VOLUME 4 □ NUMBER 10 OCTOBER 1967

IEEE spectrum

features

41 **Spectral lines: Voluntary page charges**
Among the important actions taken by the Board of Directors at their August meeting was the approval of a new source of revenue for Group publications—the voluntary page charge

44 **The laser gyro** Joseph Killpatrick
Recent developments in the laser field are evolving a new breed of gyro guidance instruments that show promise of greater accuracy and reliability than formerly possible

56 **Television's role in tomorrow's world** George H. Brown
The computer and closed-circuit television are natural partners, because television can act as an organ of vision whereas the computer supplies a rudimentary form of intelligence

59 **Compatible EMI filters** H. M. Schlicke, H. Weidmann
The frustrations so often encountered in filtering against electromagnetic interference can be overcome by discarding much of our conventional thinking regarding filter design

73 **Nuclear power—the next decade of development** Milton Shaw
The next ten years should witness substantial progress in expanding the uses of nuclear power from electric power generation to desalination and many other process applications

81 **Supervisory control of remote manipulation**
William R. Ferrell, Thomas B. Sheridan
Complex robot-computer systems are being considered as a means of extending man's ability to manipulate remote objects, be they in the ocean depths or in outer space

89 **The digital computer as a creative medium** A. Michael Noll
The famous machine that only does what you tell it to do produces some unexpected outputs after all

96 **The scanning electron microscope** R. F. W. Pease
In the SEM, the signal resulting from the interaction of a beam with the specimen modulates the brightness of a cathode-ray tube and thus a picture of the specimen is built up on the tube face

42 **Authors**

Departments: *please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

the cover

What's man, what's machine? A digital computer generated the straight lines by a pseudorandom algorithm. White areas were blocked in partly by machine algorithm and partly by Yale graphic design student Aaron Marcus, who was experimenting with computer design at Bell Labs. Man-machine art is the theme of the article starting on page 89.

Frank Haneman, head of AIL's Special Systems and Techniques Department, and Ted Flattau, consultant, this month discuss a solution to the problem of maintaining phase synchronization between two oscillators which are many miles apart.

Remotely Synchronizing Oscillators Over Time Varying Paths

AIL has developed a system (now undergoing field tests) that will phase-synchronize two oscillators that are separated by a multimile time-varying propagation path. The technique employed also provides continuous data describing the nature and magnitude of any path variations that occur within the control bandwidth of the system.

The phase synchronization system has application wherever coherent phase-stable signals are required at widely separated locations, and where quantitative information is desired about the propagation delay variations between the locations. This system is useful in long-baseline interferometers; for establishing a remotely synchronized time base or clock (ultra-stable oscillators are not needed); and propagation path studies.

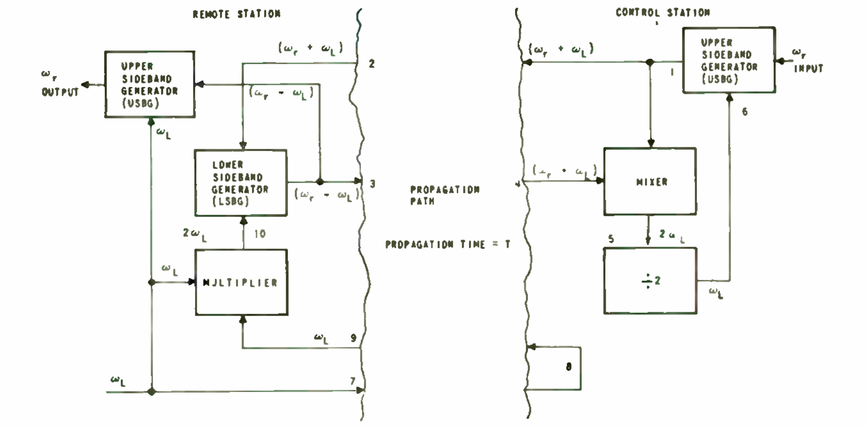
The operation of the system can be understood with the aid of Figure 1—a simplified block diagram of the system, and the table included therein. The object is to have the phase of the frequency ω_r at the OUTPUT in the remote station equal to the phase of the frequency ω_r (the system reference) at the INPUT in the control station. The system uses two two-way RF control links to measure the propagation characteristic of the transmission path, and to control the phase of the OUTPUT oscillator.

For this description we will assign an arbitrary phase θ_r to ω_r at the INPUT, and label the phase of ω_L at point 6, θ_L. The output of the Upper Sideband Generator (USBG) at point 1 is then, in vector notation, $(\omega_r + \omega_L)\underline{/\theta_r + \theta_L}$. After transmission, the signal received at remote station, point 2, is then $(\omega_r + \omega_L)\underline{/\theta_r + \theta_L - (\omega_r + \omega_L)T}$, where T is the one-way propagation time.

Now let us consider the sub-loop 7, 8, 9, 10. This loop is used to obtain the frequency ω_L shifted in phase by the round trip time delay of the propagation path. This signal $\omega_L\underline{/-2\omega_L T}$ is mixed with the output of the ω_L frequency generator $(\omega_L\underline{/0})$ to produce $2\omega_L\underline{/-2\omega_L T}$.

The Lower Sideband Generator (LSBG) produces the signal $(\omega_r - \omega_L)\underline{/\theta_r + \theta_L - (\omega_r - \omega_L)T}$ at point 3. After transmission to the control station, this signal is then $(\omega_r - \omega_L)\underline{/\theta_r + \theta_L - 2(\omega_r - \omega_L)T}$ at point 4. It is mixed with output of the USBG to produce $2\omega_L\underline{/2(\omega_r - \omega_L)T}$. Dividing by two produces $\omega_L\underline{/(\omega_r - \omega_L)T}$.

Thus, we see that the phase of the signal originally labeled θ_L is $(\omega_r - \omega_L)T$ and therefore, the signal at point 3 can be written as $(\omega_r - \omega_L)\underline{/\theta_r}$. This signal and the signal $\omega_L\underline{/0}$ are used in the OUTPUT USBG to produce $\omega_r\underline{/\theta_r}$.



CONDITIONS IN LOOP

6	ω_L	$\underline{/\theta_L}$
1	$(\omega_r + \omega_L)$	$\underline{/\theta_r + \theta_L}$
2	$(\omega_r + \omega_L)$	$\underline{/\theta_r + \theta_L - (\omega_r + \omega_L)\ T}$
3	$(\omega_r - \omega_L)$	$\underline{/\theta_r + \theta_L - (\omega_r - \omega_L)\ T}$
4	$(\omega_r - \omega_L)$	$\underline{/\theta_r + \theta_L - 2(\omega_r - \omega_L)\ T}$
5	$2\omega_L$	$\underline{/2(\omega_r - \omega_L)\ T}$
6	ω_L	$\underline{/(\omega_r - \omega_L)\ T}$
∴	$\theta_L = (\omega_r - \omega_L)\ T$	
7	ω_L	$\underline{/0}$
8	ω_L	$\underline{/-\omega_L T}$
9	ω_L	$\underline{/-2\omega_L T}$
10	$2\omega_L$	$\underline{/-2\omega_L T}$

FIGURE 1.

The system now being field tested uses phase-locked loops as the sideband generators. The reference frequency $f_r = \omega_r/2\pi$ is 1347.5 MHz and the offset frequency $f_L = \omega_L/2\pi$ is 100 kHz. The system synchronizes phases to within ±3 degrees over a line of sight path of up to 20 miles, with a control bandwidth of 500 Hz. It requires that the propagation time delay T be the same at all of the transmitted frequencies. This requirement is satisfied by using frequencies that are not widely separated.

The signals in the control station at points 6 and 8 can be compared in a phase detector (not shown) to obtain the propagation phase delay ω_rT.

Figure 2 is a photo of the equipment.

Phase variations of thousands of electrical degrees can be encountered over a propagation path of 20 miles. Thus, the reduction to ±3 degrees is an important improvement in both accuracy and sensitivity. A ±3 degree phase variation contributes an error of angular position of about 10 milliseconds of arc in a 20-mile baseline interferometer operating at 1300 MHz.

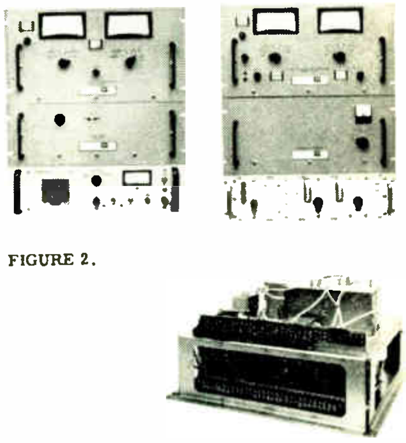


FIGURE 2.

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


For Production Efficiency

CRM FINE WIRE for ULTRASONIC BONDING — ALUMINUM & ALUMINUM ALLOYS

featuring

Cleanliness • Strength • Uniformity

Contact CRM engineers for recommendations for your application.

Write for latest catalog.

CONSOLIDATED REACTIVE METALS, inc.
115 Hoyt Avenue — Mamaroneck, N. Y. — OWens 8-2300 — TWX 914-835-3380

At least one of your interests is now served by one of IEEE's Groups

Group Transactions and Journals publish specialized papers in their specialized field. In addition, Groups sponsor conferences and symposia, as well as sessions of the Convention. Through Group Chapters, members have frequent opportunity to participate in Section activities pinpointed to the Group's technical specialty.

Aerospace and Electronic Systems (G-10)	Fee $5
Antennas and Propagation (G-3)	Fee $6
Audio and Electroacoustics (G-1)	Fee $3
Automatic Control (G-23)	Fee $6
Broadcast & Television Receivers (G-8)	Fee $4
Broadcasting (G-2)	Fee $3
Circuit Theory (G-4)	Fee $6
Communication Technology (G-19)	Fee $5
Computer (G-16)	Fee $6
Education (G-25)	Fee $3
Electrical Insulation (G-32)	Fee $5
Electromagnetic Compatibility (G-27)	Fee $4
Electron Devices (G-15)	Fee $7
Engineering in Medicine and Biology (G-18)	Fee $4
Engineering Management (G-14)	Fee $5
Engineering Writing and Speech (G-16)	Fee $3
Geoscience Electronics (G-29)	Fee $3
Human Factors in Electronics (G-28)	Fee $3
Industrial Electronics and Control Instrumentation (G-13)	Fee $3
Industry and General Applications (G-34)	Fee $6
Information Theory (G-12)	Fee $5
Instrumentation and Measurement (G-9)	Fee $3
Magnetics (G-33)	Fee $5
Microwave Theory and Techniques (G-17)	Fee $5
Nuclear Science (G-5)	Fee $3
Parts, Materials and Packaging (G-21)	Fee $4
Power (G-31)	Fee $6
Reliability (G-7)	Fee $3
Sonics and Ultrasonics (G-20)	Fee $3
Systems Science and Cybernetics (G-35)	Fee $4
Vehicular (G-6)	Fee $3
Journal of Quantum Electronics (option G-15A or G-17A)	Fee $5
Journal of Solid State Circuits (by subscription or option G-15B) (For 1967 G-4 members receive free)	Fee $6

Membership in IEEE Groups at the above annual fees is open only to individuals who are already members of the IEEE. Information on non-member subscription rates for Group *Transactions* or on joining the IEEE is available from IEEE Headquarters at the address given below.

**The Institute of
Electrical & Electronics Engineers
345 East 47th St., New York, N.Y. 10017**

IEEE spectrum

Editorial Board

C. Chapin Cutler	*Editor*

Colin Cherry	B. M. Oliver
C. C. Concordia	Herre Rinia
A. C. Dickieson	Shigebumi Saito
Peter Elias	J. J. Suran
E. G. Fubini	Charles Süsskind
E. W. Herold	L. A. Zadeh

Ronald K. Jurgen	*Managing Editor*
Robert E. Whitlock	*Senior Editor*
Gordon D. Friedlander	*Staff Writer*
Nilo Lindgren	*Staff Writer*
Evelyn Tucker	*Assistant Editor*
Harold L. Hoffman	*Assistant Editor*
Roberta A. Reynes	*IEEE News Editor*
Stella Grazda	*Editorial Assistant*

Herbert Taylor	*Art Director*
Ruth M. Edmiston	*Production Editor*
Jeff Derecki	*Assistant to the Art Director*
Morris Khan	*Staff Artist*

IEEE PUBLICATIONS BOARD

F. K. Willenbrock	*Chairman*
E. L. Harder	*Vice Chairman*

C. L. Coates	E. C. Starr
C. C. Cutler	J. E. Thomas
E. E. David	D. Van Meter
W. A. Higinbotham	M. E. Van Valkenburg
D. D. King	G. Wade

PUBLICATIONS OPERATIONS

Ralph H. Flynn
Director, Publishing Services

Elwood K. Gannett
Director, Editorial Services

Helene Frischauer *Administrative Editor*

Alexander A. McKenzie
Assistant to the Director, Editorial Services

ADVERTISING

William C. Copp *Advertising Manager*

Lillian Petranek
Assistant Advertising Manager

Carl Maier
Advertising Production Manager

VOLUME 4 □ NUMBER 10 IEEE spectrum OCTOBER 1967

Spectral lines

Voluntary page charges. Among the important actions taken by the IEEE Board of Directors at their August meeting in San Francisco was the approval of a new source of revenue for Group publications. This source of revenue, which has been utilized by a growing number of other scientific and technical societies in the United States for a number of years, has been called the "voluntary page charge" system. Essentially, it is a method by which the organization with which the author or authors of a paper are associated is invited to assist in defraying a part of the cost of publication.

It is important to recognize what a "voluntary page charge" is and what it is not. Let's start with what it is not. It is not a fee paid for space in a publication; it is not a factor in the editorial handling of a paper; it is not related to the decision of whether or not a paper is accepted for publication. Rather, it is a recognition that the costs associated with a research or development project might also include the cost of its dissemination to the scientific or technical community concerned. Indeed, the U.S. Government recognizes the payment of voluntary page charges as a legitimate part of the cost of performing research and development work under Government contract.

The mechanics are as follows: After a paper has been accepted for publication through whatever review procedure a particular Group uses, the author is sent, with the galley proofs, an order form for reprints on which his institution is asked to indicate in advance whether or not it plans to pay the page charge. If the answer is affirmative, the IEEE sends a bill to the institution (not to the author) after publication of the paper. If the answer is negative, no bill is sent, nor is one sent in such cases where the author is without an organizational affiliation.

The bill is calculated on the basis of the number of published pages a paper uses. It is planned to charge $50 per page in 1968 for papers published by letterpress printing, for those IEEE publications that choose to adopt the voluntary page charge system. If an author's organization pays the bill, the funds are transmitted to IEEE Headquarters and become part of the funds available to the Group to pay for publication costs. If not, no charge is made to the scheduled publication of the paper. It is customary to provide 100 reprints free to those organizations that pay the bill; reprints are available at cost to organizations that do not.

Payment of the bill is entirely voluntary and the position taken by the author's organization is privileged information; it is not made available to the editor, to his reviewers, or to others involved in the editorial procedures of determining whether or not a paper is accepted. Thus the payment or nonpayment of the bill is unrelated to the paper's publication. As the editors and reviewers do not know which bills are paid and which are not, their editorial decisions are independent of the policy adopted by the author's organization.

An *ad hoc* committee* of the Technical Activities Board and the Publications Board, which was chaired by IEEE Treasurer John V. N. Granger, carefully studied the question of whether or not the Institute should approve this system. It was found that the publication costs of many IEEE Groups are increasing as the square of the number of Group members whereas the Groups' major sources of income increase directly with the number of members. In a number of important fields, the number of papers to be published increases directly with the number of members. In considering the implications for the future, the committee concluded that the use of voluntary page charges is essential for some Groups as they have no other means to meet the rapidly rising costs of publishing the increasing quantity of papers that they judge worthy of publication.

It is important to note that the action of the Board of Directors was permissive, not mandatory. Groups must request the voluntary page charge and the approval of the Operating Committee of TAB and the Publications Board is required in each case.

It will be recalled that approval was given recently for Groups to include display advertising in their publications if it could be established that it was both technically and economically advisable to do so. This option has not been exercised by many Groups as yet but it represents another possible source of revenue. It is anticipated that, in general, those Groups for which voluntary page charges are most appropriate will not be suitable vehicles for advertising, and vice versa. Thus, the two systems—page charges and advertising—are somewhat complementary.

The decision to approve voluntary page charges was not taken lightly, but has been under discussion and study for more than two years. The Committees and Boards that considered it were sensitive to the fact that it could conceivably be a source of embarrassment to some authors or their organizations. By careful handling we expect to avoid such possible embarrassment. The system is being initiated to provide a solution to the continuing problem of growing publication costs. More efficient techniques of information dissemination may evolve but we need to keep our present methods economically viable and strong until they are replaced. The IEEE cannot afford to perform its primary function—information dissemination—poorly. *F. Karl Willenbrock*

* Members include J. Earl Thomas, R. W. Sears, Werner Buchholz, Calvin Mooers, and E. K. Gannett.

Authors

The laser gyro (page 44)

Joseph E. Killpatrick (M), Electro-Optics Section chief at Honeywell Inc., St. Paul, Minn., is currently engaged in directing research in the areas of CO_2 lasers, gas lasers, and the laser gyro. The Co_2 laser work is in the area of laser communication, radar, and Doppler applications.

Mr. Killpatrick's previous experience includes work on the theory and application of electromagnetic fields and radiation, and he has conducted studies of theoretical and practical applications of optical techniques to the design of horizon scanners, sun sensors, and star trackers. The initial design of the OSO sun seeker requiring intricate scanning and pointing was under his direction. He has been awarded several patents and has applied for others in the fields of horizon scanners, signal processing, colorimetry, and the laser gyro. He is a member of the Optical Society of America.

Television's role in tomorrow's world (page 56)

George H. Brown (F), executive vice president for research and engineering at RCA, Princeton, N.J., serves on the board of directors of RCA and of RCA Communications, Inc. He is a Fellow of the American Association for the Advancement of Science, and he has received the IEEE Edison Medal and the Modern Pioneer Award. He is included in *American Men of Science* and in Dunlap's *Radio's 100 Men of Science*.

After receiving four degrees from the University of Wisconsin, Dr. Brown joined the RCA Manufacturing Company, Camden, N.J., in 1933, in the position of research engineer. There he developed the turnstile antenna for television, FM radio, and facsimile transmission. In 1942 he transferred to the research center at Princeton, and he has been a vice president since 1959.

Compatible EMI filters (page 59)

H. M. Schlicke (F), manager of the Laboratories for Applied Electronic Research at Allen-Bradley Company, Milwaukee, Wis., is concerned with conceiving, planning, and realizing (to the point of pilot production) new products for the company. He is the originator of new ceramic low-pass filters and antenna multiplexers and, recently, he has been concentrating on true thin-film microelectronics and nonconventional filters. He is a member of the IEEE G-EMC Administrative Committee.

H. Weidmann has been serving as a development/project engineer at Allen-Bradley Company since 1962. He has been concerned with various problems of VHF multiplexers, and with the development of active interference filters for dc and ac power lines.

He received the diploma in electrical engineering from the Federal Institute of Technology, Switzerland, and then served in the Institute for Industrial Electronics. He joined Allen-Bradley in 1958 and the Swiss National Railroads in 1960.

Nuclear power—the next decade of development (page 73)

Milton Shaw has been serving as director in the Division of Reactor Development and Technology, U.S. Atomic Energy Commission, Washington, D.C., since 1964. He is responsible for programs dealing with the development and improvement of nuclear reactors and isotopic systems and associated equipment for civilian and assigned military applications. During the three years prior to his present appointment, he was the senior technical assistant to the Assistant Secretary of the Navy for Research and Development, and was in charge of test and evaluation matters for the Navy and Marine Corps, and for liaison for engineering, scientific, program, and planning offices with the Department of Defense and other Government agencies. Formerly, he served for 12 years in the Naval Reactors program under Vice Admiral Rickover.

Supervisory control of remote manipulation (page 81)

William R. Ferrell (M) is presently serving as professor of mechanical engineering at the Massachusetts Institute of Technology, Cambridge. After service in the United States Army Signal Corps, he received the bachelor of arts degree, with honors, from Swarthmore College in 1954. He earned the bachelor of science and master of science degrees in mechanical engineering from M.I.T. in 1961, and the Ph.D. degree in 1964.

Dr. Ferrell has been engaged in work on machine and product design and, since 1961, he has been associated with the Man–Machine Systems Laboratory at M.I.T. There he has investigated delayed sensory feedback and other problems related to remote manipulation.

In his present position, Dr. Ferrell includes among his teaching and research interests a concern for human performance in decision and control tasks as well as engineering design.

Thomas B. Sheridan (M) is associate professor of mechanical engineering at the Massachusetts Institute of Technology, is head of the Man–Machine Systems Laboratory, and he also has been serving in the capacity of editor of the IEEE TRANSACTIONS ON HUMAN FACTORS IN ELECTRONICS since 1964. His current research activities include human manipulation and control, human and machine learning, and sensory aids for the blind.

He received the bachelor of science degree in mechanical engineering from Purdue University in 1951 and the master of science degree in engineering from the University of California at Los Angeles in 1954. He was awarded the Sc.D. degree from M.I.T. in 1959 after participating in an interdisciplinary program combining engineering and psychology. Between 1951 and 1953, while serving in the U.S. Air Force, he worked in the Aeronautical Laboratory at Wright Air Development Center on the design of high-altitude escape systems.

