be achieved through control of precipitation in the relevant watershed could make this kind of weather modification an attractive economic possibility. Successful management of the water entering a watershed could increase the net economic productivity of hydroelectric utility systems by permitting more efficient use of storage capacity; it could also facilitate a favorable change in the ratio of peak-to-average plant capacity and thus further reduce the overall capital intensity of the complex. In coming to these conclusions, Crutchfield cited estimates made by Eberly [12] in a Pacific Gas and Electric Company test program in precipitation enhancement. This study indicated that the increase in average runoff required to recover all costs of a cloud-seeding project might range from less than two percent to ten percent or more. In a significant number of cases, however, the break-even point would be reached if runoff could be increased by amounts lying in the lower part of this range. Just how much the runoff would have to be increased in any particular situation would depend on the storage capacity of reservoirs in the system, the characteristics of the runoff, and the size of the watershed.

It is important to remember, however, that increases in average runoff are not the same as increases in precipitation; runoff is invariably much less because some precipitation is either reevaporated directly or transpired by plants and some infiltrates underground. Precipitation increase—*not* runoff increases—of the order of ten percent apparently may be produced by ground-based silver iodide seeding of winter storms in areas where warm, moist air masses are forced to ascend mountainous barriers, as on the Pacific-facing slopes of the western U.S. Though this suggests that runoff increases in the lower part of the < 2 to > 10 percent range cited by Eberly may be attainable, a caveat is necessary. Under the somewhat different meteorological conditions characteristic of the eastern and midwestern U.S., silver iodide seeding also appeared to increase rainfall by about 10–20 percent in the seeded areas. But follow-on studies in these regions have shown that increases in precipitation comparable to those induced in the seeded area may also occur up to 240 km downwind of the nominal target area. This water may not fill your reservoir but someone else's cellar. Still more disturbing to contemplate are statistically randomized cumulus seeding experiments in Missouri, which indicate, tentatively, that although perhaps 5 to 10 percent more rain may indeed form in the clouds in the target area, comparable or greater decreases in rain reaching the ground may occur as a "shadow" effect for a long distance downwind from the target area. This could also increase the litigation load in your legal department at a faster rate than it increases your water supply.

The meteorological results are clearly not all in; and the manifold legal problems connected with rainmaking—even for research purposes—are just coming under systematic investigation.[13] In spite of these uncertainties and potential hazards, programs devoted principally to increasing or redistributing rainfall and other forms of precipitation have developed rapidly in the Soviet Union, Australia, Japan, Israel, India, France and Germany, as well as in the U.S.

Where do we go from here? And why?

In the United States the pace and scale of federally sponsored and coordinated efforts in weather modification can be expected to accelerate sharply—from an annual funding level of about $10 million in Fiscal 1967 to a funding level of $90 million or more in Fiscal 1970, or more likely 1971 or 1972—if the new Administration and Congress heed the recommendations contained in the so-called Newell Report.[14] Authored by Homer Newell, who at the time (1966) was associate administrator for space science and applications of the National Aeronautics and Space Administration (an agency not conspicuously involved in weather modification per se), and issued under the auspices of the Federal Council for Science and Technology in the Executive Office of the President, the report recommends a much more vigorous and highly coordinated national program. It also discusses administrative, scientific, budgetary, facilities, and educational elements of the program it proposes.

Science administrators, even those with no visible bureaucratic axe to grind, may propose, but the President and Congress dispose. The Newell Report, however, will probably figure prominently in any decisions that may be made, if only because it is one of the most recent and comprehensive of nearly a dozen separate studies of the weather modification effort that have been conducted since 1960 by federal agencies, think tanks, advisory groups, and Congressional committees and staffs.

Decisions about the pace and direction of the weather modification program cannot be postponed much longer. The exhortatory reasons usually offered in support of such an assertion include a number of familiar arguments: we are running out of fresh water; losses of life and property in hurricanes, hailstorms, and tornadoes are intolerable; and so on. In the United States, at least, such assertions cannot be defended on economic grounds, at least not in view of the present rather incipient state of the art. Professor Crutchfield dampened audience spirits considerably at the First National Conference on Weather Modification in Albany a year ago, by sounding this essentially negative note in his assessment of the economics of weather modification. In attempting to ease the impact of his assessment, however, he may have inadvertently touched upon the most compelling reason for proceeding with weather modification research on an urgent basis. He said, "I should like to make it as clear as possible that these words of caution are not an argument against expanded support for research in weather processes and the physical aspects of weather modification, nor do they reflect any lack of appreciation of the discontinuities in research that may require large-scale efforts if any discernible results are to be achieved and identified. I do feel that we are on safer and more honest grounds in arguing that the principal output of such research, for some time to come, will be scientific knowledge rather than direct economic benefits."