The digital computer as a creative medium (page 89)

A. Michael Noll is in the Speech and Communication Research Department at Bell Telephone Laboratories, Inc., Murray Hill, N.J. He joined the Laboratories in 1961 and, initially, worked on the assessment of telephone quality. In 1965 he transferred to the Acoustics Research Department, where he was concerned with computer simulations and investigations of short-time spectrum analyses and new methods for vocal pitch determination. Presently, his interests include computer-generated three-dimensional displays of data, the application of computer technology to the visual arts, and psychological investigations of human reactions to pseudorandom patterns. He is a member of the Acoustical Society of America, the Audio Engineering Society, and the American Society of Aesthetics.

The scanning electron microscope (page 96)

R. F. W. Pease is presently on industrial leave from the post of assistant professor of electrical engineering at the University of California at Berkeley, and he is working at Bell Telephone Laboratories, Inc., Holmdel, N.J., on problems concerning the electrical communication of optical images.

After receiving the B.A. degree in natural sciences and electrical engineering from Cambridge University, England, in 1960, he did research at the Cambridge University Engineering Laboratory, there designing and building a scanning electron microscope with a resolving power better than 100 Å. For this work, he was elected to a Research Fellowship at Trinity College, Cambridge, in 1963, and he received the master of arts and Pd.D. degrees in 1964.

Since that time, he has been serving as assistant professor at Berkeley, and he has been engaged in research on the application of the scanning electron microscope to a wide variety of problems in both the physical and the biological sciences.

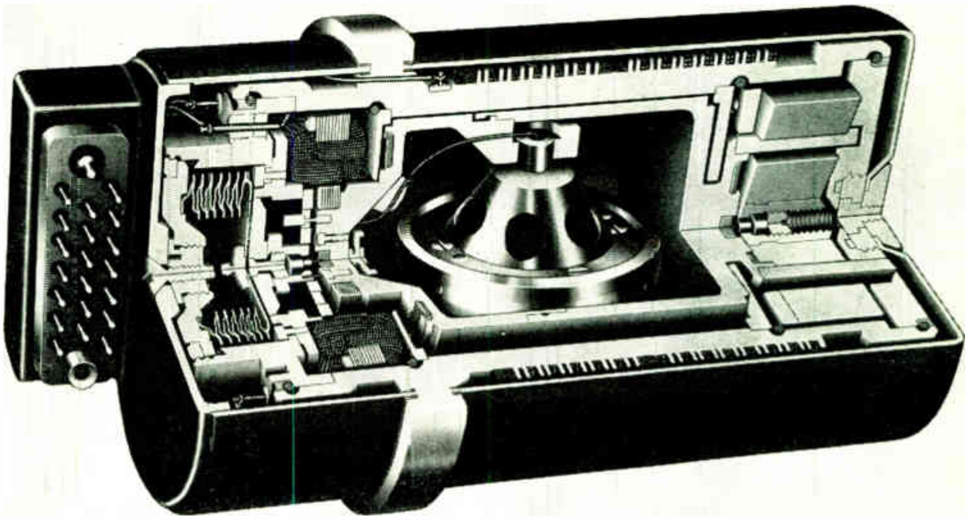


FIGURE 1. Conventional gyro.

The laser gyro

Now in the research and development stage, the laser gyro shows promise of replacing many conventional inertial gyroscopes. It is not subject to prolonged starting time or to the effects of sudden acceleration—ideal specifications for our space-age aeronautic society

Joseph Killpatrick *Honeywell Inc.*

One of the most dramatic recent developments in optical technology is the laser gyro, which combines the properties of the optical oscillator, the laser, and general relativity to produce an integrating rate gyroscope. This gyro measures rotation in inertial space, but does not use a spinning mass as conventional gyros do. Because of the absence of spinning mass, the gyro's performance is not affected by accelerations; and it can sense very high rates with great accuracy. Other important advantages of the laser gyro are lack of special cooling, low power consumption, and simplicity of construction.

The ideal sensor for inertial rotation (angular motion with respect to the "fixed" stars) would be a device whose output gives angle increments, just as an angle encoder measures rotation between two mechanical elements. The device would have a well-defined input axis about which the rotations were measured. The angle sensed also would be free of acceleration effects, both angular and linear, independent of the rate at which the angle is measured, and independent of special environmental factors such as temperatures. The output of the instrument should be, if possible, high-resolution digital signals that are easily handled. Such a device is available in the form of a single-axis rate-integrating gyro, and it is currently used in many different control systems. Inertial guidance systems control airplanes, helicopters, the stabilization of guns, cameras, and radar antennas, the guidance of ships, aircraft, and other vehicles by using gyroscopes of this type to provide attitude references.

In order to sense attitude changes for motions in any direction, three such sensors must be used to independently sense the rotations about three axes, usually orthogonal, and use this information to determine the precise angles between the coordinate system defined by the gyros and the "fixed" stars.

In general, the gyro is fixed to the vehicle and, as such,

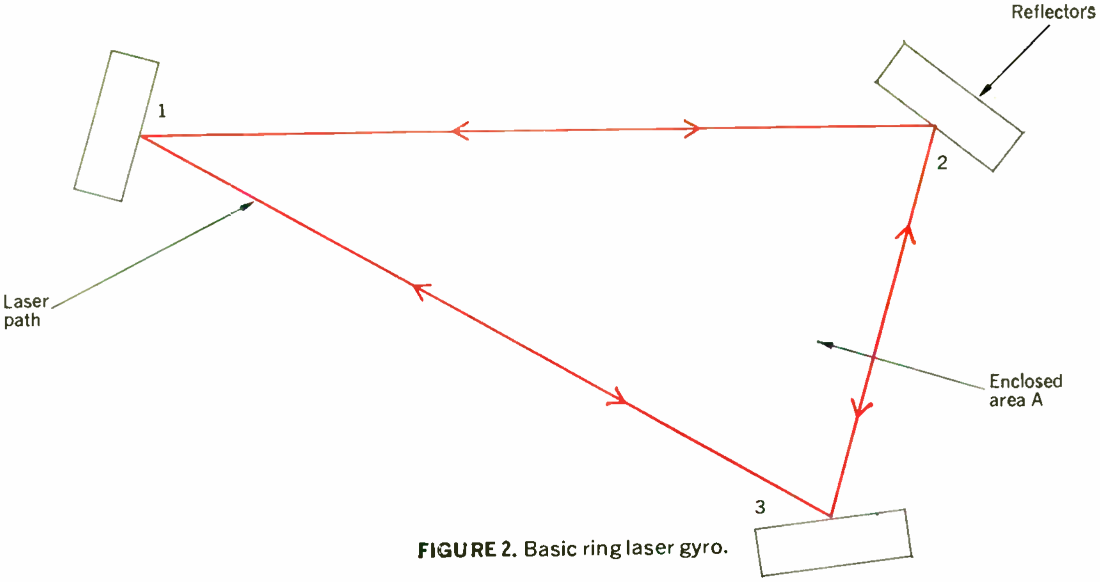


FIGURE 2. Basic ring laser gyro.

will find its input axis orientation varying with time. The output is in the form of a phase shift between two electric signals, and is the integral of the input rate about its input axis:

$$\Delta\theta = \int_{t_1}^{t_2} \Omega(t)n(t)dt \qquad (1)$$

$$\theta_{t_2} - \theta_{t_1} = \int_{t_1}^{t_2} \Omega_{\text{in}} \cdot \mathbf{n} dt \qquad (2)$$

where Ω is the input rate, $\mathbf{n}$ is the input-axis unit vector, and $\Delta\theta$ is the phase shift between the two signals.

Why a laser gyro?

Conventional gyros using a spinning mass as a sensing element suffer inaccuracies due to many error sources. These include mass unbalance and structure instabilities that give rise to acceleration and acceleration-squared effects, and high rate effects. High costs are associated with the many elements, as well as the precision machining, assembly, and cleanliness required in their manufacture (Fig. 1). The laser gyro will avoid many of these problems, and will probably be very low in cost.

In addition, the run-up time of this gyro is the time it takes to ionize the gas discharge and start the current through the cavity. Times of the order of several milliseconds are typical. Inherently, this gyro is capable of operating within several milliseconds after pushing a start button. The "spin-up" of a mass, as used in a conventional gyro, is completely avoided.

The laser gyro is an integrating-rate gyro that senses inertial angles with high precision. Its most dramatic features are its ability to sense those angles in the presence of high accelerations, and high input rates as well as low accelerations. Its input is digital, and in its simplest form it is a device with an input axis and two electric outputs, one for positive and the other for negative angle increments. Whenever the rotation about the input axis equals the size of the output increment (typically one arc second) a pulse appears. If the input rate is constant, the output pulses will appear at a constant rate.

The rate at which the pulses occur is obviously directly proportional to the input rate to the gyro. By counting these pulses, the total angle through which the gyro has rotated can be measured.

It is paradoxical that, although simple in construction and operation, the laser gyro's description and analysis require consideration of extremely small effects. In fact, this is one of the most precise instruments resulting from the development of laser technology. It must measure path differences of less than 10^{-6} Å, and frequency changes of less than 0.1 Hz (a precision of better than one part in 10^{15}) in order to read rotation rates of less than 0.1 degree per hour. The instrument is simply a laser that has three or more reflectors arranged to enclose an area.* The three mirrors, together with the light-amplifying material in the laser path, comprise an oscillator (laser), as in Fig. 2. In fact, there are two oscillators, one that has energy traveling clockwise, and one that has energy traveling counterclockwise around the same path.

The frequencies at which these oscillators operate are determined by the optical length of the path they travel. In order to sustain oscillation, two conditions must be met: The gain must be equal to unity at some power level set by the amplifying medium, and the number of wavelengths in the cavity must be an exact integer (that is, the phase shift around the cavity must be zero). If the first condition is to be achieved, the laser frequency must be such that the amplifying medium has sufficient gain to overcome the losses at the reflectors and other elements in the laser path. In addition, the wavelength must be an exact integer fraction of the path around the cavity. This last condition actually determines the oscillation frequency of the laser.

When the enclosed ring is rotated in inertial space, the clockwise and counterclockwise paths have different lengths. (This effect is *not* due to a Doppler shift.) The path difference in these two directions causes the two

---

* It is interesting to note that the word "gyro" is derived from the Greek "gyros," meaning ring. In this context the laser gyro is perhaps more of a true gyro than the conventional rotating wheel, for it is simply a laser arranged in a ring, by using three or more reflectors.

I. Path difference as related to area and input rate

Area, square meters	Ω, degrees per hour	ΔL, angstroms
1	10	66.7×10^{-4}
10^{-2}	10	66.7×10^{-6}
10^{4}	10	66.7
10^{5}	10	667
1	10^{5}	667

oscillators to operate at different frequencies. The difference is proportional to the rate at which the ring is rotating, since path difference is proportional to inertial rotation rate. The readout of the gyro is accomplished by monitoring the frequency difference between the two lasers. The total phase shift, $\Delta\phi$, between the oscillator, cycles of output $\times 2\pi$, is the integral of the input rate, Ω:

$$\Delta f \propto \Omega_{\text{in}} \tag{3}$$

$$\frac{d(\phi_2 - \phi_1)}{dt} \propto \Omega_{\text{in}} \tag{4}$$

$$\Delta\phi \propto \int_{t_1}^{t_2} \delta\, \Omega_{\text{in}}\, dt \tag{5}$$

The laser gyro assembly consists of the gain media, the reflectors defining the path and enclosed area, and a readout method for monitoring the difference between the two oscillators. This is accomplished by evaluating the phase shift between the two oscillators, using two detectors to separate positive and negative phase shifts (to separate clockwise inertial rotation from counterclockwise rotations).

Basic principle: how the same path can be different

The key to the operation of the laser gyro is the description of the phenomena by which the path around the ring can be different for observers (photons) traveling *with* the direction of rotation than for observers traveling *against* the direction of rotation.

General relativity[1] predicts that two observers, traveling around a closed path that is rotating in inertial space, will find that their clocks are not in synchronization when they return to the starting point (traveling once around the path but in opposite directions). The observer traveling in the direction of rotation will experience a small increase, and the observer traveling in the opposite direction will experience a corresponding small decrease in clock time. The difference in the readings of these clocks, Δt, depends upon the inertial rotation rate Ω, the area A enclosed by the path, and the speed of light c.

$$\Delta t = \frac{\Omega}{c^2} 2A + \frac{\Omega}{c^2} 2A = \frac{4A\Omega}{c^2} \tag{6}$$

Since the laser gyro uses photons, traveling at the speed of light, for observers, the time difference appears as an apparent length change in the two paths of ΔL. This length change is equal to the observers' velocity times their time difference.

$$\Delta L = c\Delta t = \frac{4A\Omega}{c} \tag{7}$$

Thus, even though both observers travel the same physical path, they see an apparent length change, which is proportional to the enclosed area of the path and its rotation rate in inertial space. The relation of ΔL to A and Ω (see Table I) shows that, in order to obtain path differences large enough to measure by conventional interferometric techniques, either the rotation rate must be quite high or the enclosed area must be very large. In 1913 Sagnac[2] conducted an experiment to investigate this effect by using a small enclosed path about 0.1 m², rotated at 2 revolutions per second, and was able to detect length changes of 100 to 200 Å. In 1925 Michelson[3] conducted an experiment using the rotation rate of the earth (10°/h) and an enclosed area of 2×10^5 m². In this experiment the detected length change was about 1000 Å. In both experiments good agreement was obtained with the equation relating area, rotation rate, and change in path length. In order to measure low rates with a small enclosed area, however, a different method of measurement was clearly needed.

Measuring small length differences by laser

For measuring small length changes the use of an electrical oscillator was proposed, in which the cavity dimensions and lengths determined the oscillation frequency.[4] In this manner a small length change is transformed into an easily measured frequency difference between oscillators. The laser gyro uses two oscillators at high frequencies (3×10^{14} Hz). Exact frequencies are determined by the length of the two cavities, one for clockwise and the other for counterclockwise traveling radiation.

The laser oscillator operates at light frequencies and, as in all oscillators, it must have a gain mechanism arranged in such a way that the losses are equalized. It must also operate at a frequency controlled so that the phase shift for a trip around the cavity is zero.

The gain is usually provided in a laser gyro with an electrical gas discharge in a mixture of ten parts of helium to one part of neon at a total pressure of 1 to 10 mm of mercury. This gas discharge is like that of a neon sign. Because of a combination of energy transfer from the helium to the neon, wall effects, and decay rates, a gain results. In other words, at certain wavelengths, a light ray will experience a gain of from 1 to 5 percent when it passes through the gas discharge. Principally because of Doppler effects, this gain exists over a range of frequencies about 1.5 GHz wide. In general, the losses are determined by the reflectivity of the laser reflectors, diffraction losses that occur as a function of the cavity size, and the manner in which the light radiation travels around the cavity. Typically, the losses due to the reflectors are less than 0.3 percent per mirror or 0.9 percent for three mirrors. Since the gain must be greater than the loss, there is a limited range over which oscillation may occur. This range is typically 1 GHz.

In addition to the oscillator conditions of gain and loss, the condition of zero phase shift must also exist. Another way of saying this is that the number of wavelengths in the cavity must be equal to an integer; in the laser gyro, this integer is about one million. Many frequencies will satisfy these conditions of zero phase, but they are separated by an amount equal to c/L (the speed of light divided by the total length of the cavity). For a total length of one meter the frequency separation is 300 MHz. Thus, for a cavity of this size, there may be two or three possible frequencies of oscillation—that is, frequencies at

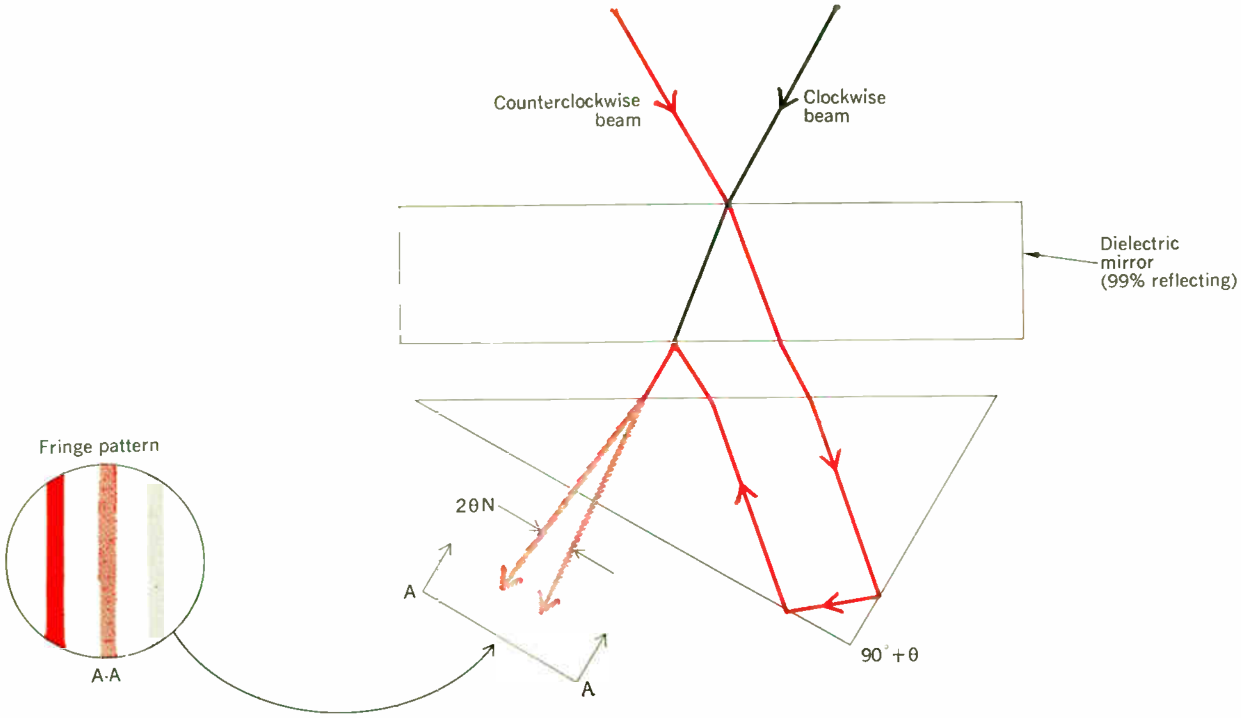
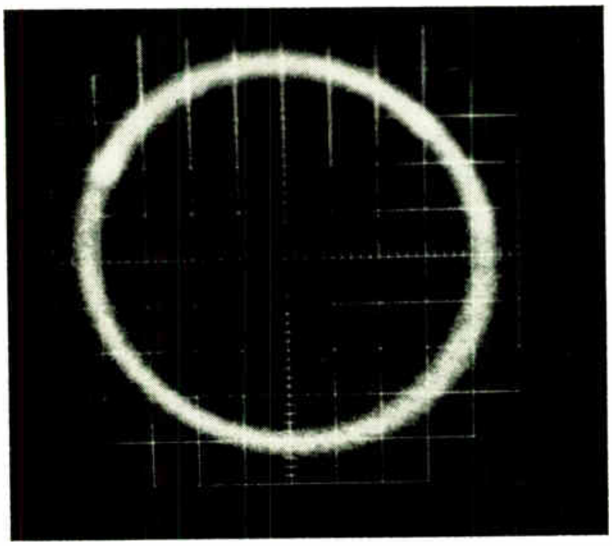


FIGURE 3. Laser gyro readout optics.

FIGURE 4. Lissajous pattern of two detector outputs.



which both the gain is greater than the losses and the phase shift around the cavity is zero. The laser frequency is determined primarily by the length of the cavity. The frequency band in which the gas medium has gain determines which of the many possible frequencies will "lase" or oscillate.

In practice, the exact phase shift must be considered, and this involves a study of phase-shift terms due to the gain, nonlinear saturation of the gain, scattering, and the role of various isotopes of neon.

The length differences in the two paths due to inertial rotation rates cause a difference in the frequency of these two oscillators, whereas physical changes in length caused by temperature, vibration, etc., do not cause major frequency differences. The fundamental boundary condition is that the laser wavelength λ must be equal to an integer fraction of the optical length around the cavity. N is an integer typically in the range of 10^5 to 10^6, and

$$\lambda = \frac{L}{N} \qquad (8)$$

A length change of ΔL will cause a wavelength change of

$$\Delta\lambda = \frac{\Delta l}{N} \qquad (9)$$

The corresponding frequency change is given as

$$\frac{\Delta f}{f} = \frac{\Delta L}{L} \qquad (10)$$

Therefore, given small length differences Δl and reasonable cavity lengths L, the operating frequency should be as high as possible.

The relation between inertial input rates Ω and apparent length change ΔL has been given as

$$\Delta L = \frac{4A\Omega}{c} \qquad (11)$$

The relation between Δf and Ω, in terms of the gyro size and wavelength, is determined by substituting Eq. (10) into Eq. (11), giving

$$\Delta f = \frac{4A\Omega}{\lambda L} \qquad (12)$$

Readout of the laser gyro

Frequency difference is monitored quite simply, as shown in Fig. 3. A small amount of the laser energy is transmitted through the mirror. This transmitted energy is uniquely related to the frequency and phase of the energy in the cavity. The two transmitted beams are superimposed to create a fringe pattern of alternate interference, and if the phase between the two oscillators remains fixed, the fringe pattern will also remain fixed. However, if the phase between the two oscillators changes (a frequency difference between the two oscillators), the fringe pattern will appear to move to the right or left. The direction of motion is determined by the direction of phase change (which oscillator is at the higher frequency), and the magnitude is measured by the number of fringes. Each fringe represents a phase change of one cycle, or 2π radians.

The output consists of two detectors mounted at the readout prism and spaced so that they are about a quarter of a fringe spacing apart. With this spacing, their outputs are phased so that the direction, as well as magnitude, of the fringe motion can be monitored. By forming a Lissajous pattern of the two detector outputs (Fig. 4), the phase shift of 90° is evident. When the gyro is rotated clockwise, the fringe pattern moves in one direction, causing the oscilloscope pattern to travel around a circle in a fixed direction, one revolution for each fringe shift.

When the gyro input is reversed, the fringe motion will reverse, as will the dot observed on the oscilloscope. These detector signals are simply processed to count the number of complete fringe motions, one direction as positive and the other as negative. (An oscilloscope is used only for monitoring and testing purposes, and is not needed for actual operational readout.)

II. Output frequency and fringes per arc second as related to various parameters ($\lambda = 1\ \mu m$)

Area, m^2	Ω, deg/h	Δt, seconds	ΔL, angstroms	Δf, Hz	Fringes/ Arc Second
1	10	22×10^{-22}	66.7×10^{-4}	50	5.0
10^{-2}	10	22×10^{-24}	66.7×10^{-6}	5	0.5
10^{-2}	0.10	22×10^{-26}	66.7×10^{-8}	0.05	0.5

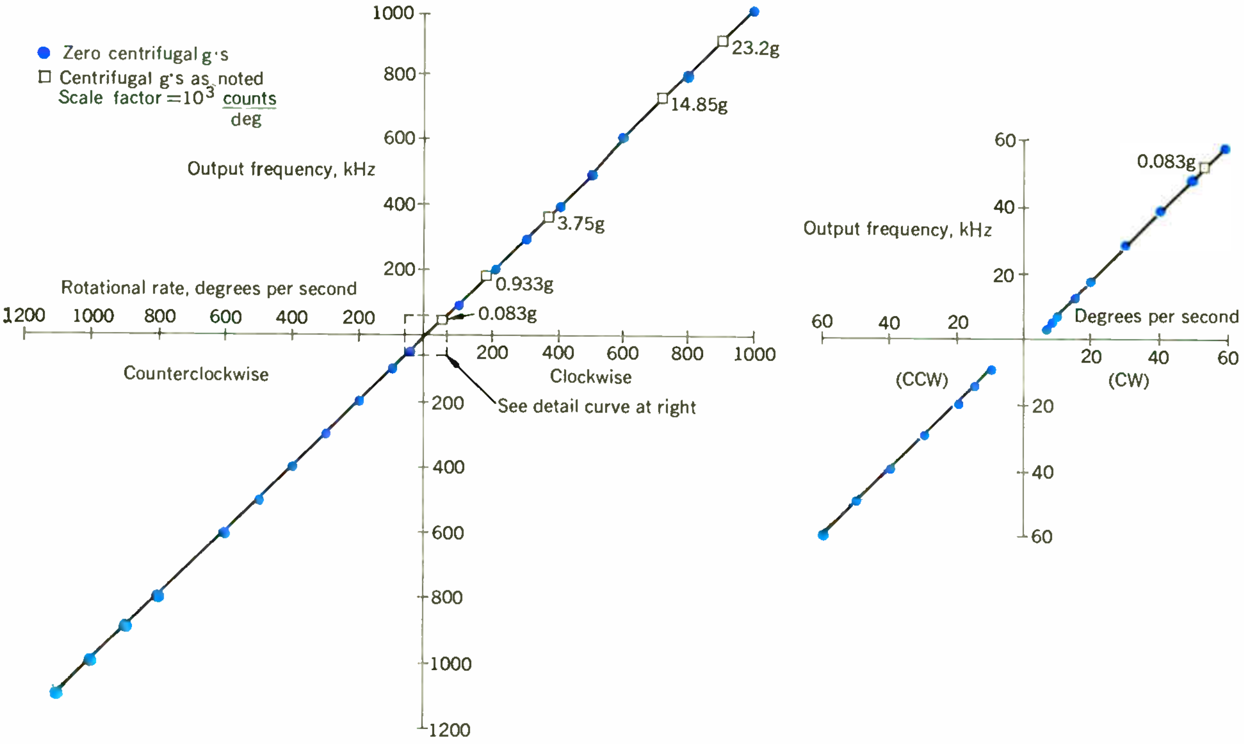
The signal from each detector is amplified and used to trigger digital counters that monitor plus and minus counts. Each count represents a phase change of 2π radians, or one cycle (resolution of one quarter of this value can be easily achieved), between the two oscillations in the laser gyro cavity. The size of each count is dependent upon the gyro relation between input rate and output frequency difference. An input rate of 1°/h (1/15 of the earth's turning rate) will typically produce a frequency difference of 1 Hz. One degree per hour is exactly one arc second per second of time; therefore, each second an inertial angle of one arc second has been generated, producing an output phase change of one cycle, or 2π radians. Each count has a weight of one arc second, and turning the gyro through an angle of 360°, or one revolution, would produce an output of 1 296 000 pulses. For rotations in one direction, these pulses are identified as positive, and in the opposite direction they are identified as negative. (The logic is identical to that used in digital incremental-angle encoders.)

Table II clearly shows the role of the laser oscillator in converting small length differences of approximately 10^{-6} Å into measurable quantities of frequency and phase.

An integrating-rate gyro

The fact that the output frequency difference is proportional to the input rate, as shown in Fig. 5, often leads to the impression that the laser gyro is a rate gyro. However,

FIGURE 5. Output frequency vs. rotation rate for laser gyro.



its output is the phase difference observed by the position of the interference pattern created by superimposing the energy from the two oscillators in the readout. The total phase shift change between the two oscillators over a given time interval is the integral of their frequency difference expressed as the instantaneous rate of phase change:

$$\Delta f = \frac{d(\phi_2 - \phi_1)}{dt} = \frac{4A}{\lambda L}\,\Omega \tag{13}$$

By integrating over a time interval, the total phase change is found as

$$\Delta\phi = \frac{4A}{\lambda L}\int_{t_1}^{t_2} \Omega\,dt \tag{14}$$

The inherent resolution (1 to 1/4 of a fringe) of about one arc second or less is sufficient to satisfy almost all conceivable applications. Of course, the resolution size and the smallest rate that can be sensed are not directly related, except when a minimum time to make that

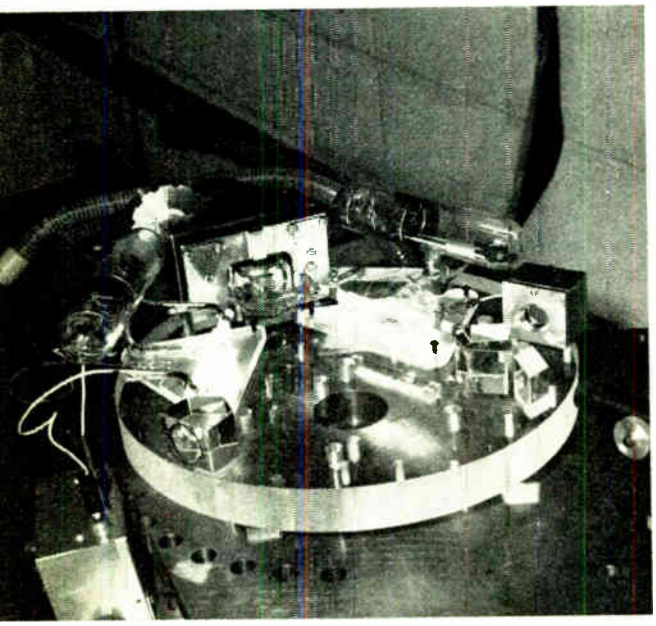


FIGURE 6. Early laser gyro model.

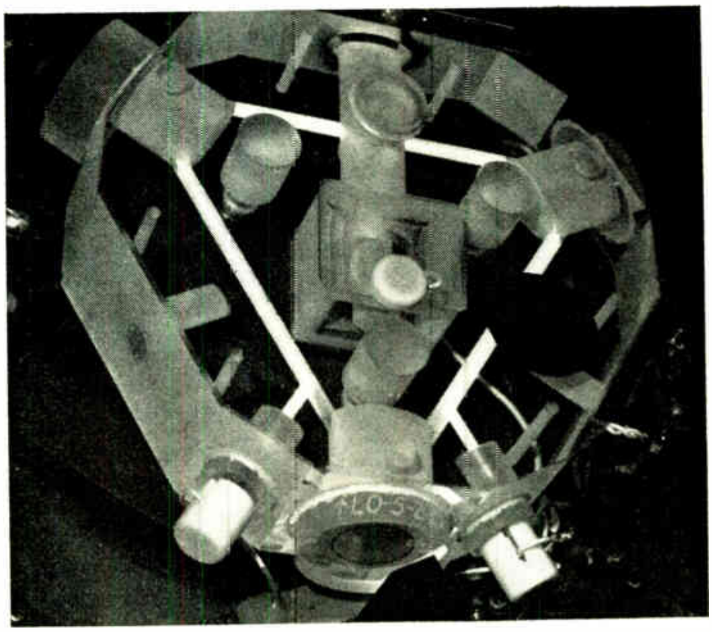
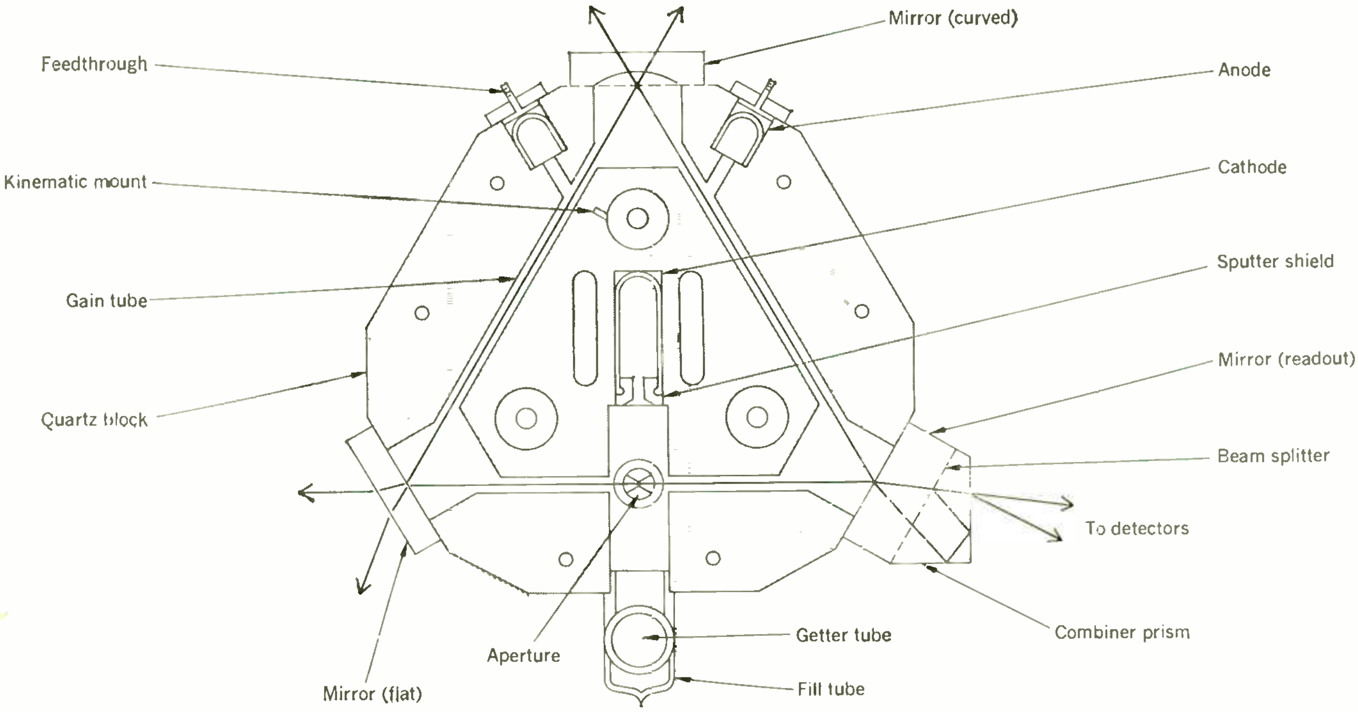


FIGURE 7. Single-axis laser gyro.

FIGURE 8. Construction detail of solid-block laser gyro.



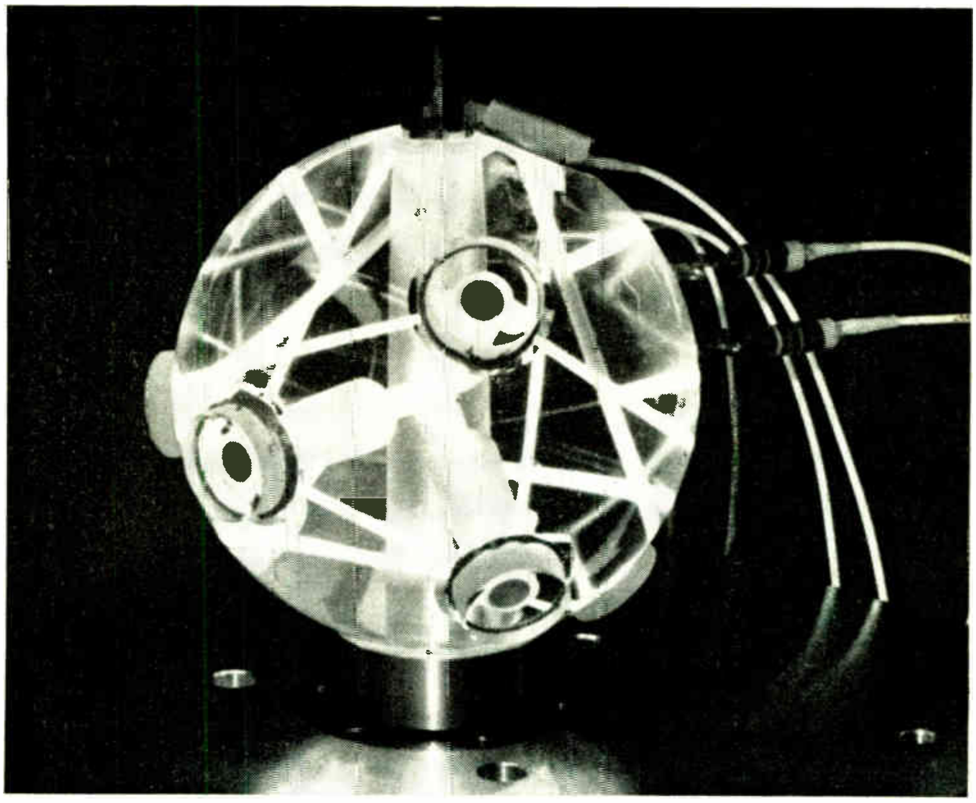


FIGURE 9. Three-axis laser gyro in solid quartz.

measurement is made. In general, the equation is written:

$$\Delta\phi = \frac{4A}{\lambda L}\int_{t_1}^{t_2} \Omega\cdot\mathbf{n}dt \qquad (15)$$

which describes the fact that the input is only that component of rate which is about the input axis of the gyro. The instrument is thus a single-axis rate-integrating gyro. Inertial rotations of the gyro about an axis orthogonal to the input axis will not create an apparent path difference, and thus will not produce any output fringe motion.

Construction of the laser gyro

The actual construction of the laser gyro varies from laboratory to laboratory, application to application, and with time as technology advances. (Figures 6 to 8 illustrate various laser gyro configurations.) However, certain elements must be observed in all cases:

Sealing. The laser cavity must be sealed, or shielded, from dust and other sources of scattering and from loss, which can affect performance in dramatic ways.

One transverse mode. The laser cavity must be designed, constructed, and aligned to support only one transverse mode of oscillation; the beam must have only one intensity maximum in its energy cross section. The reason for this is that the oscillations traveling in one direction may pick a mode distribution that is different from the mode distribution of those traveling in the other direction. The choice of mode distribution for either direction is random; it changes with vibration and small cavity perturbations if more than one transverse mode is permitted to exist. This switching, as well as the fact that these modes have different phase shifts and hence different frequencies, contributes to large noise fluctuations.

Mode position. The mode position relative to the basic laser gain distribution must be controlled to achieve maximum accuracy. The system must be simple, rugged, free of adjustments, and low in cost. The use of a solid block of quartz to define the laser cavity has proved quite successful. The anodes and cathode are attached to the housing, as are the corner mirrors defining the cavity dimensions. The entire gyro is filled with helium and neon and sealed. The gyro needs no alignment, and is both rugged and low in cost (Figs. 7 to 9).

The solutions to problems located to date do not detract from the basic simplicity or add substantially to the cost of the gyro. It is apparent that this instrument will be used in many future applications because of its many advantages over more conventional gyroscopes.

An accelerometer?

The question arises as to the use of this instrument as an accelerometer. The fundamental sensed property is a path difference arising from a rotation of a closed path in inertial space; therefore, linear acceleration and even large angular accelerations have negligible effects on the laser performance. Acceleration effects would arise only through a physical bending of the gyro housing that could break the assembly. Figure 5 shows the lack of acceleration effects as two runs are compared, one with no loading, and the other on a centrifuge with linear accelerations up to 23 *g*.

Error sources

The laser gyro is not perfect, and it has several factors that must be considered if it is to achieve the high performance necessary for many applications. It operates by sensing the relative frequency differences between two lasers occupying the same physical cavity and traveling in opposite directions. Therefore, any factor that can cause changes in the apparent length difference of the two cavities, other than inertial input rate, will introduce errors into the laser gyro output. The startling fact is that such terms equivalent to length changes of less than 10^{-6} Å can be identified in magnitude and source. The following discussion briefly reviews some of these error terms and the methods by which they are reduced or eliminated.

The predominant effect in the laser gyro is the shift in frequency due to input rates, with the error terms being perturbations in the basic sensed parameters. Accuracies of better than one part per million, and rate sensitivities of far less than one degree per hour, are important to the description of the laser gyro. Thus, strict attention must be paid to basic laser fundamentals. It is a mark of the rapid advance and understanding of the laser that equations and descriptions of this oscillator can be made to accuracies of far better than 0.1 Hz, even though the oscillator frequency is 3×10^{14} Hz.

Ultimate limits of performance

Before we explore the practical problems, it is important to examine the fundamental limit for the gyro, in order to establish how close to this limit we are today, and to determine whether basic factors exist that may restrict its ultimate potential.

The inherent frequency limit of a laser is determined by spontaneous emission from the laser gain medium, just as thermal noise determines the inherent limit in a conventional electrical oscillator. Each spontaneous-emission photon has a magnitude of $h\nu$, where h is Planck's constant and ν the frequency of the radiation. At a wavelength of 1 μm, each photon has an energy of 1.2 electron-volts ($h\nu = 1.2$ eV). When this emission of $h\nu$ occurs (it must be in the proper direction in order to interact with the laser mode energy), it may bear any phase relation with the oscillator energy, and thus may cause both magnitude and phase fluctuations in the laser output. The output power can be represented by a vector whose power magnitude is $Nh\nu$ per second, where N is the number of photons in the laser mode, a number equal to 10^{18} or greater. The spontaneous-emission photon can add a phase to this original oscillator of a magnitude of $1/N$. This phase shift, $\Delta\phi$, is quite small, but occurs in a random manner. This variation gives rise to a frequency spread in the laser output equal to[5]

$$\Delta f_n = \frac{8\pi h\nu(\Delta\nu_0)^2}{P_m} \qquad (16)$$

For typical conditions, this frequency spread Δf_n is 1 to 10^{-3} Hz, and arises from phase fluctuations as small as 10^{-18} radian that occur at rates up to 10^{18} per second. The spontaneous-emission photons occur about every 10^{-18} second, each one causing a phase fluctuation of about 10^{-18} radian in the oscillator energy. The frequency spread would be about 0.1 Hz:

$$\Delta f = \frac{\Delta\phi}{\Delta t} \doteq \frac{1}{2\pi}\,\frac{10^{-18}}{10^{-18}} \approx 0.1 \text{ Hz} \qquad (17)$$

However, since this is a random process, the actual measured frequency fluctuations will decrease with the square root of the time of measurement. Townes gives a clear illustration of this point,[6] and gives the following equation relating the apparent or observed line width $\Delta\nu_0$ in a oscillation of maximum line width $\Delta\nu_{\max}$ in a time t and with an oscillation power of P:

$$\Delta\nu = \Delta\nu_{\max}\left(\frac{h\nu}{Pt}\right)^{1/2} \qquad (18)$$

With conservative estimates for $\Delta\nu_{\max}$ of 1 Hz and a P of 10^{-4} watt, a measurement time of 1 second will measure frequency fluctuations of only 4×10^{-7} Hz.