In the opinion of many environmental scientists, the need for such scientific knowledge is precisely the most urgent reason for sharply expanding the national commitment in this area. Their sense of urgency is engendered by fragmentary but growing evidence that suggests that by polluting the lower atmosphere as extensively as we have, we may already be engaged in a massive, uncontrolled experiment in weather modification on a global scale. The outcome of this experiment is not yet predictable (as Thomas Malone noted), but there are no early indications that it will be pleasant. This inadvertent experiment may turn out to have such serious and perhaps irreversible geophysical—and therefore social—consequences, they argue, that it behooves us to find out precisely what we are doing to natural weather systems as soon as we can.

The climate changes that may already have been set in motion by pollution could have equally serious implications for the electric power industry. On a less dramatic and less speculative scale, though one no less laden with implications for the electrical community, the possibility exists that by unintentionally overseeding the lower atmosphere with pollution particles, we may not only be inhibiting natural precipitation but undercutting future attempts to alter cloud behavior and precipitation processes by conventional cloud-seeding techniques. If these fears turn out to be valid, and they are being investigated in a program sponsored jointly by ESSA and NSF, the ante for developing effective electrical techniques of cloud modification could be raised considerably. We shall discuss these matters in the next, and final, installment of this article. We shall also explore the professional implications of the fact that the research needed to transform weather modification from a national goal into an operational technology is not the same as that needed to use such a technology wisely.

REFERENCES

1. "Weather modification, ninth annual report, 1967," National Science Foundation, Publication No. 68-21, U.S. Government Printing Office, Washington, D.C., 1968.

2. "Weather and climate modification, problems and prospects," two vols., National Academy of Sciences–National Research Council, Publication No. 1350, Washington, D.C., 1966.

3. Blanchard, D., "The electrification of the atmosphere by particles from bubbles in the sea," in *Progress in Oceanography*, vol. 1, M. Sears, ed. New York: Pergamon, 1963.

4. Imyanitov, I. M., and Chubarina, E. V., *Electricity of the Free Atmosphere*. Jerusalem: Israel Program for Scientific Translations, 1967. (Available from the U.S. Department of Commerce, Clearinghouse for Federal Scientific and Technical Information, Springfield, Va. 22151.)

5. Pierce, E. T., and Whitson, A. L., reply to correspondence, *J. Atmospheric Sci.*, vol. 23, pp. 452-453, July 1966.

6. Dolezalek, H., "Discussion of an Atmospheric Electricity Ten-year Program" Tech. Rept. D-2, Contract Nonr N00014-66-C0303, Boston College, Chestnut Hill, Mass., June 1967.

7. Pierce, E. T., "The monitoring of global thunderstorm activity," in *Proceedings of the Fourth International Conference on Univeral Aspects of Atmospheric Electricity, Tokyo, May 1968*, S. C. Coroniti and J. Hughes, eds. New York: Gordon and Breach, in press.

8. Horner, F., "Radio noise from thunderstorms," in *Advances in Radio Research*, vol. 2, J. A. Saxton, ed. New York: Academic Press, 1964.

9. *Proceedings of the International Conference on Cloud Physics, August 26 30, 1968, Toronto, Canada.* Boston: American Meterological Society, 1968.

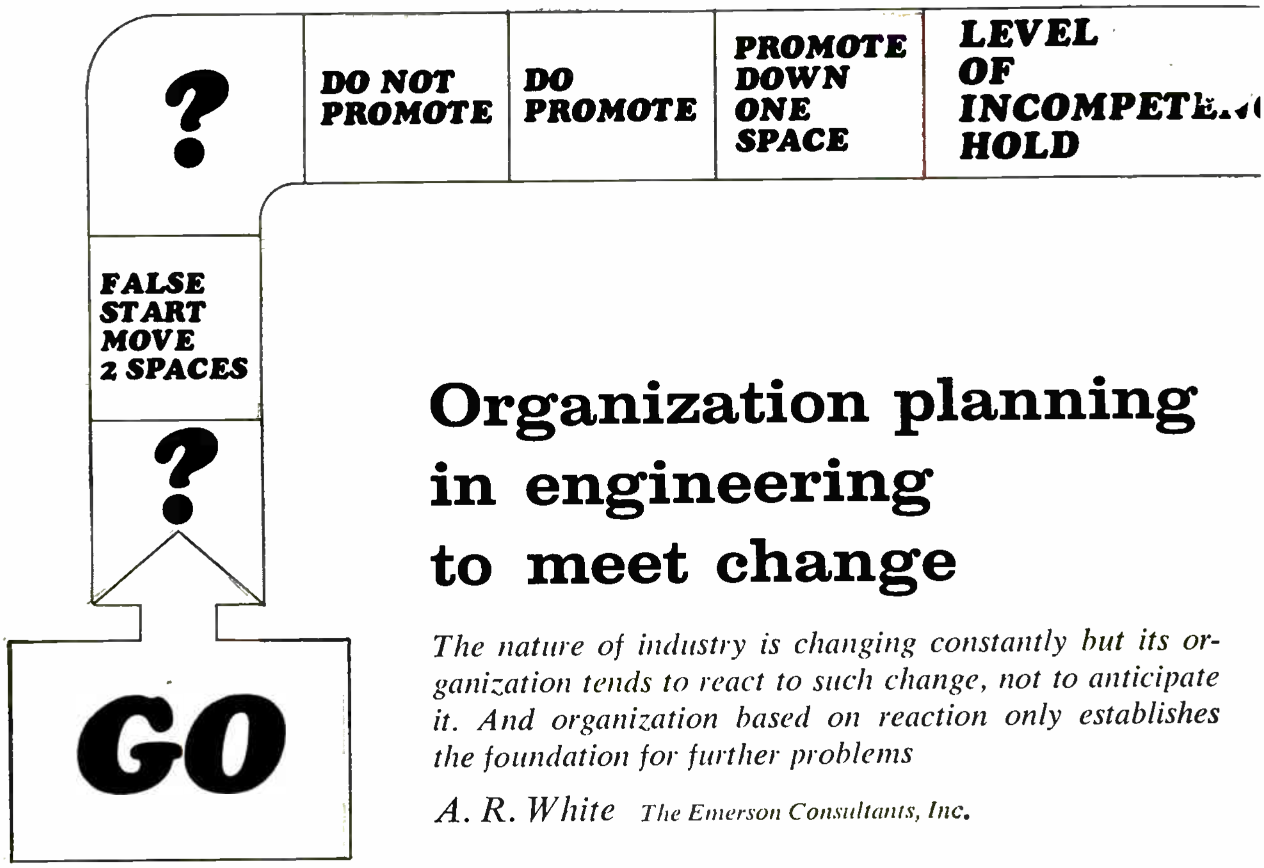
10. Malone, T. F., "New dimensions of international cooperation in weather analysis and prediction," *Bull. Am. Meteorol. Soc.*, vol. 49, pp. 1134–1139, Dec. 1968.

11. Crutchfield, J. A., "The economics of weather modification," in *Proceedings of the First National Conference on Weather Modification, April 28–May 1, 1968, Albany, N.Y.* Boston: American Meteorological Society, 1968, pp. 181–189.

12. Eberly, D., "Weather modification and the operations of an electric power utility," in *Human Dimensions of Weather Modification*, W. R. D. Sewell, ed. Chicago: University of Chicago Press, 1966, pp. 209–226.

13. Taubenfeld, H. J., and Taubenfeld, R. F., "The law and weather modification," in *Proceedings of the First National Conference on Weather Modification, April 28–May 1, 1968, Albany, N.Y.* Boston: American Meteorological Society, 1968, pp. 190–201.

14. Newell, H. E., "A recommended national program in weather modification," Federal Council for Science and Technology, Interdepartmental Committee for Atmospheric Sciences, Rept. 10a, Washington, D.C., Nov. 1966.



Organization planning in engineering to meet change

The nature of industry is changing constantly but its organization tends to react to such change, not to anticipate it. And organization based on reaction only establishes the foundation for further problems

A. R. White *The Emerson Consultants, Inc.*

During the early part of the century various precepts for effective industrial management were set up that, although still basically valid in many respects, are being challenged, either in principle or in application. It is now appropriate to ask whether organizational change can any longer be evolutionary. This article points out the need for organization that anticipates change, not reacts to it. Specifically, it discusses the misuse of the engineer in the current scheme of things and suggests ways in which the situation can be remedied.

One hundred years ago Harrington Emerson,[1] founder of the consulting firm that still bears his name, witnessed the conquest of France in 45 days by the Prussian general, von Moltke. He wrote:

"It is not the pomp and glory of that campaign that appealed to me as I intimately and personally . . . watched it . . . but the calm skill of the play showed me what principles could do when carried into effect by a suitable and competent organization . . . planned by the master organizer of the . . . century. It was not the soldiers, . . . drill or tactics, ... not the German equipment that won the war. . . . It was not the money. It was von Moltke's principles and organization that won. . . . He perceived the theory of a general staff (in which) each topic that may be of use to an army shall be studied to perfection by a separate specialist."

Early in this century Emerson introduced von Moltke's line- and staff-organization theory to U.S. industry, beginning with the Santa Fe Railroad. This concept is still considered by many to be the cornerstone of industrial organizational theory.