In practice, the working limit of the frequency fluctuations is determined by the minimum phase that can be measured. This phase is usually sensed by the voltage on a detector or detectors that are measuring the relative phase between two oscillators (as in the laser gyro case). If a signal-to-noise ratio of 10^4 is obtained in the detected signal, the minimum detectable phase shift is 10^{-4} radian. In one second, the apparent frequency fluctuation caused by detector noise is approximately 10^{-5} Hz:

$$\Delta\nu_{\text{apparent}} = \frac{10^{-4}}{2\pi} \text{ Hz} \qquad (19)$$

This number is greater than the actual laser frequency fluctuation for this measurement time. For these particular parameters, the error caused by the phase fluctuation of the oscillator would not exceed the minimum detectable value until the time of measurement exceeded 10 minutes. The exact figures will vary from one laser configuration to another; but, for times up to several hundred seconds, the fundamental limit of rate measurement is always determined by the phase resolution. The measured fluctuation for times of only a few milliseconds would turn out to be about 10^{-4} Hz, if it could be measured. Since a rough rule of thumb is that 1 Hz is equivalent to an input rate of 1 degree per hour, the inherent capability of this gyro far exceeds 10^{-3} degree per hour. In fact, for periods of one hour, the fundamental accuracies exceed 10^{-7} to 10^{-9} degree per hour. Clearly, if such fundamental accuracies can be achieved, this gyro would be far superior to any gyro in operation today: An angle of 10^{-8} degree is the angle subtended by 10 cm at a distance equal to that from the earth to the moon, or 386 000 km.

Another, and perhaps more meaningful, expression of this fundamental limit is the total phase error between the two laser oscillators. The gyro output is the phase difference between the two oscillators; and therefore descriptions of this phase error can be directly related to measured outputs. This error will grow with the square root of time as in any random process, and will have a minimum value (as given in the previous example) of 10^{-18} radian when the time is less than about 10^{-18} second.

For times less than several days the inaccuracy in the laser gyro output phase is less than 10^{-3} radian, or about 10^{-4} arc second. (A phase error of one count, 2π radians, is equal to an inertial error angle of only one arc second.)

In summary, the gyro output, which is the phase angle between the two oscillators, is affected by the fundamental noise process of spontaneous emission by amounts so small that their detection is virtually impossible. A laser gyro system operating at this fundamental limit would be able to measure inertial angles for times up to several days, with errors not exceeding 10^{-4} to 10^{-2} arc second. In particular laser gyro applications these numbers may vary; but in all cases the laser gyro, as limited by this fundamental process, would be orders of magnitude more accurate than any existing gyro. It is doubtful that these limits will be reached in the foreseeable future.

After examining the most basic limit to the gyro performance (spontaneous-emission noise) and establishing that this term does not in any way limit the laser performance, those factors actually encountered must be considered as practical limits. These factors include lock-in, null shifts due to direct current, null shifts due to differential loss mechanisms, and the bias techniques used to overcome lock-in phenomena.

Lock-in

The first of these factors that affect the frequency difference between the two oscillators is a phenomenon called "lock-in," which arises from coupling between the two laser oscillators. Lock-in in a laser gyro was initially observed in early 1963,[7] and has been observed for many decades in conventional electrical oscillators.

When the output of the laser gyro is observed as a function of the rotation rate, it can be seen that the difference frequency is proportional to the input at high rates. However, as the input rate is reduced, the frequency difference between the two oscillators will fall to zero before the input rate goes to zero, which is apparent from the fact that the output fringe pattern does not move. The input rate at which this lock-in, zero-difference frequency, occurs is called the lock-in rate.

The lock-in rate is found to depend upon the wavelength, the coupling factor r, and the enclosed area A. Thus in order to reduce the lock-in rate these rules apply:

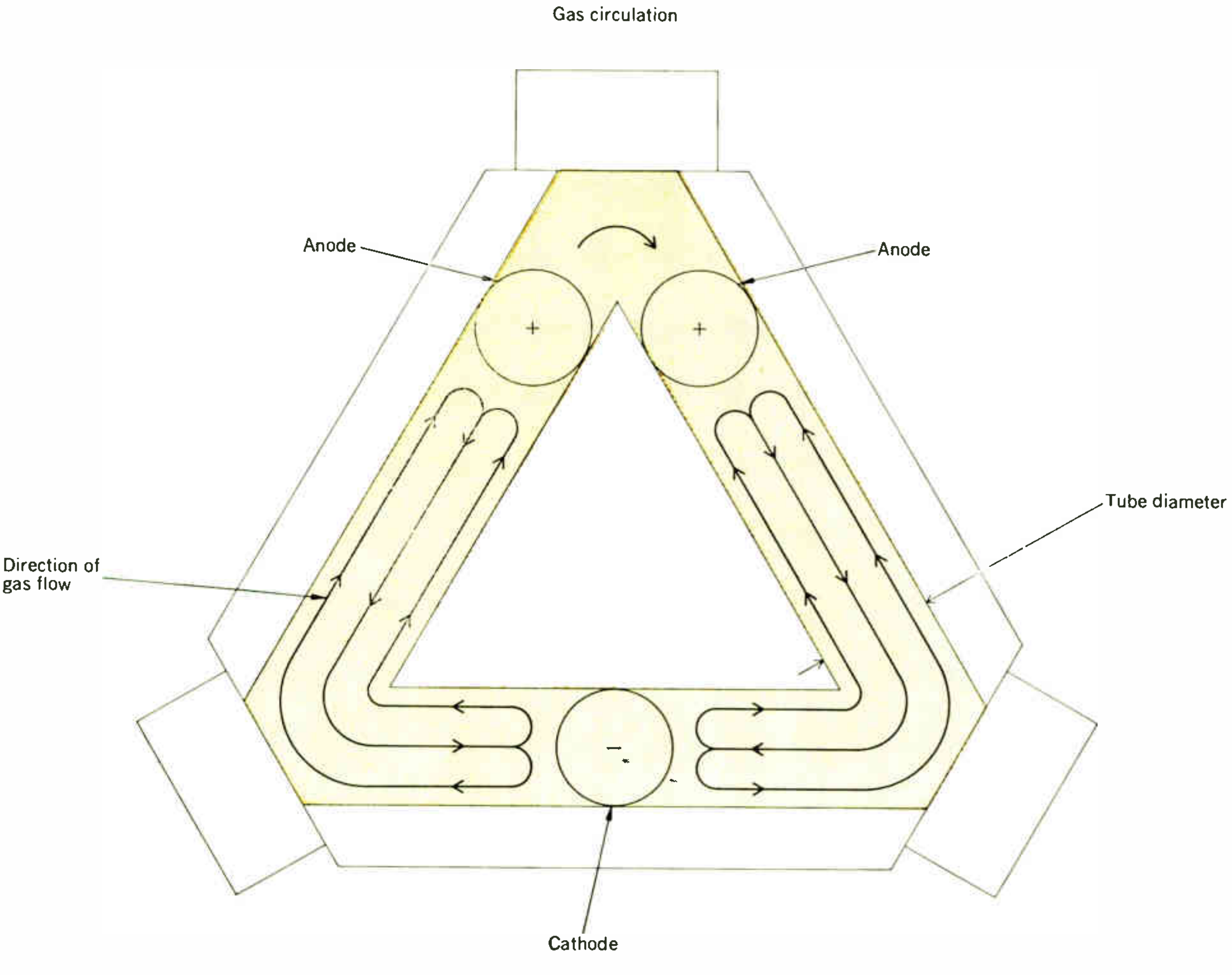
1. The smaller the wavelength, the better.
2. The bigger the gyro unit, the better.
3. The smaller the coupling factor r, the better.

In practice, the choice of wavelength is restricted to values at which laser oscillation can occur and by the fact that r is not completely independent of wavelength. The limiting values of r are determined by Rayleigh scattering, which increases rapidly as the wavelength decreases, thus prohibiting the application of rule 1 indefinitely.

Rule 2 also presents problems, since for practical considerations of size, weight, and power, the gyro should be as small as possible. Also, very large gyros create other laser locking effects that complicate operation.

Rule 3 has practical limits that are defined by the quality of the laser cavity. Scattering from windows, mirrors, and other elements within the cavity must be reduced to an absolute minimum. In practice, coupling factors of less than 10^{-5} have been achieved by very careful design and assembly. By examination of all the factors contributing to coupling, reduction in its value by many orders of magnitude does not look promising. Coupling arising from the scattering of the gas molecules in the laser gain medium will limit the value of r so that lock-in rates

FIGURE 10. Gas flow in a dc laser discharge



of less than a degree per hour may be very difficult to reach. (The word impossible would be used if it weren't such a dangerous technical term.)

In practice, the magnitude of the lock-in rate is typically 100 Hz, or 100°/h for a ring of about 0.1 m^2 enclosed area and a wavelength of 6328 Å (red light).

It should be emphasized that it is the scattering (coupling) that leads to lock-in, and not the loss or absorption of energy. (The fact that the gain or loss in a laser cavity is not uniform generates lock-in terms that are low in magnitude, and are usually hidden by other factors.)

Null shifts due to direct current

A second source of errors is the null shift, equivalent to a fixed bias torque in a conventional gyro, which arises from the direct current used to excite the laser gyro. Null shifts resulting from this parameter must both be considered and reduced. When a gas discharge is sustained with a direct current the gas flows in the discharge cavity. The flow is established by the combination of several effects, including wall collisions, charge distribution on the wall, and the electric field along the discharge. The result is a flow of gas toward the cathode in the center of the discharge, and a flow back toward the anode in a region close to the laser cavity walls (Fig. 10). The laser energy is concentrated in the center portion of the cavity, and therefore passes through gas that is flowing toward the cathode. The flow produces a shift in the index of refraction that depends upon the relative directions of the laser energy and the gas flow. Therefore, the cavity will appear longer in one direction than the other, and will cause an apparent null shift in the input rate sensed by the gyro. As can be seen from the diagram, Fig. 7, this effect is reduced by constructing the gyro in a balanced configuration, with a single cathode and two anodes. This tends to cancel the current effects, as energy traveling around the cavity passes through gas traveling both with and against the laser energy. By balancing the two anode currents, the null shifts due to this effect can be greatly reduced. If only one anode is used, apparent input rates of several hundred degrees per hour are introduced. This effect can be used to an advantage, since by unbalancing the currents a null shift can be introduced into the gyro either to cancel other null shift terms or to introduce known input rates purposely.

High rates

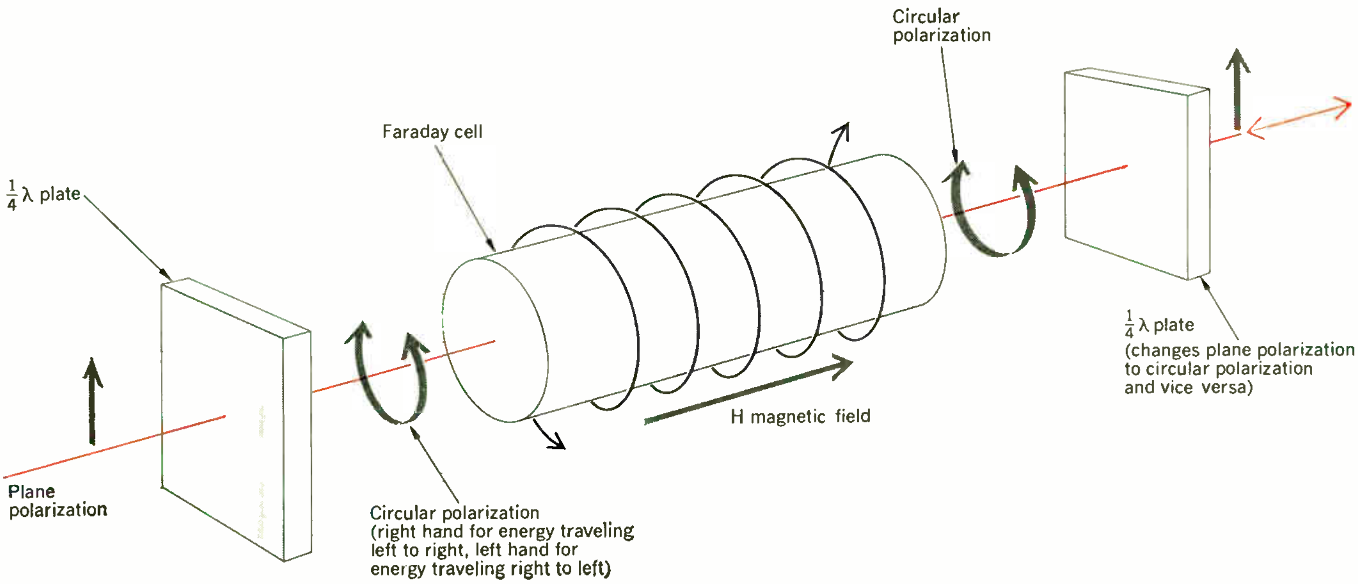
Another potential problem is high input rates; but this turns out to an advantage for the laser gyro, as it can sense high rates very accurately. The gyro's ability to sense high rates is determined by the inherent frequency capability of the laser and the detector–amplifier. The laser gain medium has a width of approximately 1 GHz. Since the frequencies in the laser change as the input rate changes, both frequencies must lie within the permitted gain region of the laser. Since an input rate of one degree per hour typically produces an output frequency difference of 1 Hz, an input rate of up to 10^8 degrees per hour could be sensed by this device. In practice, when the frequency difference becomes a large fraction of the gain distribution width, deviations in linearity are to be expected. For rates up to 10^7 degrees per hour, however, the laser gyro performance should be virtually ideal. Current test results support this statement, with performances at rates over 10^6 degrees per hour having been demonstrated. Figure 5 shows the excellent linearity for input rates up to 4 million degrees per hour.

The predicted problem is the detector–amplifier combination since, at 10^8 degrees per hour, bandwidths of 100 MHz are required in the amplifying circuits. Such bandwidths are achieved, but not without some very careful design and assembly. Here the high resolution works to our disadvantage.

Biasing of the laser gyro

Lock-in phenomena prevent the laser gyro from accurately measuring low input rates unless some means is provided to overcome this effect. Lock-in rates vary for gyros of different size, quality of the surfaces in the laser cavity, and wavelength of the laser oscillations; but for

FIGURE 11. Biasing through rotation, using Faraday cell.



practical sizes and other factors, lock-in rates of less than a few hundred to tens of degrees per hour, at the best, are observed. These lock-in rates agree with theory, and cannot be expected to drop rapidly in the future.

Clearly, in order to measure rates of less than a degree per hour, some means of avoiding this effect must be employed. One technique is to introduce another known, or effective, input rate to the gyro. Such biasing moves the operating point of the laser gyro away from very low rates, where lock-in occurs, to much higher rates, where the operation of the laser gyro approaches the ideal. The total input rate to the gyro is the sum of the true input rate plus the bias rate. The actual input can be measured by merely subtracting the known bias from the gyro output.

In general, biasing is accomplished by either of two ways. One method involves physically rotating the gyro, thus producing an input rate, and then subtracting the angle of rotation. A second method is to introduce into the laser cavity an optical element having a index of refraction dependent upon the direction in which the radiation is passing through the element. The magnitude and direction of this index difference are usually controlled by a magnetic field in a Faraday cell; see Fig. 11. Since the path around the laser gyro cavity is larger in one direction than the other, the laser gyro is biased away from the lock-in region.

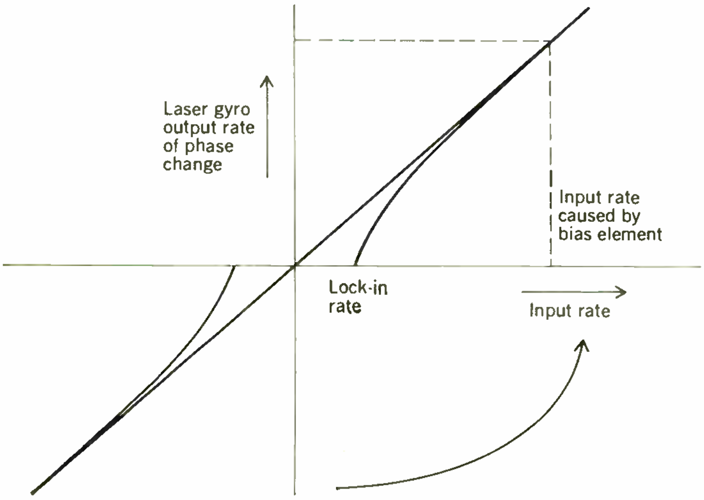
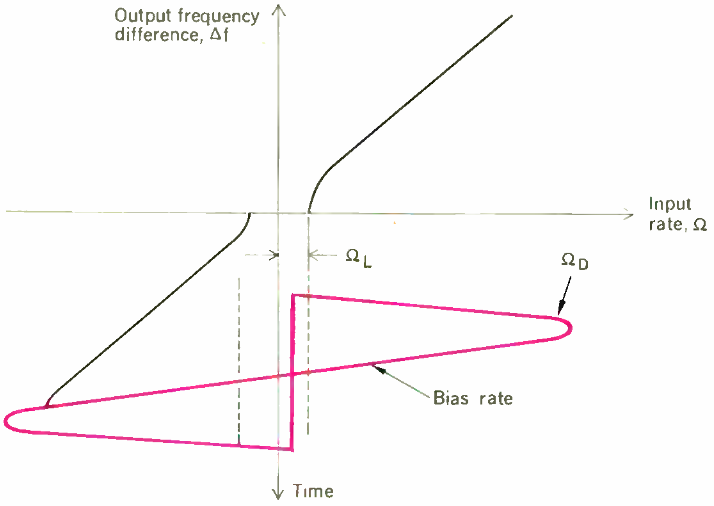


FIGURE 12. Laser gyro fixed bias.

FIGURE 13. Laser gyro alternating-bias technique.



Either of these general techniques can be employed in two ways: by holding the bias fixed, measuring the input rate or angle (Fig. 12) and observing differences in the laser output due to input rate; or by introducing a negative-to-positive alternating bias (Fig. 13).

One method of producing a fixed bias is to rotate the entire gyro, producing an input rate and angle, and thus biasing the gyro away from the lock-in region. The gyro output is proportional to this bias, plus any input rates to the entire gyro. Therefore, by subtracting the bias angle vs. time from the gyro output, only the desired input angle is left. This technique also can be employed using elements in the cavity to produce a fixed bias, and subtracting that bias in the output in order to measure the actual input rate and angle.

Fixed-bias techniques require strict stability, and are subject to errors arising from drifts or changes in the bias magnitude or the inherent gyro linearity. For example, an internal bias of 10^6 degrees per hour must be stable to one part in 10^8, to measure input rates of 10^{-2} degree per hour.

In an effort to reduce this absolute requirement for stability, the bias may be oscillated from positive to negative states. Since the gyro is an integrating-rate gyro, only the net rotation angle appears in the output. Therefore, the laser is biased out of the lock-in region most of the time, and the gyro operation integrates the input rate. This oscillation reduces the requirement for absolute magnitude stability. The bias technique can be employed using either optical means, such as varying the magnetic field of the internal bias element, or by mechanical means, such as oscillating the entire gyro.

One method of accomplishing this ac bias is to oscillate (dither) the entire gyro at a frequency of 20–40 Hz, at an amplitude of several hundred arc seconds. Dithering can be accomplished by a number of electromechanical methods, causing the input rate to the gyro to vary sinusoidally from a maximum positive value to a maximum negative value. An additional input to the gyro will add to this bias rate, increasing its value during one half of the bias cycle, and decreasing its value during the other half. Since the laser gyro is an integrating-rate gyro, the bias integrates to zero. The output, due to the input rate, will appear as a net accumulation of the output cycles. The slope of this accumulation of counts, or angle, is the input rate. Experiments and techniques for biasing are being pursued in many laboratories; the bias technique is one of the most important design aspects of the laser gyro.

Other problems

In addition to these problems of lock-in, current balance, etc., there are several other factors that must be considered. The role of multimoding* is particularly important in the operation of the laser gyro; the role of differential loss and gain factors can be of dominant importance in some laser gyro systems.[8,9] There are other fac-

---

* Multimolding is the phenomen of more than one frequency (mode of oscillation) traveling around the cavity in each direction.

tors—such as linearity, stability, and alignment—each of which is an interesting discussion in itself. For our purposes it is important to note that these additional problems and factors all appear to have solutions in keeping with the general objective of a simple, low-cost device.

The future

At present, the laser gyro is in advanced research and development, making promising progress toward an accurate, low-cost instrument. The fate of the laser gyro is difficult to predict beyond three to five years in the future; but it is clearly not limited by fundamental considerations (basic line-width analysis shows that the gyro has a limit of less than 10^{-6} degree per hour), but by construction, cleanliness, bias techniques, and other factors. The progress of the laser gyro, as measured by its ability to measure low rates accurately, has been quite rapid. Since the first announced laser gyro, in early 1963,[7] of 1 m² in area and a threshold of 50 degrees per hour, the published performance has improved to about a 0.1 degree per hour, with gyro areas more than an order of magnitude smaller than the first gyro. Further improvements are to be expected, as well as additional problems that have not as yet been identified or solved. The future should see laser gyros in many applications requiring low cost, lack of acceleration sensitivity, and high rate capability; such requirements will make the laser gyro the indisputable choice over other, more conventional, sensors.