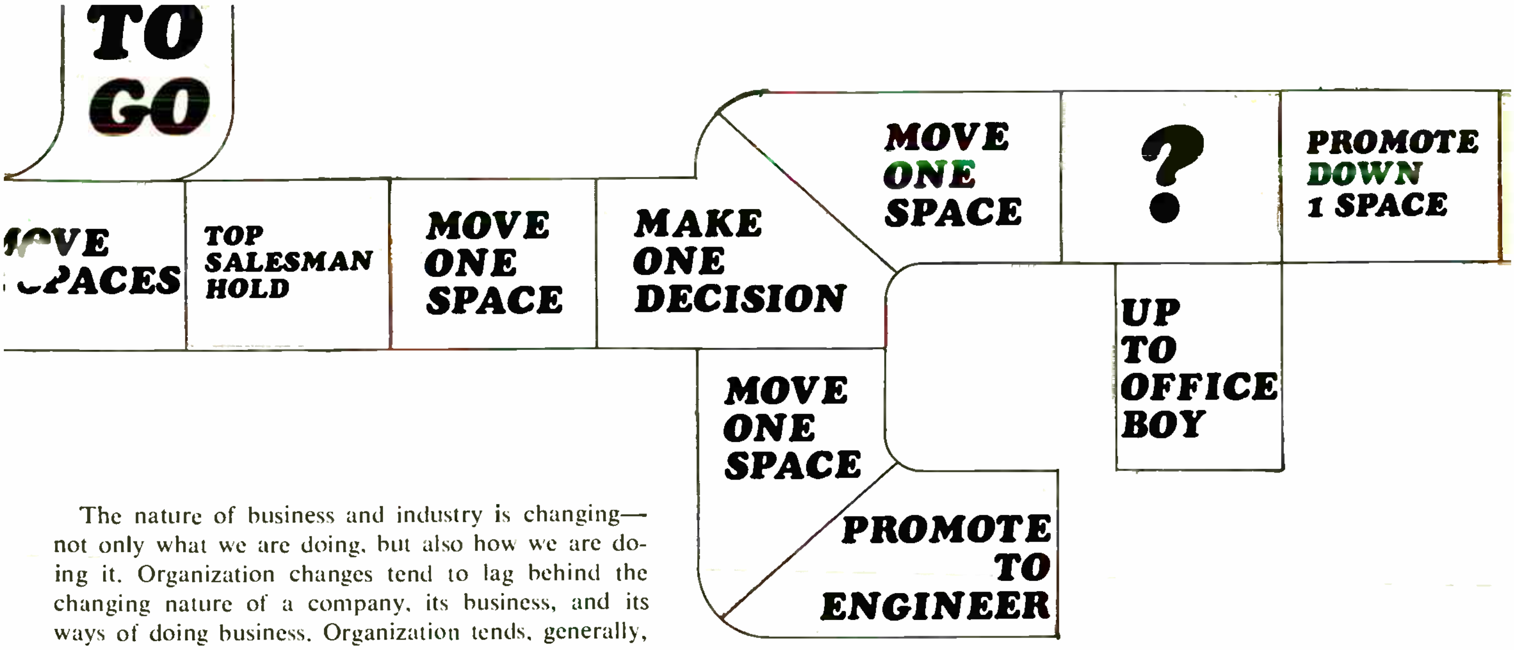
Almost immediately, however, Emerson recognized difficulties. Staffs proliferate. Coordination and cooperation become problems. So do the relationships between line and staff managers in carrying out what each considers his responsibility within his authority.

Emerson and his contemporaries evolved a number of precepts for effective industrial management that have been refined and extended by later generations of management thinkers. Most of these precepts have stood the test of time. But like everything else today, they are being challenged, in principle or application.

An example is the organizational tenet ascribed to Henri Fayol: "Each person in the organization shall have only one boss." But in engineering it is not at all uncommon to work for several bosses—

- If one is a member of a project team, there will be the project manager and one will have a functional supervisor as well.
- There may be a specialist technical manager and an administrative manager.
- There may be a financial manager and an engineering manager.
- If one is a draftsman, there will be the drafting supervisor and the engineer for whom the drafting is being done.

Another example is the "span of control" principle which Peter Drucker questions. Fifty years ago it was stated that "A manager should have no more than six subordinates reporting directly to him." Today, many managers have more—with no apparent problems.



The nature of business and industry is changing—not only what we are doing, but also how we are doing it. Organization changes tend to lag behind the changing nature of a company, its business, and its ways of doing business. Organization tends, generally, to react to change; it does not anticipate change. And organization based upon reaction succeeds only in establishing the foundations for a fresh set of problems for the next generation of managers.

Reaction usually comprises the undertaking of numbers of small, localized organizational changes. In effect, Band-Aids are applied to the corporate body, which may be in need of surgery. But the possibilities of effecting discrete change in one area of our society, or within a corporation, without repercussions occurring throughout a very much wider area, are increasingly remote. A corporation is a system of interrelated activities and people. It is the most precisely assembled system devised by man. It reacts accordingly.

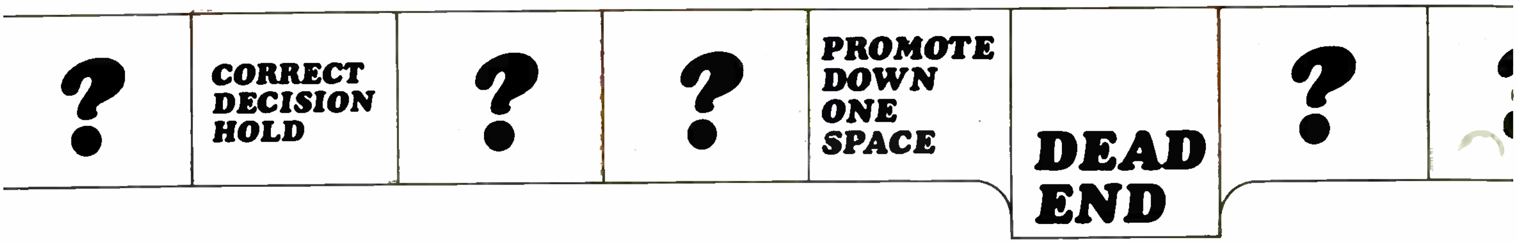
Change is, in fact, a constant of a technological society. When this is not recognized, the results can be failure. In the past 17 years, an estimated 2000 technology-based companies have been founded, of which only one in eight has lasted more than ten years. A study of these failures has shown that the predominant reason they occurred was that the companies had been started to exploit a particular technological change, but their founders failed to understand that their continued existence depended upon their ability to anticipate *continued* technological change, which, in time, would displace present technology. This "predominant" reason obscures the real reason. These firms failed because they were not organized to stay in business.

Piecemeal, patchwork organization changes can cause more problems than they will solve, unless an overview approach is applied, unless more than the obvious ramifications are taken into account. An entire corporate reorganization is not necessary every time the "vendor drawings expediter" function is changed, but such a change does entail a review of influences in a wider area than just the document expediter's job. Nor must organization formations be irrevocably structured, with the organization chart carved in stone. No structure must be considered permanent. Creation of that tradition succeeds only in establishing an atmosphere of resentment against, and resistance to, change. There is too frequently the feeling that an incumbent has a divine right to his position because he has held it for ten years; and because he has never been required to train a replacement, he thinks of himself as irreplaceable. Today we are witness to the growing disenchantment with, and rejection of, organizations that have become self-perpetuating bureaucracies immune to the interests of those who comprise or who are served by them.

Young people are rejecting the gloomy prognosis of George Orwell's 1984. They will not tolerate the mechanistic, totalitarian, and monolithic organization of their lives in college, in government, in society, or in their work. As new engineers and future managers from the post-'68 campuses join industry, this rejection of structural rigidity will grow. They may be willing to start at the bottom, but they are no longer willing to have all decisions made for them or to stay at the bottom very long. Organizational structures that do not consider this are headed for trouble.

It is valid to ask whether organizational change can any longer be evolutionary. To employ a point already made: Organization must anticipate change, not react to it. And the question is: Are we in an evolutionary society or a revolutionary one? Servan-Schreiber says that we are in what amounts to a permanent industrial revolution. The chief executive must recognize the possibility and include this in his thinking whether or not he accepts the dictum as fact.

Corporate organization planning must have time goals. Changes must be planned to achieve long-term organization objectives in pace with, and as part of, the long-term objectives of the corporation itself. As corporate objectives change, so must organizational objectives. The training and development of potential managers must be planned in anticipation of these organizational changes. An all too common problem in developing organizations is the lack in the engineering department of special technological talents other than engineering. When a new organization plan is proposed, there is no one who can fill vital positions in, for example, systems planning, work scheduling, engineering project management, critical path scheduling, or mathematical model building. An equally common problem encountered in organization studies is the similarity of age among top management. Frequently, all the top echelon is due for



retirement within two or three years and often there is virtually nobody available and able to take over. In both cases the results are crash training programs, personnel brought in from outside (which can have a very bad effect on present staff morale), or promotion of people into jobs for which they are not fit.

Dr. Laurence Peter has an interesting theory about this, as quoted in the *Wall Street Journal*[2]:

"In each hierarchy, whether it be government, business, etc., each employe tends to rise to his level of incompetence. Every post tends to be occupied by an employe incompetent to execute its duties.

"Dr. Peter's theory is that each employe is promoted up the rungs of the organization ladder until he reaches a position for which he is incompetent to fulfill his duties. He then, of course, is not promoted any further, but neither is he removed. He remains for the rest of his career in a job which he is not capable of handling.

"Dr. Peter notes that some critics have mistakenly seen his theory as placing in the hands of lower-level employes a means of ridiculing their superiors.

"Not so, says Dr. Peter. The lowest rank of every bureaucracy or hierarchy must contain its own share of incompetents. True, he says, some competent people are being added at this lower level through recruitment, but they are being rapidly drawn off by promotion. Only the incompetent remain.

"How is it, then, that any work is done at all? Dr. Peter explains that the work is done by people who have not yet attained final placement at their level of incompetence.

"Also, it may be wondered why we occasionally see a competent person at the very top of a hierarchy. Dr. Peter says it is simply because there are not enough ranks for him to have reached before he finds his level of incompetence.

"As a rule, however, says Dr. Peter, this paragon of competence eventually sidesteps into another bureaucracy—from business to government, from the armed forces into industry, from law to politics—and there finds his level of incompetence."

As business grows more complex, as higher degrees of specialization are needed, as Harrington Emerson's problem of coordination becomes more acute, it becomes increasingly important to recognize the need to plan for an organization trained and capable of handling the sophistication of today's and tomorrow's business. And the people best suited to these techniques are not necessarily engineers; they may be systems analysts, economists, mathematicians, logicians. Whether engineers like it or not, these new technicians must be taken in as partners in the engineering function.