REFERENCES

1. Landau, L., and Lifshitz, E., *Classical Theory of Fields.* Cambridge, Mass.: Addison-Wesley, 1951.
2. Post, E. J., "Sagnac effect," *Rev. Mod. Phys.*, vol. 39, pp. 475–493, Apr. 1967.
3. Michelson, A. H., and Gale, H. G., "The effect of the earth's rotation on the velocity of light," *Astrophys. J.*, vol. 61, p. 140, Apr. 1925.
4. Rosenthal, A., "Regenerative circulatory multiple-beam interferometry for the study of light-propagation effects," *J. Opt. Soc. Am.*, vol. 52, pp. 1143–1148, Oct. 1962.
5. Schawlow, A. L., and Townes, C. H., "Infrared and optical masers," *Phys. Rev.*, vol. 112, pp. 1940–1949, 1958.
6. Townes, C. H., "Some applications of optical and infrared masers" in *Advances in Quantum Electronics*, X. X. Singer, ed. New York: Columbia University Press, 1961, pp. 9–11.
7. Macek, W. M., and Davis, D. T. M., Jr., "Rotation rate sensing with traveling-wave ring lasers," *Appl. Phys. Letters*, vol. 2, pp. 67–68, 1966.
8. Lee, P. H., and Atwood, J. G., "Measurement of saturation induced optical nonreciprocity in a ring laser plasma," *IEEE J. Quantum Electronics*, vol. QE-2, pp. 235–242, Sept. 1966.
9. Hutchings, T. J., Winocur, J., Durrett, R. H., Jacobs, E. D., and Zingery, W. L., "Amplitude and frequency characteristics of a ring laser," *Phys. Rev.*, vol. 152, pp. 467–473, 1966.

BIBLIOGRAPHY

Adler, R., "A study of locking phenomena in oscillators," *Proc. IRE*, vol. 34, pp. 351–357, June 1946.

Aronowitz, F., "Theory of traveling-wave optical maser," *Phys. Rev.*, vol. 139, pp. A635–A646, 1965.

Aronowitz, F., and Collins, R. J., "Mode coupling due to backscattering in a He-Ne traveling-wave ring laser," *Appl. Phys. Letters*, vol. 9, pp. 55–58, 1966.

Bagaev, S. N., Kuznetsov, V. S., Troitskii, Yu. V., and Troshin, B. I., "Spectral characteristics of a gas laser with traveling wave," *JETP Letters*, vol. 1, no. 4, pp. 114–116, 1965.

Bagaev, S. N., Troitskii, Yu. V., and Troshin, B. I., "The polarization of radiation and the frequency characteristics of ring lasers with a triangular resonator," *Opt. Spectr. USSR*, pp. 768–769, 1966.

Basov, N. G., Belenov, E. M., and Letokhov, V. S., "Diffraction synchronization of optical masers," *Sov. Phys.-Doklady*, vol. 10, pp. 236–238, 1965.

Batifol, E., and Pecile, D., "Effect of the flow of the active medium of a ring gas laser," *Phys. Abstr.*, vol. 72, pp. 3128–3146, 1967.

Belenov, E. M., Markin, E. P., Morozov, V. N., and Oraevskii, A. N., "Interaction of traveling waves in a ring laser," *JETP Letters*, vol. 3, pp. 54–58, 1966.

Bershtein, I. L., "Frequency-pulling band of a laser oscillator," *Sov. Phys.-Doklady*, vol. 10, pp. 607–609, 1966.

Bershtein, I. L., and Zaitsev, Yu. I., "Operating features of the ring laser," *Sov. Phys. JETP*, vol. 22, pp. 663–667, 1966.

Bertein, F., and Petit, F., "Sur l'effet de perturbations apportées aux miroirs dans un laser," *Compt. Rend. Acad. Sci. (Paris)*, vol. 259, pp. 2980–2983, 1964.

Bridges, T. J., and Rigrod, W. W., "Output spectra of the argon ion laser," *IEEE J. Quantum Electronics*, vol. QE-1, pp. 303–308, Oct. 1965.

Catherin, J. M., and Dessus, B., "Traveling wave laser gyrocompass," presented at 1967 Conf. on Laser Engineering and Applications, Washington, D. C., June 6–9.

Christiansen, D., "Laser gyro comes in quartz," *Electronics*, 1966.

Collins, S. A., Jr., and Davis, D. T. M., Jr., "Modes in a triangular ring optical resonator," *Appl. Opt.*, vol. 3, pp. 1314–1317, 1964.

de Lang, H., "Derivation of the relation between two weakly coupled nonlinear optical oscillators," *Appl. Phys. Letters*, vol. 9, pp. 205–207, 1966.

Dessus, B., and Catherin, J. M., "Effets non reciproques et couplages dans un laser en anneau," *Comp. Rend. Acad. Sci. (Paris)*, vol. 262, pp. 1691–1694, 1966.

Fenster, P., and Kahn, W. K., "Single-beam-laser rotation-rate sensor," *Electron. Letters*, Sept. 1966.

Hercher, M., Young, M., and Smoyer, C. B., "Traveling-wave ruby laser with a passive optical isolator," *J. Appl. Phys.*, vol. 36, p. 3351, 1965.

Itzhan, I., "A circular ring laser," *Proc. IEEE (Correspondence)*, vol. 53, p. 164, Feb. 1965.

Khromykh, A. M., "Ring generator in a rotating reference system," *Sov. Phys. JETP*, vol. 23, pp. 185–186, 1966.

Klass, P. J., "Laser unit challenges conventional gyros," *Aviation Week*, Sept. 12, 1966.

Klimontovich, Yu. L., Kuryatov, V. N., and Landa, P. S., "Wave synchronization in a gas laser with a ring resonator," *Sov. Phys. JETP*, vol. 24, pp. 1–7, 1967.

Letokhov, V. S., and Markin, E. P., "On the statistics of laser emission," *Sov. Phys. JETP*, vol. 20, pp. 509–510, 1964.

Macek, W. N., Schneider, J. R., and Salamon, R. M., "Measurement of Fresnel drag with the ring laser," *J. Appl. Phys.*, vol. 35, pp. 2556–2557, 1964.

Podgorski, T., and Aronowitz, F., "Langmuir flow effects in the laser gyro," presented at 1967 Conf. on Laser Engineering and Applications, Washington, D.C., June 6-9.

Post, E. J., and Yildiz, A., "Cavity resonances in accelerated systems," *Phys. Rev Letters*, vol. 15, pp. 177–178, 1965.

Proctor, E. K., "On the minimum detectable rotation rate for a laser rotation sensor," *Proc. IEEE (Correspondence)*, vol. 51, p. 1035, July 1963.

Raterink, H. J., Stadt, H. v. d., Velzel, C. H. F., and Dijkstra, G., "Development of a ring laser for polarimetric measurements," *Appl. Opt.*, vol. 6, pp. 813–820, 1967.

Rigrod, W. W., "The optical ring resonator," *Bell System Tech. J.*, pp. 907–916, 1965.

Rigrod, W. W., and Bridges, T. J., "Bistable traveling-wave oscillations of ion ring laser," *IEEE J. Quantum Electronics*, vol. QE-1, pp. 298–303, Oct. 1965.

Smith, R. C., and Watkins, L. S., "A proposed method for reducing the locking frequency of a ring laser," *Proc. IEEE (Correspondence)*, vol. 53, p. 161, Feb. 1965.

Tang, C. L., and Statz, H., "Phase-locking of laser oscillators by injected signal," *J. Appl. Phys.*, vol. 38, pp. 323–324, 1967.

Thomson, A. F. H., and King, P. G. R., "Ring-laser accuracy," *Electron. Letters*, Oct. 1966.

Walsh, P., and Kemeny, G., "Laser operation without spikes in a ruby ring," *J. Appl. Phys.*, vol. 34, pp. 956–957, 1963.

White, J. A., "Stability of traveling waves in lasers," *Phys. Rev.*, vol. 137, pp. A1651–A1654, 1954.

Zaitsev, Yu. I., "Radiation fluctuations in a gas laser," *Sov. Phys. JETP*, vol. 23, pp. 349–354, 1966.

Zhelnov, B. L., Kazantsev, A. P., and Smirnov, V. S., "Mode interaction in a gas laser," *Sov. Phys. JETP*, vol. 23, pp. 1291–1295, 1966.

Zhelnov, B. L., Kazantsev, A. P., and Smirnov, V. S., "Stimulated traveling-wave radiation in lasers," *Sov. Phys.-Solid-State*, vol. 7, pp. 2276–2279, 1966.

Television's role in tomorrow's world

Any prediction concerning the future of communications must recognize today's trend toward unification of the various electronic media—which suggests that we should revise our traditional idea of television as a broadcasting service

George H. Brown

Radio Corporation of America

In this discussion, which was originally presented at the Fifth International Television Symposium, the author, in his role as prophet, suggests at the outset that the concept of such a meeting may be doomed for obsolescence; a more apt name for the next symposium might well be the First International Information Conference. He points out that a unified system of electronic communications media, combined with a data processing system built around the computer, may well be the principal information source of the future. Viewed in such a context, tomorrow's television techniques promise many benefits, both for the industry and for the consumer.

There are certain rules that govern successful forecasting. For example, it is best to cast any specific prediction so far into the future that if it should fail to come true, no one will remember just what was originally said. And if one is forced to make near-term predictions, they must be couched in ambiguous language to allow freedom of interpretation when the future becomes the present.

One seldom can scan a newspaper or magazine these days without finding at least one article describing the world of tomorrow. I have observed, however, that almost all the devices and systems that are imagined each day actually depend very little on future research, but could be devised with our present technology. The common ingredients needed to obtain results in every case are money for design and development and a conviction that the object of our attentions, the potential customer, will exhibit an overwhelming desire to partake of the bounty we wish to provide. For example, much of the basic research has already been done for the really significant changes in communications technology that will come about during the next ten years. In view of the acceleration of research and engineering, almost anything we can imagine will be feasible after that time.

The most notable feature of the communications technology today is its tendency to abolish the traditional distinctions among the various electronic media: television, radio, telephone. Instead, we are moving toward a unified system that brings them all together, and adds another factor—the electronic data-processing system built around the computer. This unified system may well be the principal information medium of a future that is not far away. When that future arrives, our seemingly complex television technology will become just one part, although a very important part, of something much more sophisticated.

This situation raises a number of interesting possibilities, and it is not too early to begin adjusting our outlook to a whole new set of circumstances for the television art and industry.

Television's dual nature

Television is really two very different things. It is a mass medium for public broadcasting, and it is a specialized medium of closed-circuit communications. These forms are distinct from one another in concept and purpose and they will be affected differently by the changes that are coming. However, they do partake of the same technology and they will gain similar advantages from progress in equipment and techniques.

The outlook for the technology is fairly clear in view of the present trends in research and engineering. It is a safe guess that solid-state innovations are going to bring about still smaller, lighter, and more flexible types of equipment in every part of the system, from camera to receiver. In another decade, we should have tubeless television cameras, thin-panel receivers, and solid-state transmitters no larger than filing cabinets.

The reach of television transmission will probably be extended to its limits around the world. Synchronous satellites already can handle a single channel across the Atlantic and Pacific Oceans. The new models, to be launched in the early 1970s, will be able to carry scores of channels and have room left over for thousands of voice-grade channels for other services. Program distribution through satellites therefore will be feasible on a large scale, both domestically and internationally.

This development is likely to have interesting effects on the broadcasting industry. Broadcasting is generally regarded today as an entertainment medium, an aspect that almost everywhere outweighs its use as a means of instruction, analysis, or practical service—such as weather forecasting, farm information, shipping news, and traffic reports. But when broadcasting becomes so universal and so pervasive, there will be both the demand for and the facilities to take care of all of the more specialized needs. The new technology is going to generate new concepts of programming, and it is difficult to predict which of the two trends will produce more radical innovations.

Closed-circuit television is already giving a virtuoso demonstration of the new technology. Solid-state circuit and miniature devices have extended human vision beyond the earth to the moon and Mars. Before another decade has passed, there will be more advanced television systems—and the vehicles to carry them—for on-the-spot inspections of Mercury, Venus, and Jupiter. Then, within another few years, we will probably carry on visual two-way communication with men exploring these bodies.

Although these are the applications that will make headlines, I believe that the greatest impact of closed-circuit television will be felt on a more mundane level. The great channel capacity of satellites, and cheap, miniaturized cameras and receivers, can make of television a communications utility equivalent in its common use to the telephone or two-way radio services.

One need not be much of a prophet to foresee sweeping effects upon education, to choose the most notable example of closed-circuit usage. Today, television extends the reach of a single teacher to scores of classrooms simultaneously. Tomorrow the same teacher will be able to reach out to other continents.

I believe that two-way television services will become both practical and common in some areas of education. When this happens, there will be an added facility for direct discussion or for questions and answers involving students and teacher. The only real barriers seem to be economic. We already know fairly well how to accomplish these things, but our technology still has some distance to go before we can bring the equipment down to a reasonably inexpensive level.

I need not enumerate the many other familiar uses of closed-circuit equipment in industry, medicine, government, and research. But for every present application, there are likely to be several more tomorrow. All of them will be able to reach, through satellites or wide-band cable, over any distance on earth.

The new technology

So far, I have been discussing television essentially in its present guise, as a distinct form of broadcasting and communication service. Yet, as I indicated earlier, we are not going to be able to think about television in just this way very much longer. All that is happening in this field fits into the context of changes that are shaking the foundations of all communications technology.

The further development of synchronous satellites is going to remove any serious limits on the volume of information that can be transmitted domestically or internationally over wide-band channels. There will be no technical barrier to a practically unlimited flow of data, voice, message, television, and other traffic over and among all of the continents.

Coincidentally, the computer is evolving into a potent communications instrument. Computers are working their way into the terminals of communications systems. They sort and switch messages at extremely high speed in immense volume, providing a fortunate complement to the expansion of channel capacity through satellites.

Computers are also taking on a number of new information-processing jobs in business management, foreign trade, education, industrial control, and a host of other daily activities. They are starting to communicate with each other in their own esoteric languages over the wide-band channels. A central computer at the focus of a communications network can now provide service to a large number of scattered users simultaneously by means of time-sharing techniques. This is an innovation that we may confidently expect to grow rapidly in scope and volume.

The association of the computer and closed-circuit television also foreshadows many significant developments. The two are natural partners, because television can act as an organ of vision whereas the computer supplies a rudimentary form of intelligence. The principle was demonstrated in rather breathtaking fashion by the Mariner spacecraft that flew past Mars. To save bandwidth and thus conserve power, the images of the Martian surface were transmitted to the earth in digital form over long periods of time. A computer received the information and used it to reconstruct the pictures. Computers also have enhanced the televised pictures from the Surveyor vehicles on the moon, giving us far more information than could have been extracted by the unaided eye.

Electronic imaging devices that incorporate television principles are now used to read characters, such as the writing on bank checks, and signal the information to computers for automatic adjustment of accounts, and even for detection of forgeries. Specialized adaptations of the television picture tube have been coupled with electronic printing equipment, and the results show promise of coping with the increasing output speeds of computers.

Evolution of the information center

I believe we are only at the threshold of many major developments that will see the combination of television and computers for complex functions that involve the sequence of sensing, analysis, information processing, and printing.

To round out the picture, hovering around the edges of these systems is a growing swarm of lesser devices that offer new ways of converting one form of information to another, devices such as image intensifiers, tape storage units, and analog-to-digital converters. They have in common the ability to receive and process data in various ways in order to pass it on to another system or to present it in a desired form.

This ferment offers an embarrassment of riches to the communicator who continues to think in terms of present conventional technology. It makes better sense, however, when we begin to piece everything together in the new concept of a unified system that can accept any type of information in any form—print, voice, or picture—and transmit it over any distance to be used for any reasonable purpose in any other desired form.

The ultimate expression of this concept is a comprehensive information center, compact enough to fit into an office or a home, and serving all information and communication purposes—from telephone-cum-television communication to delivery of electronically printed papers and books. Such a center could perform the functions that now require a television set, radio, and telephone, and it could furnish a multitude of new services through direct connections to banks, utilities, computer centers, and libraries. And this comprehensive information center can be expected to have many offshoots in the form of other personal and business services combining two or more communication and computing techniques.

When we do reach the point of unified communication, it seems unlikely that any of our present methods will be utilized to any great extent as independent systems. Even the broadcasting industry will be able to diversify its services with the help of electronic printing and other information storage and reproduction techniques.

The consequences of progress

Although it is tempting to continue to indulge in prophecy because the theme has endless variations, it is time to get down to the serious business of considering the impact of the developments I have described.

This is a more complex matter than we are sometimes led to believe. Many ardent advocates of progress imply that new technology showers its blessings equally and simultaneously upon all lands and all people. According to this view, we have only to distribute the latest technological achievements around the world and a golden age will begin.

Things do not work out in quite this way, however. Even now one may catch sight, in an odd corner somewhere, of a rusting tractor that was pressed upon a society that really needed steel edges for its wooden plows.

The fruits of a new technology are determined by the soil in which it is planted. For example, a truly sophisticated advance, such as computer programming of machine tools, is critically important to a highly industrialized society; it is largely meaningless anywhere else. And so it is in the realm of future communications. Their impact will be great everywhere, but it will take very different forms in different regions.

In the industrialized nations of North America, Europe, and the Far East, I expect that we shall see rapid progress toward the more sophisticated use of television, computers, satellites, and all the other paraphernalia of the budding information revolution. It is in these parts of the world that the unified information system will first emerge. Those who now provide television products and programs will find themselves dealing in new kinds of mass and individual communications. Within the next decade they should begin to branch out into a number of interesting new services that will impinge upon broadcasting: electronic newspapers and magazines broadcast or piped into the home; educational systems for personal and group instruction; computer-oriented information services for many professional users.

It is hard to surpass this concept in terms of sophistication, and the consequences will be extremely important. Yet the greatest impact of the new technology will probably be felt at the other end of the scale, with the use of simpler equipment in the developing nations of Africa, Asia, and Latin America. In these regions, the new satellite channels, and better and cheaper terminal and receiving equipment, can be expected to speed the education that is so urgently needed for progress. All of the growing information complex—programmed instruction, central electronic reference facilities, broadcasting from satellites—can be mobilized to build up literacy, knowledge, and skill where they are lacking.

The possibilities present an immense challenge today, simply because there are so few facilities in many of these areas to provide even the most elementary information services. According to a UNESCO estimate, the basic requirements for a country to supply adequate information to its people include at least ten copies of one daily newspaper, five radio and/or television receivers, and two cinema seats per every 100 persons. These standards rule out all of Southeast Asia as well as large parts of Africa and South America.

One of the consequences of the lack of information services in these regions is widespread illiteracy. Under such conditions, where, in some cases, more than three fourths of the population is illiterate, it would seem likely that electronic sight and sound will take the place of books and newspapers in educating and informing. I would expect that broadcast and closed-circuit television will lead the way for some years to come in communicating throughout the developing nations, using satellite relays, simplified terminal equipment, and low-cost receivers. Visual presentation is still the most effective means for communicating the most information to the largest number of people in ways they can most easily understand.

The outlook for the future

These are the highlights that I see in looking into the future—and I imagine that they offer enough challenge to satisfy the most energetic among us. Obviously, there is much that we can do to improve the techniques and services of television—in broadcasting, in education, and in the growing multitude of closed-circuit uses. Now there is also room for fresh concepts that place television into the context of the change that is sweeping through the whole structure of communications and information.

There are great opportunities in both directions. There are even greater benefits, both for the television industry and for those to whom the new systems will bring better ways of learning, improved methods of business and commerce, and greater mutual understanding.

Essential text of a paper, "The Outlook Into the Future," presented at the Fifth International Television Symposium, Montreux, Switzerland, May 26, 1967.

Compatible EMI filters

Conventional thinking has to be discarded when one is tackling electromagnetic-interference power-line filters. Active, lossy, and ceramic filters offer sensible solutions to problems in this often-frustrating field

H. M. Schlicke, H. Weidmann *Allen-Bradley Company*

A vexing problem in electromagnetic compatibility is the effective filtering of conducted interference from power-supply lines. Because of unavoidable and severe mismatch, conventional suppression filters operate only conditionally; such filters are often so large they are omitted from the system. Three new classes of power-line filters without these limitations are described. They are active, truly lossy, and ruggedized ceramic filters, covering the frequency range from direct current to microwaves. Present filter test methods are shown to be misleading and are replaced by a rather simple, realistic test.

There is a set of filter classes that is rapidly growing in importance and, by necessity, is characterized by the absence of impedance matching. Without matching, all the elegant filter theories developed[1] invalidate the very premise upon which they were based and the theories are wholly inadequate and misleading. Such conditions exist for filters inserted into power lines; power wiring is contrasted with impedance-matched cabling that interconnects subsystems for information handling and for which conventional filtering is fully adequate.

Because wires supplying energy are all-pass networks and pervade systems or interconnect subsystems, such wires (unless shielded) act as antennas for external electromagnetic fields. They may spread electromagnetic energy to places where it is undesirable and dangerous; compatible filtering then becomes mandatory.

Suppression of conducted EMI (electromagnetic interference) constitutes an essential part of EMC (electromagnetic compatibility) and has thus far been frustrating and inadequate. Conventional EMI or RFI (radio-frequency interference) filters, because of interface mismatching, are in general undependable. Before discussing EMI filters, we shall survey briefly the field of electromagnetic compatibility to put filtering for EMI in the proper perspective.