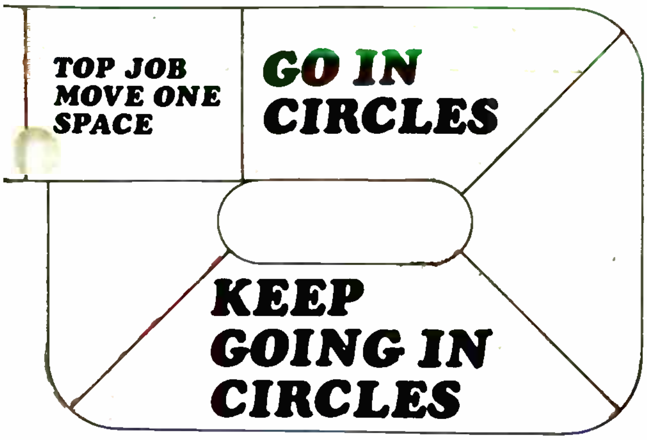
Frequently, these nonengineers are grudgingly accepted into the organization, but not on an equal status. It is almost inconceivable to some managers of engineers that anyone but an engineer can really be trusted to do a job—any job. So we see engineers as office boys, purchasing clerks, and dispatchers.

Concomitant with this misuse of engineers is the problem of recognition of status—or advancement. Historically, promotion has been up the hierarchic ladder of the company. This works pretty well in an army or in a predominantly blue-collar organization, but it is likely to fail where highly technical, scientific, or professional employees are concerned. The reason is not hard to find: Professional people judge each other on their professional abilities. They expect to be judged this way.

Scientists, doctors, engineers can be expected to be interested in their work. In many cases their interest in administration is only secondary, if, indeed, any interest exists at all. They consider it a bothersome interference with their professional activities. Nor does the training and education of an engineer normally lead to the development of administrative qualities. As a result, professional people often make poor administrators.

Today, most hospitals and schools are managed not by doctors and teachers, but by professional managers. In engineering, this practice is far less common. Too often a good engineer is promoted to a key administrative position, where his time is largely consumed by administrative affairs. His talents as an engineer are lost to his company, and his administrative abilities may not be adequate.

One frequently finds that engineers are forced to judge their own success by how high up the administrative ladder they are—how many men they command, how resounding their titles appear. One solution adopted in some technologically based companies is to allow engineers and scientists to develop along two possible paths: line, or administrative, progression; or technical progression. Under this system, the engineer has an option. He can be judged as an administrator and be rewarded as such. Alternatively, he can become a better and better engineer in his specialty and be judged this way. The engineer profits, and so does the company, which benefits with regard to manpower planning, training for succession, and recruitment. In effect, planning is possible. There seems to be no serious problem of a highly paid engineer being administered by a perhaps less highly paid manager. A journeyman mechanic usually makes more than a junior supervisor. A top sergeant receives more than a second lieutenant. Board members regularly are paid less than the president.



There is another current problem that the organization planner must consider—the relationship between engineering and the computer in its use as a tool for control. Perhaps the one really new aspect of organizational planning is the recognition that information itself is a product that must be controlled, managed, and handled in no less a degree than production or inventory. It is an assignable function and includes budget planning and controls, systems development and analysis, financial and other model building, manpower planning, organization, procedures, estimating, cost controls, data processing and analysis, long-range planning, scheduling, workload forecasting, performance measurement, and similar activities. Fundamental to information handling is, of course, the computer and its attendant operations research analysts, programmers, key punch operators, and program managers—a whole network of specialties so new that one has yet to hear of a systems analyst reaching retirement.

In recognizing the existence of this information technology, there is also the need to recognize the very real danger that managers of information will become also de facto managers of an entire company, including engineering—simply by censoring information, perhaps unwittingly. This is an emerging problem for which there must be organization planning. We cannot operate without data. But the data must be accurate, useful, and timely—and needed. Information must not be measured by the ream; it is too easy to be almost overwhelmed by computer-generated data that, as far as one can tell, serve no useful purpose. Thus, there is need for specialization in information control: gathering, analysis, cataloging, generating, retrieval. A company-wide information system is essential but it must serve the individual parts of the corporation, not dominate them. And ways must be developed to ensure this.

Inevitably, there will be two organizations in the company, with key personnel playing two roles, perhaps even with two job descriptions. One organization will be managerial in structure; it will be hierarchical in the traditional pyramid shape and its purpose will be to control administration—bookkeeping, office management, salaries, training, promotions, maintenance, and so forth. It might also include sales, customer relations, and other activities.

The second structure will be functional. Within this structure, engineering project work will be undertaken, task forces will do research, systems analysis will be performed, and information will be managed.

For each organization there will be distinct channels of communication because one can no more compel unnatural lines of communication than one can expect water to flow uphill. By recognizing these two structures, every man can contribute to his best ability. An engineer no longer need be restricted by unreasonable lines of authority designed for an entirely different purpose, that of administration.

Recognition of this dual organizational reality will do much to alleviate the suspicion and dislike of large bureaucratic organizations felt by so many of today's younger engineers. It provides the sense of participation and personal involvement that these engineers often demand in return for their dedication and loyalty. It provides a great sense of satisfaction to the engineer and to the nonengineer technician as well.