There is a great and growing pollution of the electromagnetic space. This national resource has far too long been taken for granted; hence, our cities are blanketed not only with sickening smog but also with dangerous electromagnetic fog. Complete electronic systems, particularly complex military systems, are disrupted by EMI; explosion hazards occur because of EMI; people carrying pacemakers for their hearts are killed by EMI; and television viewing is marred by EMI. The appalling consequences of EMI can here be only hinted at, but they are so significant that IEEE has formed an Electromagnetic Compatibility Group, which holds annual national symposia on this subject. Although many European countries have already enforced strict EMI codes, in the United States the proposed Public Law S-1015* has to pass the Senate again.

The vast field of electromagnetic compatibility is rather

---

*Section 3.02 (b) reads: "No person shall manufacture, import, sell, offer for sale, ship, or use devices which fail to comply with regulation promulgated pursuant to this section."

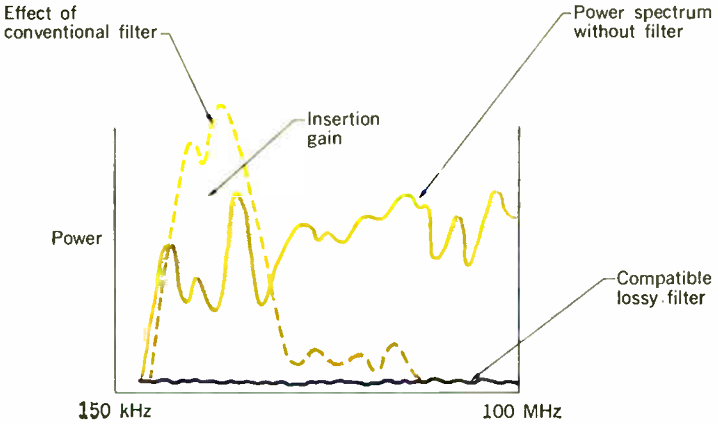


FIGURE 1. A conventional EMI filter can be worse than no filter when it causes insertion gain.

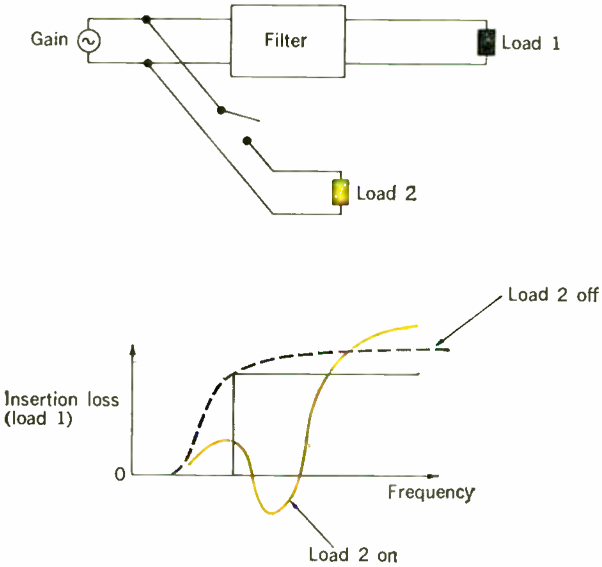


FIGURE 2. The effect on insertion loss when adding a load to the generator caused by an inoperative filter.

perplexing to the uninitiated. (As we shall demonstrate, test methods on EMI elimination are still in the archaic state.) From the beginning, the systems engineer not familiar with EMC should rely on experienced EMC men to make the system safe from electromagnetic interference; redesign can be very costly if at all possible.

The phases of EMC encompassing frequency allocation, interference prediction, shielding, etc., will not be discussed here. Rather, we shall confine ourselves to methods of broadband filtering of noise or information from power lines; this is a field of great confusion and frustration in EMC. Further, we must be made aware that filtering is not to be considered as an independent entity, arbitrarily divorced from the whole EMC complex.

Broadband noise to be filtered either may consist of many random narrow-band sources or be the result of switching transients and digital modes of operation. In the extreme, an impulse function has a power spectrum of white noise and any step function has a power spectrum inversely proportional to frequency. How can the effects of such broadband noise be eliminated? Three principal approaches, in part complementing each other, are possible:

1. Do not generate it; typical remedies are the replacement of electronic control by fluidic control or the substitution of electronic garage-door openers (the super-regenerative receivers of which play havoc with aircraft navigation near airports) by optical garage door openers In short, eliminate the use of electricity wherever possible.
2. Immunize the system against EMI; typical examples are clipping, FM, PCM, and correlation techniques.
3. Confine interference; it may be confined to its source by having a common ground point instead of a common ground wire, which produces coupling between circuits or systems. In addition, shielding is recommended for preventing radiated interference and filtering for preventing conducted interference.

Shielding and filtering form complementary methods. Because power lines (as defined here) pass through shielded enclosures, the lines must have filters mounted in the shields to prevent the shielding effect from being negated. Conversely, a filter may be completely bypassed by radiation coupling between input and output; for high frequencies, this effect can be prevented only by feeding the filter through the shield. Power-line filters differ from conventional filters with respect to four essential criteria:

1. They are not impedance matched. Power supplies and loads have no other purpose than to be highly efficient at the power frequency. To expect the terminations (generator and load) of the line to match its characteristic impedance over a large frequency band is wishful thinking. To satisfy some inapplicable theories, convenient but specious test methods were devised to make them appear correct. But only recently has the presumptive mistake been fully recognized (more about this later).
2. The suppression of EMI has to be accomplished over many decades of frequency, in some cases, up to 11 decades. Over such wide frequency ranges, one cannot expect that the filter elements will behave as predicted.
3. The dc or ac power, which has to be fed through the filter, constitutes strong biasing effects not found in ordinary filter applications.
4. Size is an important practical consideration when many filters are involved; for example, in multipin connectors, filters have to accommodate extremely small spaces. Other size problems arise for very low frequencies, where unmanageable large inductors and capacitors would be required for passive filters. In order to achieve small size, magnetic materials of high permeability and dielectric materials of high dielectric constants are often used; provision, however, must be made to prevent saturation by current or voltage.

We can now sum up the criteria for compatible power-supply-line filters: they must work under highly nonideal conditions and they must be small and light. Because both qualities are often not met by available devices, we shall treat the field of compatible EMI filtering systematically from the viewpoint of guaranteed effectiveness (stiffness) and small size. The discussion will be concerned with basic underlying principles only, not with the often critical details.

Three generic types of filters are available:

1. Miniaturized ceramic filters, which permit considerable size reduction for very-high-frequency low-pass

filters and capacitors for medium-frequency filtering. Their practicability is premised upon ruggedness and high reliability.

2. As contrasted with conventional reactive filters, lossy filters (and, in a limited way, lossy lines), which work unconditionally into all interface impedances.

3. Active EMI filters, which offer the only possible approach to obtain reasonable size where very low frequencies have to be filtered under strong power-bias conditions.

We shall start with type 2, where the primary concern is stiffening the filter against the degrading effects of mismatch, or, more exactly, detuning. Types 1 and 3, which involve very high and very low frequencies, have in common the need for size reduction. Mismatch conditions play only a secondary role because the values of Q are smaller at the frequency extremes than in the center region.

The mismatch dilemma

The following statements are premised upon supposedly perfect conventional power-supply-line filters, passing the presently prescribed and accepted tests (50-ohm system, MIL-ST-220A); that is, we exclude all poorly designed filters. Nevertheless, such certified filters may fail miserably in actual use. In the first of two situations, shown in Fig. 1, a line filter fails, particularly at the lower radio frequencies, and may result in a negative insertion loss. A filter with negative insertion loss may be worse than useless. Difficulties of this kind are now officially acknowledged.[2]

Figure 2 illustrates the second situation. Initially, the filter performs satisfactorily; but when another load is switched on (or off) the generator, the filter becomes inoperative. Negative insertion loss may again occur. A modification in load equipment, so prevalent in complex gear, may cause a similar deterioration of the assumedly good filtering.

Anything but 50 ohms. The reason for such seemingly mysterious behavior is that detrimental resonances may occur when the reactances at the filter terminals meet equal reactances of opposite sign at any frequency within the alleged stopband of the filter. Power generators, their loads, and their connecting lines, are designed by electric power engineers, who are interested in achieving high efficiency at the power frequency. If, on the other hand, RF engineers had to design the power-supply system from their point of view, namely, to meet the premises of their elegant filter theories, they might come up with a good RF system of constant generator and load impedances with corresponding transmission lines of the right characteristic impedance. However, the dc or ac efficiency would be poor; hence, the RF engineer must concede the primary objective and let the power engineer build the system according to his goals. As will be shown, the RF engineer does not really have much of a task in solving this seemingly perplexing problem, if he is willing to give up his conventional thinking.

The impedance seen by the filter may be practically anywhere on the Smith chart, as indicated in Fig. 3, which is approximately correct for frequencies near 150 kHz. For much lower frequencies, the center of the chart has a lower resistive value; for very high frequencies, the center of the chart moves toward 50 ohms. At both extremes the voltage standing-wave ratio tends to become lower (of the order of 10). These conclusions are tentatively drawn from limited probe measurements made on some power outlets and loads. Systematic statistical work may yield more exact boundaries; but for practical purposes, one can assume that the whole Smith chart represents the interface impedance for the critical frequency range of 10 kHz to 10 MHz.

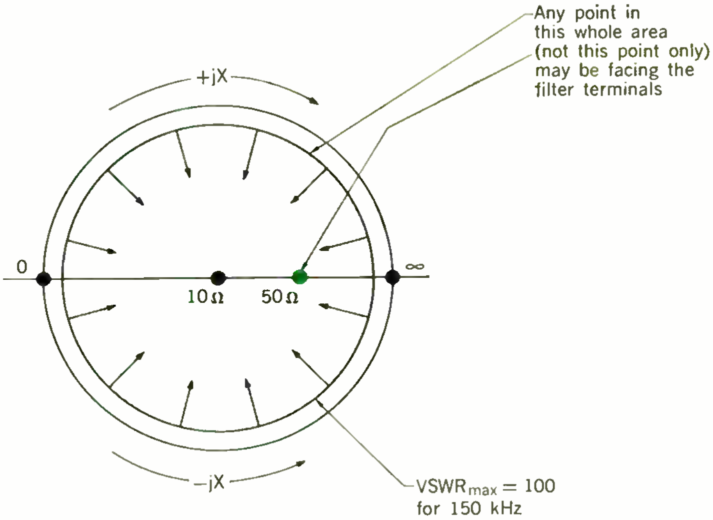
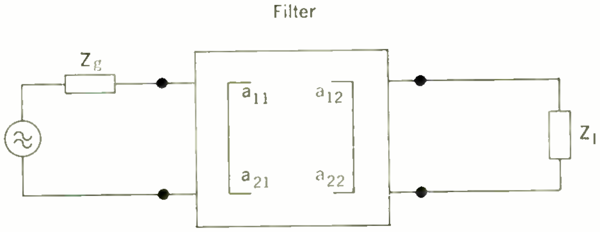


FIGURE 3. The impedance seen by a filter may be nearly anywhere on the Smith chart.

FIGURE 4. Filter represented by the A matrix.



Interfacial resonances. Power-line filters, then, are randomly and often severely mismatched. This may result in an improvement or in a degradation of predicted filter performance. We are interested here in worst-case performance.

For computer work the cascade (or A) matrix is most suitable for determining the envelope of minima of insertion loss. With the notations used in Fig. 4, the insertion loss (IL) can be derived as:

$$\text{IL (dB)} = 10 \log_{10} \left| \frac{a_{12} + Z_l a_{11} + Z_g a_{22} + Z_l Z_g a_{21}}{Z_l + Z_g} \right|^2 \quad (1)$$

Equation (1) can readily be minimized on the computer by assigning any reactive value to Z_g or Z_l, where Z_g is the generator impedance as transformed by the supply line to the filter input terminals and Z_l is the corresponding load impedance seen at the output terminals of the filter.

Before discussing the results of such truly worst-case calculations, we shall consider briefly another well-known representation of insertion loss,[3] an expression that is more transparent and permits us to gain an insight of what occurs when the filter fails to operate as a filter.

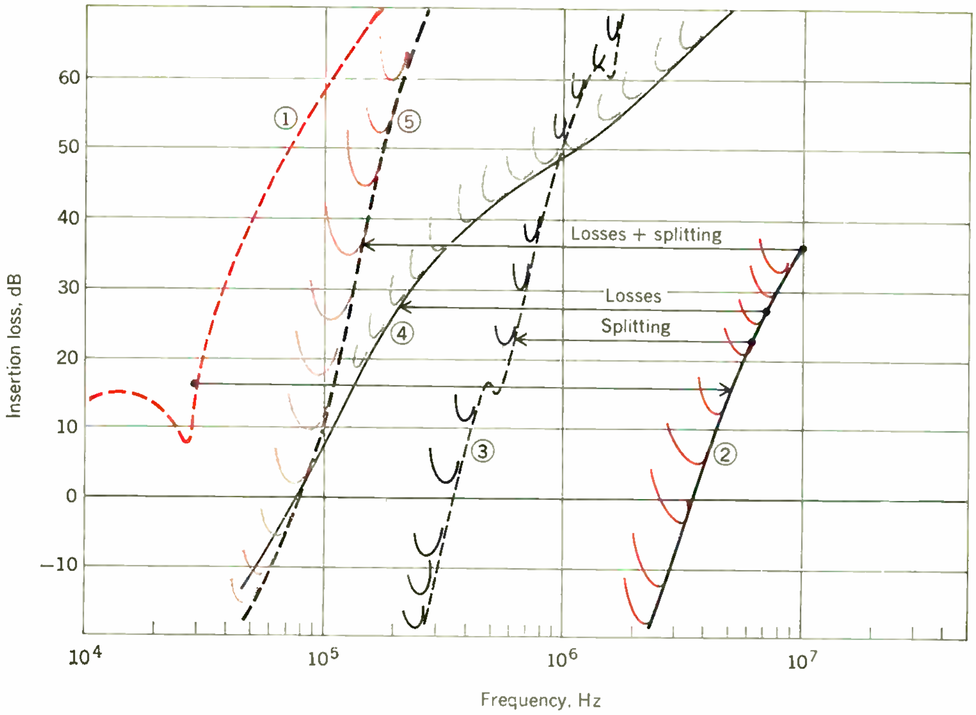


FIGURE 5. Comparison of the performance of filters in different configurations.

Consider the following equation:

$$\text{IL (dB)} = 20(\log_{10} F_2 + \log_{10} F_3 + \log_{10} F_4 - \log_{10} F_1 + 0.434a) \qquad (2)$$

where $$F_1 = \frac{Z_g + Z_l}{2(Z_g Z_l)^{0.5}}$$

The F_1 term is subtractive because generator–load impedance mismatch does not contribute to the filter effectiveness. The other terms include, besides Z_g and Z_l, the characteristic impedance Z_0 (which for simplicity is assumed to be symmetrical) and the attenuation, a, of the filter. Therefore,

$$F_2 = \frac{Z_g + Z_0}{2(Z_g Z_0)^{0.5}} \quad \text{and} \quad F_3 = \frac{Z_l + Z_0}{2(Z_l Z_0)^{0.5}}$$

Term F_4, expressing the interaction of primary and secondary mismatch, can be neglected for even a moderate attenuation a; hence, only F_2 and F_3 significantly affect the insertion loss of the mismatched filter. They can, in effect, compensate or even overcompensate the attenuation term in Eq. (2) for essentially lossless filters. Because Z_0 in the stopband is reactive, Z_g or Z_l (or both) may resonate with Z_0. Such interfacial resonances can reduce F_2 or F_3 to values close to zero; that is, their logarithms assume high negative values. This produces insertion gain instead of insertion loss, which one expects from an impedance-matched filter. Insertion gain can exist at several frequencies for a filter in a particular configuration because Z_g and Z_l form unpredictable spirals on the Smith chart.

In order to avoid detrimental interfacial resonances in the stopband of the filter, one can select a filter configuration such that L equals L and C equals C at the interfaces, thus shifting the interfacial resonances into the passband. Yet, even if the user of the filter takes the trouble to measure the generator and load impedances and selects a filter so that adjoining impedances are of equal sign, the impedances can change, as discussed in the context of Fig. 2. To ensure unconditional functioning for passive EMI filters, a more circumspect approach to filtering in highly mismatched circuits is necessary.

Resonances are suppressed by introducing heavy, frequency-selective losses, so that losses do not occur at power frequencies. These losses may be introduced into the generator, the load, the power line, or the filter. Disregarding losses in the generator and load (though not denying their possibility), we can choose between introducing heavy losses in the line or in the filter. We first consider lossy filters.

Conventional vs. lossy low-pass filters. Because low-pass filters consist of series L's and shunt C's, losses may be introduced either into the L or C branch. Inductive losses are caused by dispersion in the magnetic material or by wire losses, particularly by an enhanced skin effect. Unfortunately, both (series) loss mechanisms do not provide the desired low Q ($\leqslant 1$) for frequencies below 5 MHz. Magnetic dispersion is shifted to higher frequencies by the inherently unavoidable current bias of the power-line filter. The enhanced skin effect (to be discussed later in connection with line dampening) results in low Q for straight wires; when the wire is formed into a coil, however, L increases but not R, making $Q \gg 1$. Consequently, only C remains for providing a $Q \leqslant 1$. Without sacrificing breakdown strength, this is possible by making the capacitor an R-C line with $\omega_0 RC = 1$, where ω_0 is its cutoff frequency. The lowest value of ω_0 is chosen for which high attenuation is stipulated; simultaneously, the steepest possible ascent of attenuation is permitted (because $Q \gg 1$ for $\omega < \omega_0$). A multisection filter is required for steep cutoff of the whole filter.

The beneficial effect of properly introduced losses and of subdividing the filter is illustrated by the results of Fig. 5.[4] The insertion loss values were calculated from

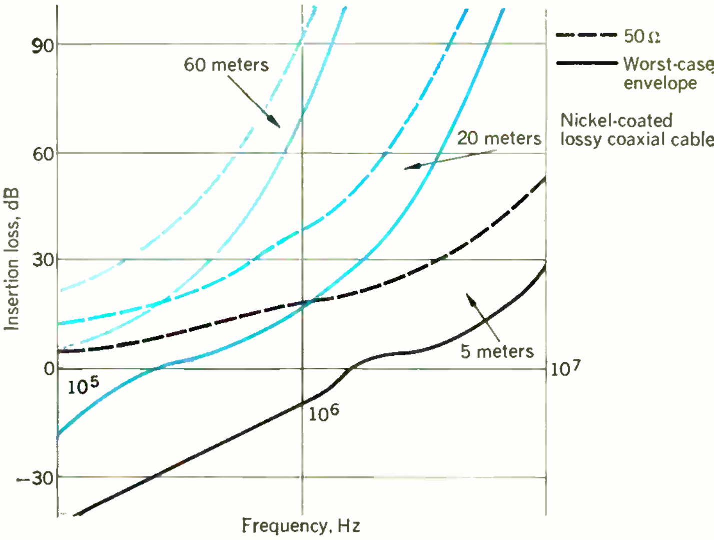
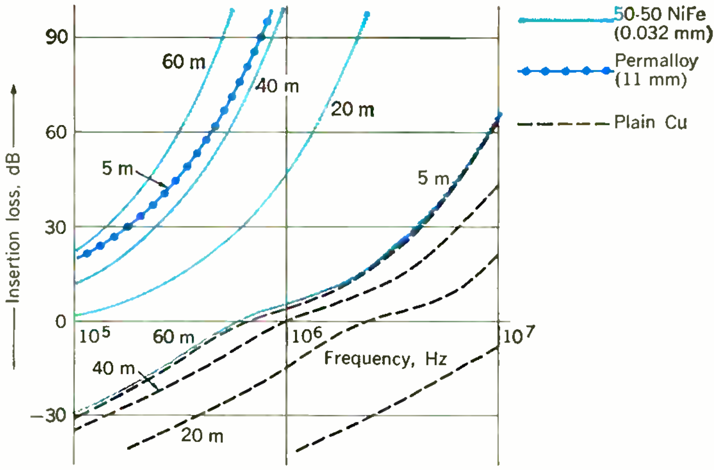


FIGURE 6. Worst-case performance of Ni-coated lossy lines

FIGURE 7. Worst-case envelopes for various coaxial lines.



(1); all five curves are for filters having a total of 40 μH and 4 μF each. Curves 1, 2, and 4 refer to simple pi networks and curves 3 and 5 to ladder networks having three equal L's of 13.3 μH each and four equal C's of 1 μF each. Curves 4 and 5 include capacitive losses with $Q = 1$ for $f > 150$ kHz. Curve 1 is a single, definite curve; all others are envelopes of worst-case conditions.

Curves 1 and 2 refer to a lossless pi network. Curve 1 is determined according to MIL-ST-220A in an illusory 50-ohm system; curve 2 is an envelope of worst-case insertion losses, the worst case being defined by the existence of interfacial resonances. Figure 5 shows that a pi network may meet the stipulations of 50 dB at 150 kHz if measured according to MIL standards, but in actual use it may have a cutoff frequency 100 times higher.

Juxtaposing curve 2 with curves 3, 4, and 5 shows clearly that a combination of splitting L and C and introducing capacitive losses results in an envelope (curve 5) that is not far apart from curve 1, for which the insertion loss is measured for $Z_l = Z_g = 50$ ohms. Curve 4 indicates that lossy capacitors in the pi filter results in a pronounced improvement.