The combination of dual ladders for individual progress and the recognition that one of them, the traditional organizational format, is primarily for administrative purposes—not functional, informational purposes—almost guarantees that engineers and technicians will be used to their maximum abilities.

The experience of management consultants has shown that there are few organization structures that cannot be improved. This is true for many reasons, not the least of which is that organizations are frequently created to fit personalities and conditions and then are distorted as personalities shift and conditions change. John W. Gardner, former Secretary of Health, Education and Welfare, said recently: "Most organizations have a structure that was designed to solve problems that no longer exist." Placing names in order of seniority on a piece of paper and connecting them by lines is not organization, nor is it planning. One must organize around some purpose.

This article is based on a talk presented at a joint meeting of IEEE, HSPE, ASCE, and ASME in Honolulu, Hawaii, Nov. 19, 1968.

REFERENCES

1. Emerson, H., "The twelve principles of efficiency," *Eng. Mag.*, 1912.
2. *Wall St. J.*, June 8, 1968.

A. R. White is a principal of The Emerson Consultants, Inc., of New York where his work pertains primarily to corporate organization and attendant control and information systems. Born in China, he was educated as an accountant in England and spent World War II as a captain in the British Commandos in Burma where he organized guerilla activities and escape routes for downed air force fliers. His organization experience continued after the war as Assistant District Commissioner for Henzada. He came to the United States to continue his studies in 1948 under the auspices of E. I. du Pont de Nemours. Subsequent enterprises took him to the Far East, and in 1955 he returned to the U.S. to organize a branch factory for a Japanese camera manufacturer. In 1960 he joined a consulting firm, becoming vice president three years later. He has been with Emerson Consultants since Emerson since 1963.

JULY 5'69

Scanning the issues

Advance tables of contents
Translated journals
Special publications

Measurements conference recorded. It is no surprise to those who are familiar with the measurements field that a recent author in this publication unequivocally stated his belief that the improvement in certain areas of the field must stand as one of the truly amazing developments of the electronic age. Referring to precision advances of the order of a billion to one over the last half century, the comment may well apply to other areas of instrumentation and measurement.

To illustrate the point, one need only peruse the December 1968 issue of IEEE Transactions on Instrumentation and Measurement, which is devoted to most of the papers presented at the 1968 Conference on Precision Electromagnetic Measurements. The opening paper, by T. Igarashi *et al.*, offers one the opportunity to deal with attofarad sensitivities in the "Determination of an Absolute Capacitance by a Horizontal Cross Capacitor." The sensitivity obtained for electrical measurement is better than 0.01 aF (0.01×10^{-18} F).

A second article, by W. B. Kendall and P. L. Parsons, examines a frequency measuring technique that was developed as a method of reducing data for the occultation experiment of the Mariner-V Venus probe flown in the summer of 1967. Designed to obtain information about atmospheric conditions by having the spacecraft fly behind the planet, the technique consists of analyzing, at a low signal-to-noise ratio, the RF signal to detect changes in its frequency. Any deviation gives a measure of the index of refraction along the signal path, thereby providing valuable information in postulating an atmosphere. Although real-time Doppler data gave useful information, a permanent record that permitted thorough analysis and multiple reductions was provided by a digitized recording, cite the authors in "A Precision Frequency Measurement Technique for a Moving Signal with a Low Signal-to-Noise Ratio." Several iterations were required to reach a level of −190 dBm, at which point the signal became lost.

If one were to consider a time dissemination system wherein time information derived from a reference clock is broadcast to users who compare the received time with that of their local clocks, the problem of measuring the time difference between the local clock and the remote reference clock would be reduced to predicting the propagation delay experienced by the radio wave. The accuracy of predictability of the propagation delay usually associated with these HF transmissions is limited to the millisecond region for most users. This limitation arises from the inability to predict the exact route a radio signal follows in arriving at the user, as well as the difficulty of defining the radio refractive index at all points along the path.