Filters have been built for 240 volts ac, 10 amperes and 100 volts dc, 10 amperes. They have a guaranteed minimum insertion loss of 45 dB at 150 kHz and 60 dB for frequencies greater than 300 kHz. Because they employ ceramic capacitors in close packaging, both filters are smaller than conventional types.

Lossy lines vs. lossy filters. Because the power-supply line has to be laid, it is an obvious idea to try to distribute a lossy filter along the line, as numerous patents issued for this testify. The practicability and economy, however, of replacing lossy filters by lossy lines is highly questionable. It seems, with presently available materials, lossy lines are useful only if either very long lines (of the order of hundreds of meters)[5] or very high frequencies (about 10 MHz)[6] are involved.

It is not so simple, as enthusiastic proponents claim, that one buys fixed frequency attenuation by the yard. Rather, the shorter the lossy cable, the higher the frequency at which it is effective. Furthermore, one can make lossy cables very effective for short lengths, but the coating on the copper wire becomes prohibitively thick, costly, and unbendable.

The results for a lossy coaxial cable, calculated from Eq. (1), are illustrated in Figs. 6 and 7. In all cases, the cable is composed of a copper wire 2 mm in diameter coated with resistive magnetic material to enhance skin effect. Over this is a plastic dielectric having a thickness of 0.012 mm and an ϵ of 3; surrounding the dielectric is the outer conductor.

For 5-, 20-, and 60-meter lengths, the upper curves of Fig. 6 are based on a 50-ohm system; the lower curves show the envelope of worst-case mismatch. Nickel (μ = 200, g = 10^7 mho/m, skin depth = 0.1 mm at 150 kHz) contributed approximately 0.5 ohm/m at 150 kHz and the dielectric contributed 13 000 pF/m to the equivalent R-C network. Replacing the nickel by 50-50 NiFe (μ = 2000, g = 5 × 10^6 mho/m, skin depth = 0.035 mm) results, with 5 ohms/m at 150 kHz, in the worst-case limits shown in Fig. 7. These curves also indicate what would happen if the magnetic layer were omitted.

The results obtained with the simple and flexible lossy lines just discussed are rather unsatisfactory; for short lengths the cable is effective only at higher frequencies, where an inexpensive ceramic filter will do equally well. For long cable lengths the filtering is effective; however, a simple lossy lumped filter would be much cheaper. Drastically reducing the conductivity of the magnetic material yields excellent worst-case performance of a 5-meter-long cable for which 2-81 Permalloy coating replaces the nickel (Fig. 7). Unfortunately, the magnetic layer now has to be over 1 cm thick; this is expensive and completely unworkable.

With the materials presently available, the practicability of lossy cables is doubtful; nevertheless, for reasons that will become clear shortly, we shall point out how to eliminate the biasing effect of power current. Without special precautions, the power-line current reduces μ, and hence the loss resistance, to completely useless values. Figure 8 depicts several practical structures for introducing air gaps in the magnetic material and maintaining the available μ for the frequencies to be attenuated. (Here, we go beyond Clark,[5] who applies this method in principle to simple, narrow, longitudinal slots.) Two effects are exploited for this purpose:

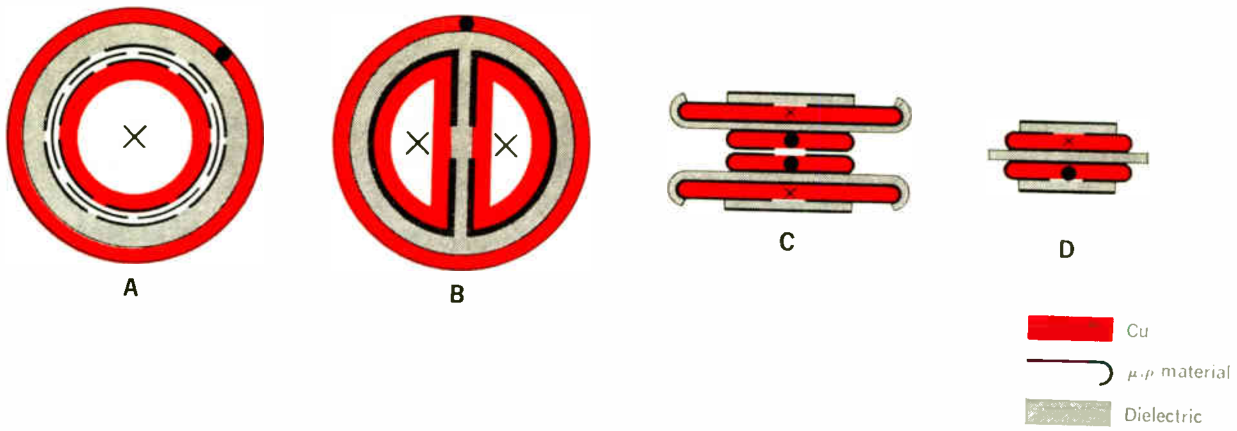


FIGURE 8. Lossy TEM transmission lines utilizing enhanced skin effect. (A and B coaxial, and C and D strip-line versions.)

1. The proximity effect, which does not permit RF to penetrate into narrow gaps.

2. The effect of positioning gaps where the RF field is zero or minimal.

For example, the inner double conductor of Fig. 8(B) taken by itself—that is, with the dielectric and outer conductor removed—can be used as lossy wiring with simple lumped capacitors. Copper wire of 2-mm diameter, provided with a very thin layer of 50-50 NiFe, exhibits 5 ohms per meter at 150 kHz and is nearly independent of current; such a wire of 1-mm diameter has a resistance of 10 ohms per meter at 150 kHz. Lossy wires of this kind reduce Q in equipment wiring and eliminate poor filtering with simple capacitor filters or lossless filters. The authors envision in a few years the widespread application of such lossy wire in equipment wiring for the essential purpose of preventing ringing in the passband, where lossy filters provide no damping.

Testing and certification

Owing to the disastrous results of unsuppressed conducted interference, it is essential that suppression devices be highly reliable. To this end, one is required to perform meaningful tests on the filter and guarantee a minimum insertion loss under all interface conditions. The Armed Forces—but unfortunately not industry—are formulating new, more significant test specifications to supersede MIL-ST-220A and other misleading specifications that pretend favorable matching conditions seldom found in practice. As far as the authors can determine, these contemplated specifications are steps in the right direction but fail to go far enough in that they use no tuning devices and do not provide high-Q circuits.

Figure 9 illustrates how these crucial conditions can be incorporated in a setup that allows testing with applied ac or dc power. The two 10-μF feedthrough capacitors shunt out accidental generator and load impedances for $f > 150$ kHz, thus permitting interfacial tuning by variable inductors of good Q. For filters having inductive interface impedances, a capacitor has to be placed in series with L_1 or L_2. Of particular importance are the current probes feeding the RF energy in and out of the system. Originally, off-the-shelf current probes were inserted that had a 7:1 turns ratio and were fed by 50-ohm generators and receivers. A value of $50/49 \approx 1$ was transformed into each side of the test circuit, thereby creating low Q's that do not constitute worst-case conditions. New current probes with a 30:1 turns ratio resulted in a much more pronounced failing of conventional filters under test (the current probes no longer dampened the interfacial resonances). With lossy filters now becoming available, they should be tested for truly (not just nominally) worst-case conditions.

Interfacial resonances can be suppressed only when losses are located in such a way that the Q is low at the interface. To prove this statement, we compared an imperfectly lossy filter (a pi network having a lossy inductor) and an experimental truly lossy compatible filter (having lossy capacitors) in the test fixture previously described. The results of these measurements are presented in Fig. 10; the difference in worst-case operation of the two filters is quite striking.

As long as lossy wires are not commercially available and applied in equipment wiring, the use of lossy filters is the only guaranteed solution for unconditional filtering. The simple test of Fig. 9 has to be applied only to one sample of a filter of a certain rating as a qualification test; filters of the same type will operate in all situations.

Alternate test methods that fit specific lossless filter configurations into a given interfacial environment must be carried out for each and every line to be filtered. Filter effectiveness can never be guaranteed because the interfacial environment is seldom invariant. New test methods of filter fitting developed to make lossless filters seemingly acceptable are time-consuming, do not permit the specification of filters in advance, and, worst of all, cannot guarantee unconditional effectiveness.

Ceramic filters and capacitors

The use of very high dielectric constants of certain ceramics, with or without the incorporation of ferrites, for very small filters is well established.[7] The basic structures are shown schematically in Fig. 11(A); their corresponding performance in a 50-ohm system is shown as the right-hand curves of Fig. 11(B).

Previously, ferrite and dielectric tubes were formed independently and mounted together; such an assembly constituted a fragile structure. New ceramic techniques,

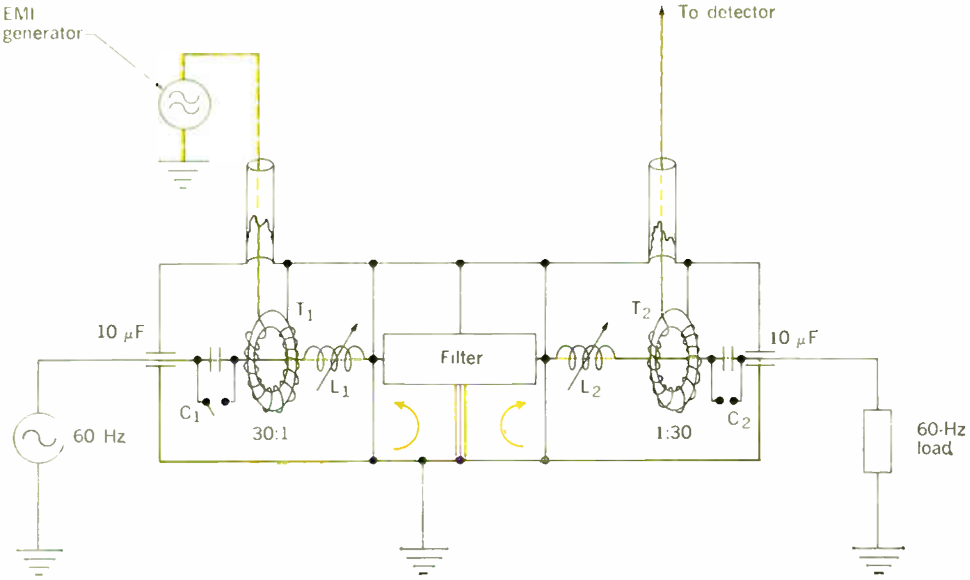
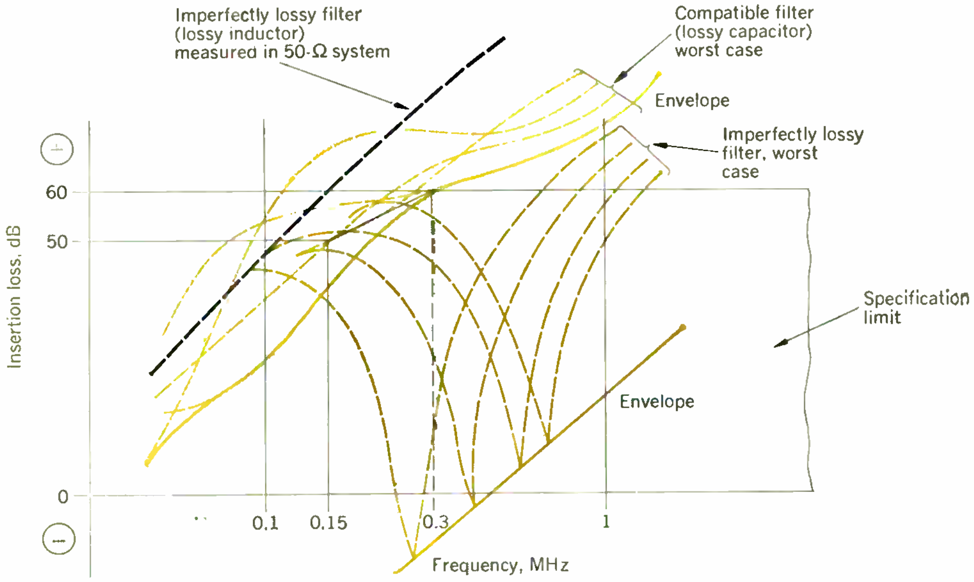


FIGURE 9. A truly worst-case test circuit for EMI filters.

FIGURE 10. Comparison of a truly lossy compatible filter with an imperfectly lossy filter.



however, permit the firing of syncrete for extremely rugged structures. Related techniques allow also for integral units of tubular multilayer capacitors of great density. For example, capacitors of 1 μF suitable for multipin connector installations are feasible; they are rated at 35 volts from −55°C to +125°C. Such a capacitor has an outer diameter of 2.5 mm, inner diameter of 1.2 mm, and length of 1.2 mm. The ruggedness achieved provides new possibilities for very reliable and small-size filtered multipin connectors and filter blocks. These filter capacitors can also be made with $Q < 1$ in the stopband.

The breakdown voltage of miniaturized ceramic filter capacitors is 10 times rated voltage; their precorona voltage is 3 times rated voltage. Such high reliability assurance is an indispensable requirement for filtered multipin connectors that are employed in large quantities for interconnecting subsystems.

Active filters

For low frequencies, EMI filters made of passive elements are prohibitively bulky and heavy because the line current flows through the filter. Active filters, using transistors, can provide large apparent values of L and C, resulting in a reduction of size and weight. Moreover, the low impedance levels existing at low frequencies in power lines can be more easily accommodated. Two major criteria exist in analyzing the filtering conditions:

1. Power flow (dc, 60 Hz, 400 Hz) with respect to the interference flow.
2. Either voltage or current attenuation as the dominant and desired feature.

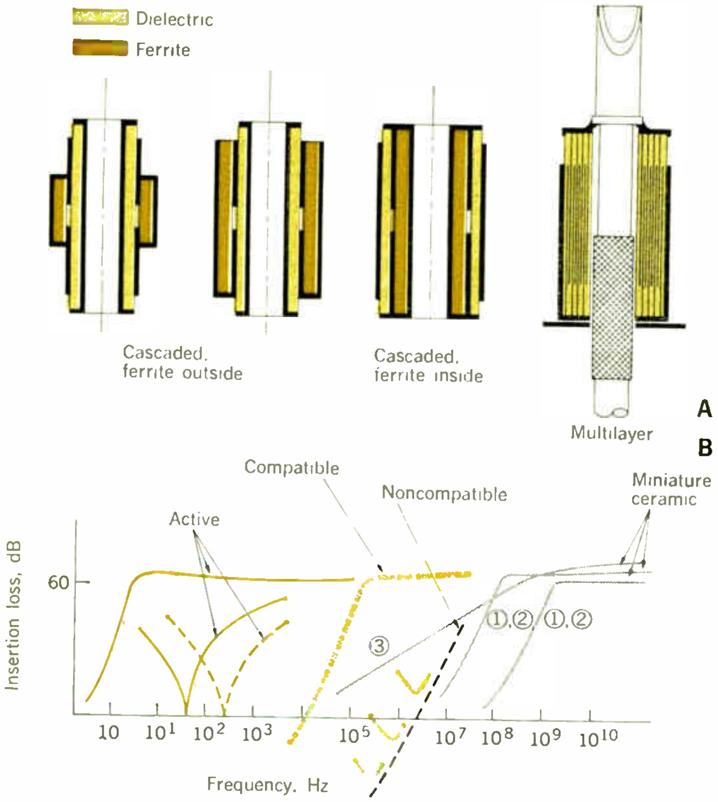


FIGURE 11. Feedthrough filters and capacitors suitable for connectors. A—Individual structures. B—Performance in a 50-ohm system.

For the power-flow criterion, three basic situations must be distinguished for ac and dc lines. The first is the interference originating in the power supply and flowing toward the load; for example, the filtering of a power line supplying a shielded room. The second situation involves interference originating in the load and flowing toward the power supply. An example is the prevention of the spreading of interference arising from SCR switching, computers, and other digital equipment or solid-state choppers, through the power-distribution system. As another example, current pulses driving a teletypewriter and carrying confidential information must be confined within a controllable area. The third situation includes cases in which interference is generated both at the power source and load, or uncertainty exists about its origin; here, bidirectional filters are required.

For the attenuation criterion, it is important to know, at the very beginning, if the filter's prime task is to provide a minimum voltage or current attenuation, independent of the terminating impedance (which may vary considerably in a power line).

Underlying principles of operation. There are four well-known means, which may be employed alone or may be judiciously combined, to prevent interference from passing through a network. These are

1. Storing and averaging out the interference (such as through the use of a capacitor in a dc line).
2. Using a large series impedance (series regulator) to obstruct the flow of interference.
3. Using a low shunt impedance (shunt regulator) to bypass interference to ground.
4. Using an equal but opposing signal to cancel the original interference signal.

In contrast to electromechanical devices or electrochemical storage devices, mere electrical means (such as capacitors) do not store the sizable amounts of energy required for high-power, low-frequency filters. There is also the difficulty of storing energy in the form of alternating current—for example, in resonant circuits. Whatever portion of the interference that cannot be stored and averaged out in a passive element must be prevented from passing through the active section of the filter. This is done by providing either a large impedance path to the interference signal (series regulator), shunting it to ground (shunt regulator), or canceling it by an opposing signal of the same magnitude. These active filters may therefore contain, for a dc line, capacitors as storage elements and modified series regulators to create a high-impedance path, and modified shunt regulators in combination with high-gain feedback systems for cancellation. In the case of ac line filters, cancellation is a most effective method for minimizing interference. In contrast to conventional regulators used in regulated power supplies, these regulators must not regulate the amplitude of the power to be passed.

Although averaging filters using only the storage principle (passive filters) have efficiencies approaching 100 percent, power-line filters incorporating active elements have a maximum efficiency of 90 to 95 percent. The active elements contained in the regulator and feedback loops cannot store energy and must therefore be biased at an operating point that will permit accommodation of the interference amplitude and the power-line current that passes through the loops. Since the power dissipated is drawn from a dc supply or taken from the power line directly, the overall efficiency of the active filter is limited. The actual efficiency depends on the design of the filter and the ratio of the interference power to the power in the passband.

Active filters must be protected against current, voltage, and power overloads in steady state, and against transients resulting from turn-on and turn-off periods. Direct solid-state protection will disconnect the filters; indirect protection causes the excessive power to bypass the filter.

Certain constraints have to be observed when an active filter is fitted into a given system. Gain–bandwidth products of available power transistors limit the obtainable attenuation; dc power supply fluctuations, power-frequency sidebands, and the maximum expected interference amplitude limit efficiency. Nevertheless, the application of active elements yields, besides size reduction, very desirable features; these include having the voltage and current attenuation independent of the externally connected impedance.

Examples

From the foregoing, it can be seen that active filters must be tailor-made in order to utilize them to their fullest advantage. The following simplified examples typify the three most often encountered filter requirements, point out the necessity for a careful problem analysis of each situation, and describe the basic operation and achievable performance for each filter.

Inverter filter. In industrial and military applications, dc/ac, dc/dc, and ac/ac inverters ranging from watts to several hundred kilowatts are rapidly gaining widespread

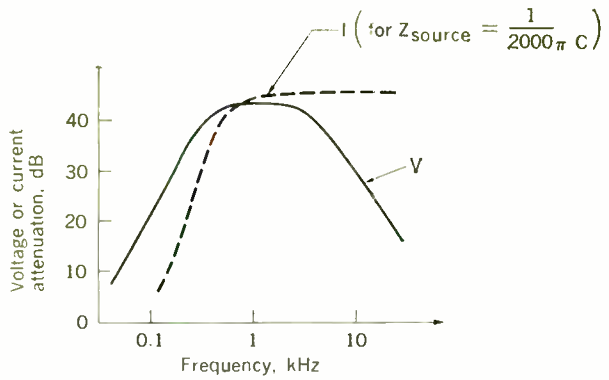


FIGURE 12. Voltage and current attenuation for 1-kHz inverter filter.

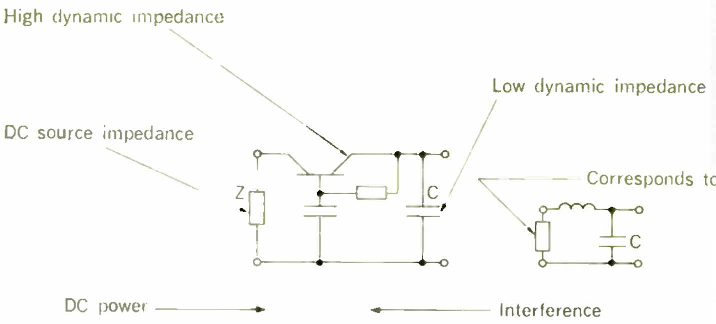


FIGURE 13. Substitution of an active circuit for an inductor in a passive filter.

use. Their operation produces repetitive current fluctuations containing a broad frequency range with the first harmonic usually between 50 hertz and several kilohertz. Because of a finite dc power source impedance, including the interconnections, corresponding voltage fluctuations are generated. These fluctuations in turn propagate as interference throughout the system and may result in the malfunctioning of associated sensitive electronic equipment. This is a case in which power and interference flow oppose each other and a given voltage attenuation is desired.