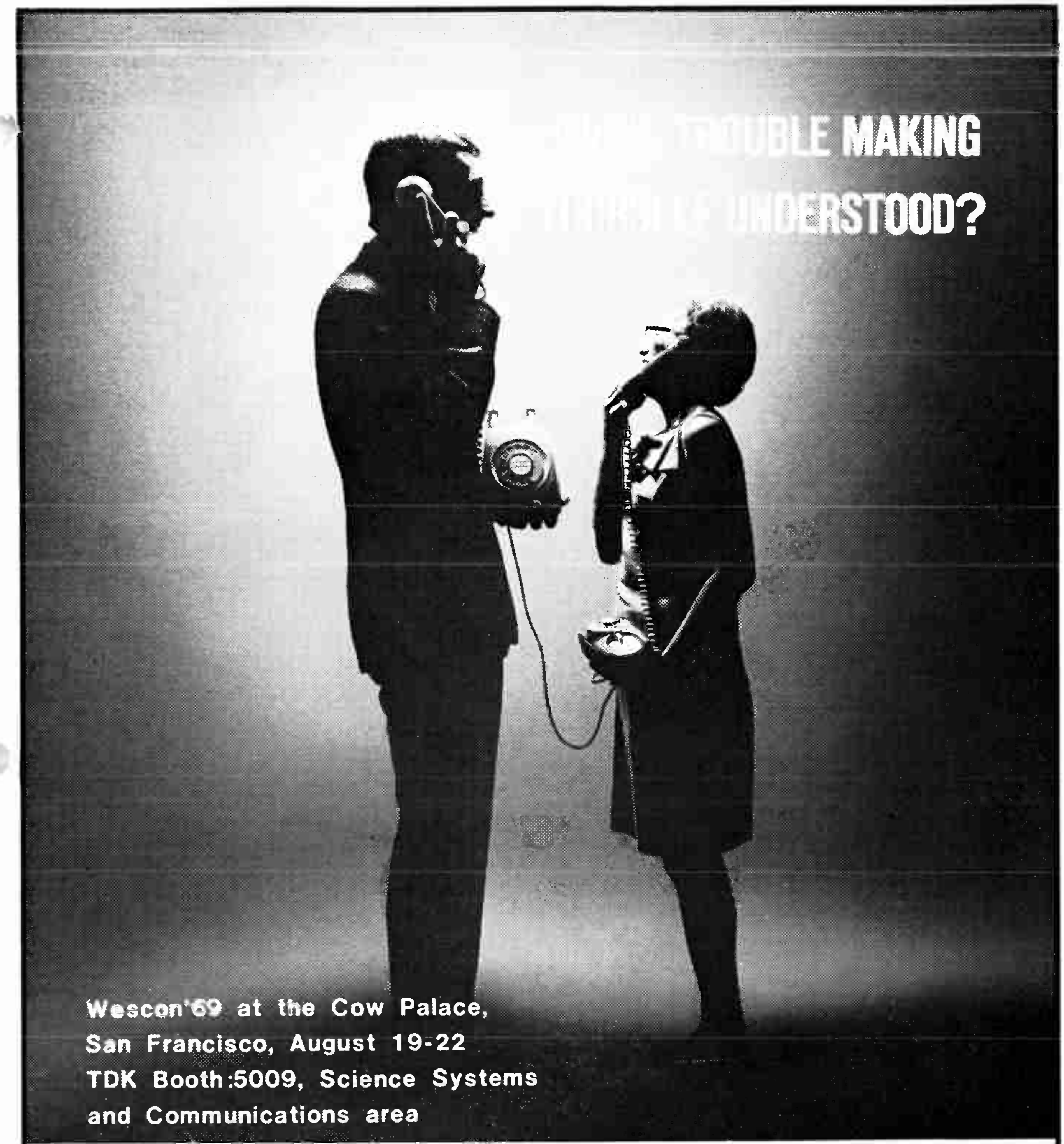
L. E. Gatterer *et al.*, in their paper entitled "Worldwide Clock Synchronization Using a Synchronous Satellite," describe a technique involving the one-way transmission of a radio time signal from a reference clock that is relayed by a geostationary satellite VHF transponder to remote clocks. The newer accuracy of predictability, and hence of clock synchronization, is found to be 60 μs if the propagation delay is computed using a satellite orbit predicted one week in advance, and 10 μs if the orbit used is updated to the time of the measurements. The method may offer an alternative to transporting atomic standards to geodetic and spacecraft tracking stations throughout the world. (*IEEE Trans. on Instrumentation and Measurement*, December 1968.)

More MIS studies listed. For those persons desiring to update the bibliography of metal-insulator-semiconductor (MIS) studies that was published in the November 1967 special MIS issue of the IEEE Transactions on Electron Devices, the December 1968 issue of these Transactions contains a listing of 158 additional papers in the field. Entered as one of 13 papers comprising another special MIS issue, the newer bibliography maintains the same purpose and scope as that of the first, and entries have been grouped according to the same classifications. (E. S. Schlegel, "Additional Bibliography of Metal-Insulator-Semiconductor Studies," *IEEE Trans. on Electron Devices*, December 1968.)

Programming nightmare. The sudden appearance over the last year and a half of an odd dozen process control languages has caused the field to examine the question of how well their process control needs were being served by the new software. With each vendor claiming to offer a panacea for the present-day ills, it became quite clear that the time was ripe for a thorough appraisal of the various new systems—with an eye on possible language standardization. With this purpose, the December 1968 IEEE Transactions on Industrial Electronics and Control Instrumentation was designed as a special issue on process control computer programming languages. In the words of Guest Editor Eric A. Weiss of Sun Oil Company:

"Users of process control computers recognized very early that the programming problem was of paramount importance. Process control computer vendors, bemused by their successes in electronic design, treated the programming problem as a minor and insignificant nuisance. Some vendors paid for this miscalculation with their corporate lives. . . .

"But the problems of economically programming and reprogramming process control computers are still unsolved and still significant. In the last eighteen months about a dozen process control programming languages have been created and introduced, chiefly by computer vendors. Each one has been announced as being easy to learn, easy to use, efficient,


HAVING TROUBLE MAKING
YOURSELF UNDERSTOOD?