The voltage and current attenuations of a typical active filter for a dc/ac inverter are shown in Fig. 12; the 1-kHz chopping frequency is voltage-attenuated by 40 dB. The circuit uses a storage capacitor in combination with a large series impedance element. With the power source impedance, an equivalent passive pi filter is formed by substituting a common-base circuit for the inductor (see Fig. 13). When an ac voltage is applied between collector and ground (interference side), only a small fraction, $1/h_{rb}$, of this voltage appears across the external impedance connected between emitter and ground (filtered side). Capacitor C, presenting a shunt impedance, is required when only the interference source is a current rather than a voltage. From the h-parameter equivalent circuit, the following expressions can be derived for this type of filter:

$$\text{Voltage attenuation (dB)} = 20 \log_{10} \frac{1}{h_{rb}}$$

$$\text{Current attenuation (dB)} = 20 \log_{10} \frac{Z}{\frac{1}{2\pi f C} h_{rb}}$$

where C = capacity, Z = power-source impedance, f = frequency, and h_{rb} = small-signal reverse voltage transfer ratio.

Obtained efficiencies of greater than 90 percent approach closely those realized with passive filters. The voltage drop is less than 2 volts over a current range of ±20 percent, of the nominal value; the supply voltage may vary ±20 percent. The filter is self-contained, requiring no external supply.

Random-pulse filter. Quite often one is faced with the problem of suppressing the conducted propagation of pulses generated by randomly switching many loads. For rectified power, filtering on the dc side is most effective; as before, the interference flows in the opposite direction to the dc power. Here, current attenuation rather than voltage attenuation is generally required.

The particular filter employs the same underlying principle of the preceding example (storage capacitor and a large series impedance element). Because the attenuation requirement of this filter is higher and has to be maintained over a wide frequency range, typically 10 Hz to 100 kHz, a high-gain transconductance amplifier is added to the circuit as shown in Fig. 14(A). Whatever interference current is still present at the emitter side of the pass transistor is attenuated by the closed-loop circuit; the attenuation is identical with the open-loop gain of the amplifier. The filter yields high current attenuation values, independent of the dc source impedance [60 dB from 10 Hz to 100 kHz, as shown in Fig. 14(B)] in a unit with a dc rating of 36 volts and 5 amperes. The attenuation of load current pulses is illustrated in Fig. 14(C).

Limitations of pulse level, duration, repetition rate, etc., have to be considered and must be specified for a given filter. The filter does not require an external power supply. Efficiencies of 75 to 80 percent can be attained and dc line voltage drops of 3 to 6 volts exist, depending on the ripple in the power supply.

Interference filters for 50-, 60-, and 400-Hz power lines. Active ac power-line interference filters pass, with high efficiency, only a narrow band about the power frequency. One of the prime applications is in connection with shielded-room filtering (Fig. 15), where interference flowing in the same direction as the 60-Hz power has to be attenuated. Without sophisticated modifications, moderate voltage-attenuation values, approximately 30 dB, are obtained even at very low load and source-impedance levels; two filters may be cascaded for higher attenuation values. In modified switching-mode operation, a 60-dB insertion loss should be realized in one stage.

The filter, shown in Fig. 16(A), is based on the principle of cancellation and, in very simple terms, operates as follows: The signal from the input appearing at A is fed into an ac-coupled amplifier through an adaptive notch filter, which is tuned to the fundamental of the power-line frequency. The amplified interference signal, with opposite polarity, is returned in series to the source through transformer T. All signals at A except the fundamental are therefore attenuated by the gain of the amplifier. Within a limited range, a separate digital control circuit

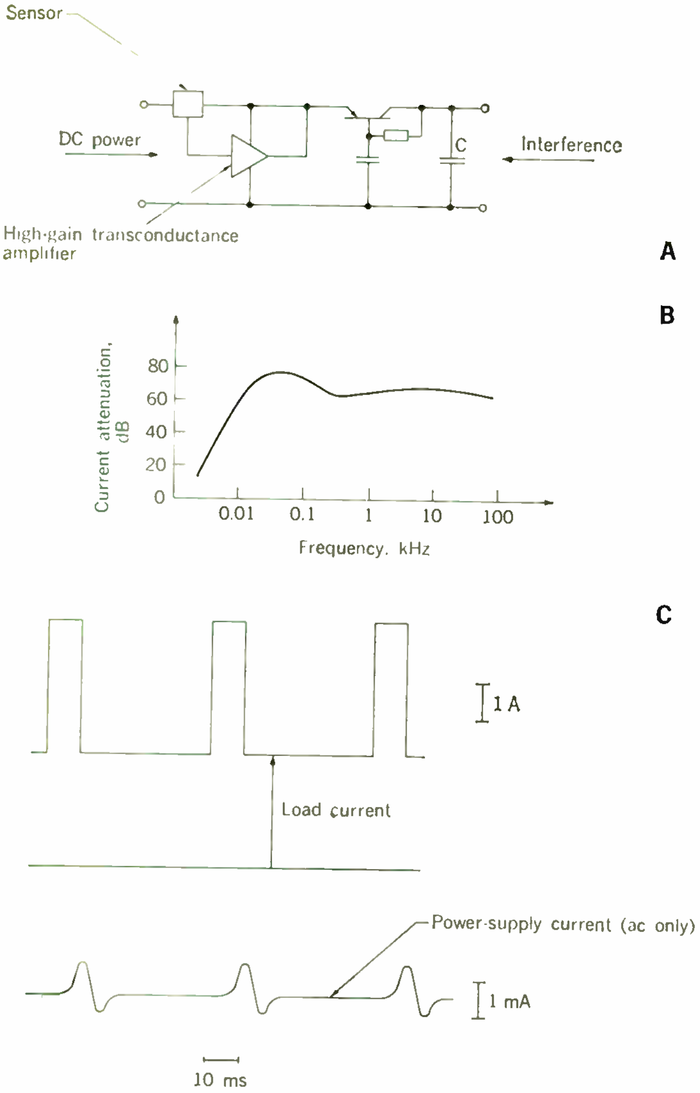


FIGURE 14. Random-pulse filter. A—Basic circuit for wide-band attenuation. B—Current-attenuation curve. C—Attenuation of 20-Hz, 2.5-ampere current pulses.

provides automatic tuning of the notch filter to the power-line frequency and corrects any changes in the filter itself. For a simple 220-volt, 20-ampere unit, the voltage-attenuation curve is given by Fig. 16(B).

When inserted in the line, the filter introduces the equivalent of a small inductance of 700 μH for the pass frequency. The change in output voltage at full current rating caused by this inductance is negligible: 1 volt for a load power factor approaching unity and up to 6 volts in the worst case for zero power factor. The filter can handle an interference voltage of 80 volts, peak to peak; this interference amplitude could be increased, but at the expense of efficiency. Efficiencies of 90 percent have been realized for 220-volt, 20-ampere filters.

Conclusions

The frustrations so often encountered in EMI filtering, namely, uncertain effectiveness and inconveniently large size, can be overcome by discarding conventional thinking in the design of such filters. The solution lies in the use of ceramic, lossy, or active filters.

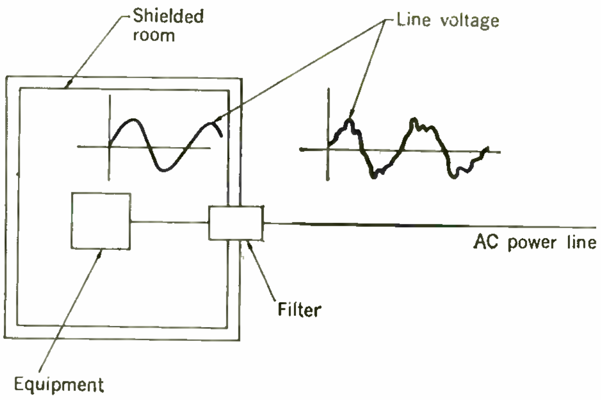
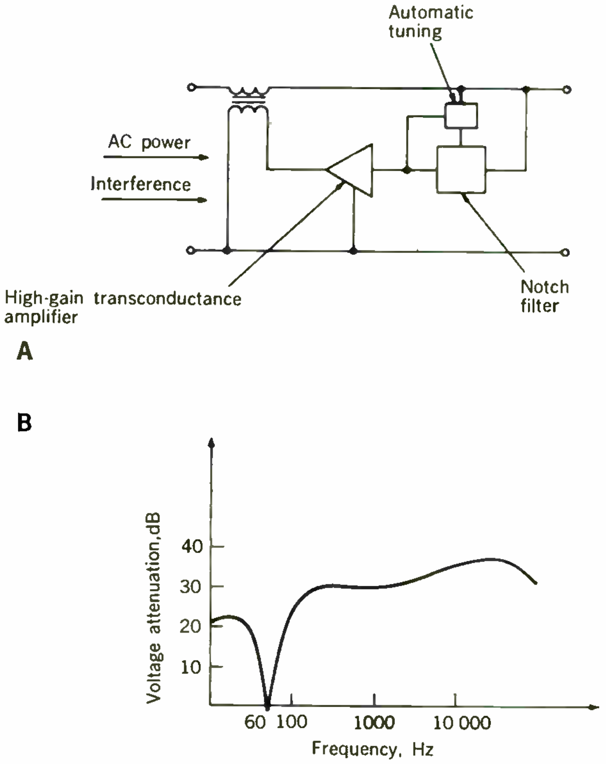


FIGURE 15. Application of a 60-Hz power-line filter for shielded rooms.

FIGURE 16. Power-line filter. A—Basic circuit. B—Voltage-attenuation curve.



REFERENCES

1. Zverev, A., "The golden anniversary of electric wave filters," *IEEE Spectrum*, vol. 3, pp. 129–131, Mar. 1966.

2. AFSC Manual No. 80-9, Transmittal No. 7, HIASD, vol. IV, p. F5-3.

3. Koehler, G., *Circuits and Networks*. New York: Macmillan, 1955, p. 119.

4. Excerpted from H. M. Schlicke's chapter on filtering in the forthcoming *Handbook on RFI*, with the permission of the editor-in-chief, R. Ficchi.

5. Clark, D. B., and Hitchcock, R. D., "Contiguous wrapping of transmission line conductors with high μ tape for large radio interference attenuation," *Proc. 5th Conf. on RFI and EMC*, Oct. 1959, p. 560. Also, U.S. Patent No. 3 219 951.

6. Mayer, F., "EMC—anti-interference wires, cables, and filters," *IEEE Trans. Electromagnetic Compatibility*, Vol. EMC-8, pp. 153–160, Sept. 1966.

7. Schlicke, H. M., "Cascaded feedthrough capacitors," *Proc. IRE*, vol. 44, pp. 686–691, May 1956. Also, U.S. Patent No. 3 035 237.

Nuclear power—the next decade of development

If the past few years are any indication, the U.S. nuclear power industry will make tremendous strides in the coming decade. The Atomic Energy Commission is encouraging close cooperation between the government, the utilities, and the reactor plant manufacturers

Milton Shaw *U.S. Atomic Energy Commission*

The Atomic Energy Commission is playing an active advisory role in the design, development, and operation of civilian nuclear power plants in the United States. The trend toward larger plant capacities is proceeding at a phenomenal rate and an increasingly significant portion of the new electric generating capacity in the United States is being committed to nuclear power. The utilities' acceptance of light-water reactors has served to speed the progress of advanced reactor projects. At present, the sodium-cooled liquid-metal fast breeder reactor (LMFBR) represents the AEC's highest-priority program.

The rapid rate of growth starting in 1965, as evidenced by the number of orders for commercial generating plants, indicates that nuclear power is becoming increasingly attractive to the utility industry. Thus, in looking ahead to the next decade of development in the field of civilian nuclear power, it appears appropriate to examine briefly the trends of this growth pattern, the assessments of the reactor concepts now in progress and related plans for the future, and some of the significant factors related to these trends.

From Fig. 1 and Table I, it is apparent that nuclear power is finding its place in the U.S. power market following a development path concentrated in regions of heavy electric load and rapid growth. It can be seen from Figs. 2, 3, and 4 that the trend toward larger capacities is proceeding at an unprecedented rate and from Fig. 5 that an increasingly significant portion of the new electric generating capacity in the United States is being committed to nuclear power. Starting in the early 1970s it now appears that about 50 percent of all new plant capacity scheduled to come on the line will be nuclear, as shown in Table II.

These figures and tables illustrate the substantial advances that have taken place in the growth of nuclear power and in the acceptance of the light-water reactors by the U.S. utility industry as a new source of electric energy. To some extent, the utility acceptance of the light-water reactors has provided a new context for the advanced reactor projects. One of the promising features

> It was exactly ten years ago, at Shippingport, Pa., when we saw nuclear power first used in this country to supply electricity to the public on a regular basis. . . . This pioneering decade that began with Shippingport has seen the hardships and disappointments, and the achievements and final triumph associated with most pioneering efforts. Success has been achieved primarily as a result of the fine cooperation between government and industry in developing and putting into operation various reactor prototypes.
>
> What has been basically significant about the era was that in it we have established the safety and reliability of the light-water reactors and come to the point where nuclear plants employing reactors of these types are now being constructed on the basis of their economics.
>
> *Glenn T. Seaborg*
> *U.S. Atomic Energy Commission*

Revised text of a paper presented at a session on nuclear power at the American Power Conference, Chicago, Ill., April 25–27, 1967. The statements appearing in boxes were extracted from the four other papers presented at this session. The material presented in this article has been updated to mid-July 1967.

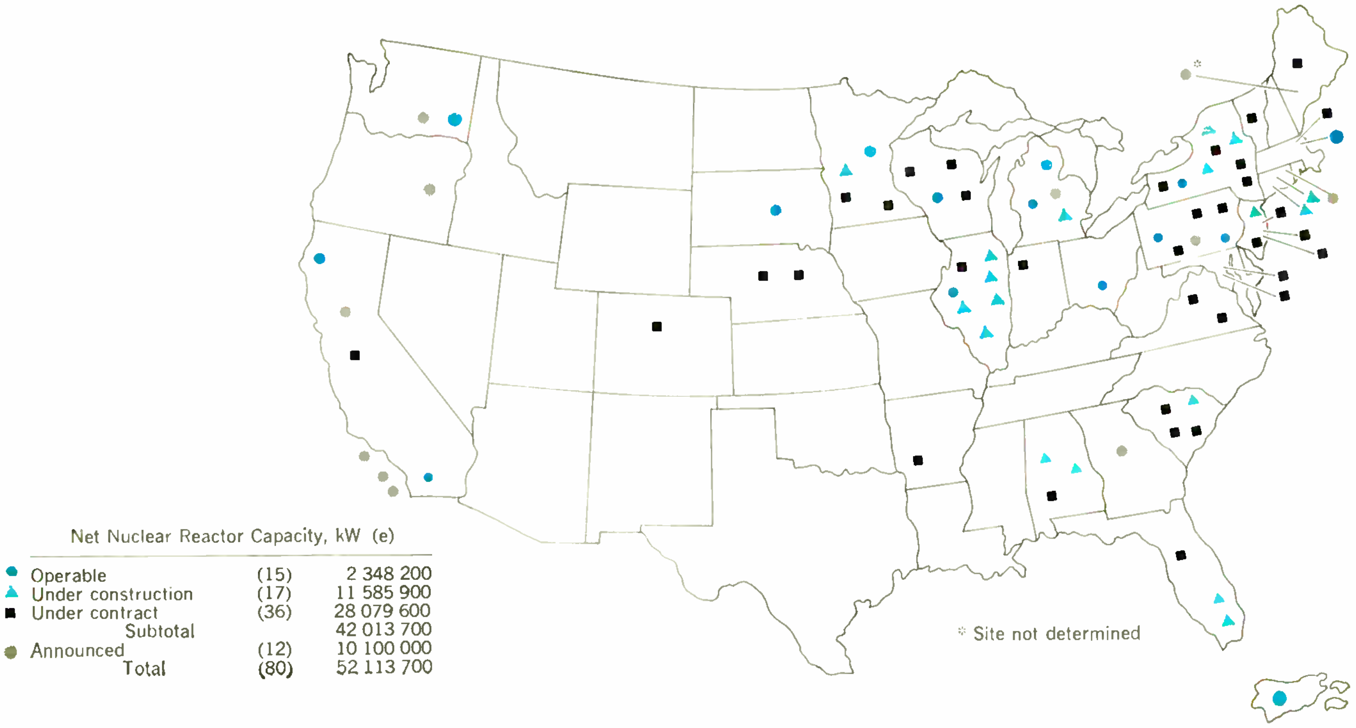


FIGURE 1. Central station power in the United States and Puerto Rico.

I. Central station nuclear power in the United States and Puerto Rico

Operable

California
Humboldt Bay
San Onofre

Illinois
Dresden No. 1

Massachusetts
Yankee

Michigan
Fermi
Big Rock Pt.

Minnesota
Elk River

New York
Indian Pt. No. 1

Ohio
Piqua

Pennsylvania
Shippingport
Peach Bottom No. 1

Puerto Rico
Bonus

South Dakota
Pathfinder

Washington
"N" Reactor

Wisconsin
LACBWR

Under Construction

Alabama
Browns Ferry No. 1
Browns Ferry No. 2

Connecticut
Conn. Yankee
Millstone No. 1

Florida
Turkey Pt. No. 3
Turkey Pt. No. 4

Illinois
Dresden No. 2
Dresden No. 3
Quad Cities No. 1
Quad Cities No. 2

Michigan
Palisades

Minnesota
Monticello

New Jersey
Oyster Creek

New York
Nine Mile Pt.
Ginna
Indian Pt. No. 2

South Carolina
H. B. Robinson

Under Contract

Alabama
Browns Ferry No. 3

Arkansas
Middle South Util.

California
Diablo

Colorado
Ft. St. Vrain

Florida
Crystal River

Illinois
Zion PW-1
Zion PW-2

Indiana
Bailley Sta.

Maine
Maine Yankee

Maryland
Calvert Cliffs No. 1
Calvert Cliffs No. 2

Massachusetts
Pilgrim

Minnesota
Prairie Island No. 1
Prairie Island No. 2

Nebraska
Ft. Calhoun
Cooper

New Jersey
Burlington No. 1
Burlington No. 2
Jersey Central No. 2
Bayside

New York
Shoreham
Indian Pt. No. 3
Easton
Milliken Sta.

Pennsylvania
Peach Bottom No. 2
Peach Bottom No. 3
3 Mile Island

South Carolina
Oconee No. 1
Oconee No. 2
Oconee No. 3

Vermont
Vermont Yankee

Virginia
Surry No. 1
Surry No. 2

Wisconsin
Point Beach No. 1
Kewannee
Point Beach No. 2

Announced

California
Bolsa Island No. 1
Bolsa Island No. 2
Rancho Seco
So. Calif. Edison

Connecticut
Millstone No. 2

Georgia
Georgia Power Co.

Michigan
Indiana–Michigan

New Hampshire
Pub. Serv. of N.H.

Pennsylvania
Susquehanna

Oregon
Trojan

Washington
WPPS

Site Not Determined
New England Elec.

of the more advanced concepts has been the prospect that they can compete economically with the light-water reactors; but in present circumstances they will face a more severe test in this regard than previously anticipated. On the other hand, the need for improved raw material utilization in order to forestall increases in cost for uranium has become more important because of the success of the light-water reactors and uncertainties in resource availability over the next few decades arising from lack of intensive uranium prospecting in the United States in recent years. Consequently, because of the many recent developments and changes, the Atomic Energy Commission is undertaking an extensive analytical effort to examine their impact on the future of the nuclear power industry and to determine what steps must be taken to insure that the competitive and self-sustaining nuclear power industry will make maximum use of our national resources of nuclear fuel.

These technical assessments and specific studies under way are intended to provide a new, thorough analysis and assessment of civilian power program prospects and of the optimum course of action for the future. The studies are being set up through the establishment of task forces to assess each of the specific reactor concepts; to synthesize the data from each concept and from a fuel recycle task force into an overall systems analysis; and to determine, with participation of utility members, the utility outlook on nuclear power and how best to factor the utilities into the national effort. The studies will consider in depth the technical and economic prospects for each of the reactor concepts under consideration, will relate these to resource projections and to overall economic considerations, and will identify the decisions that must be made in the light of changing circumstances. It is expected that the initial phase of the individual studies will be completed at various times during 1967. Some of these results may lead to changes in emphasis or direction of the program as a whole.

Changes since 1962

In response to a request from the Joint Committee on Atomic Energy, the AEC recently published a status re-

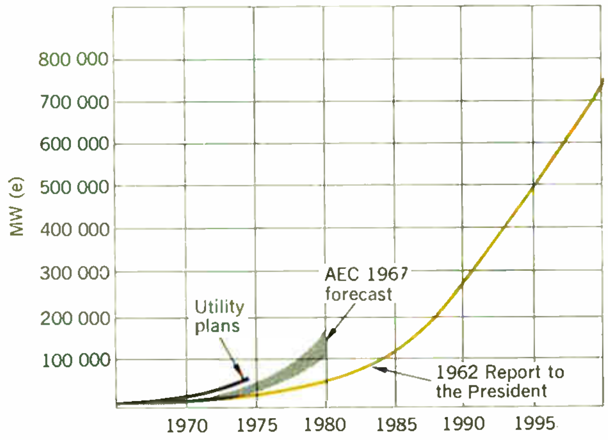


FIGURE 2. Estimated nuclear generating capacity in the United States through the year 2000.

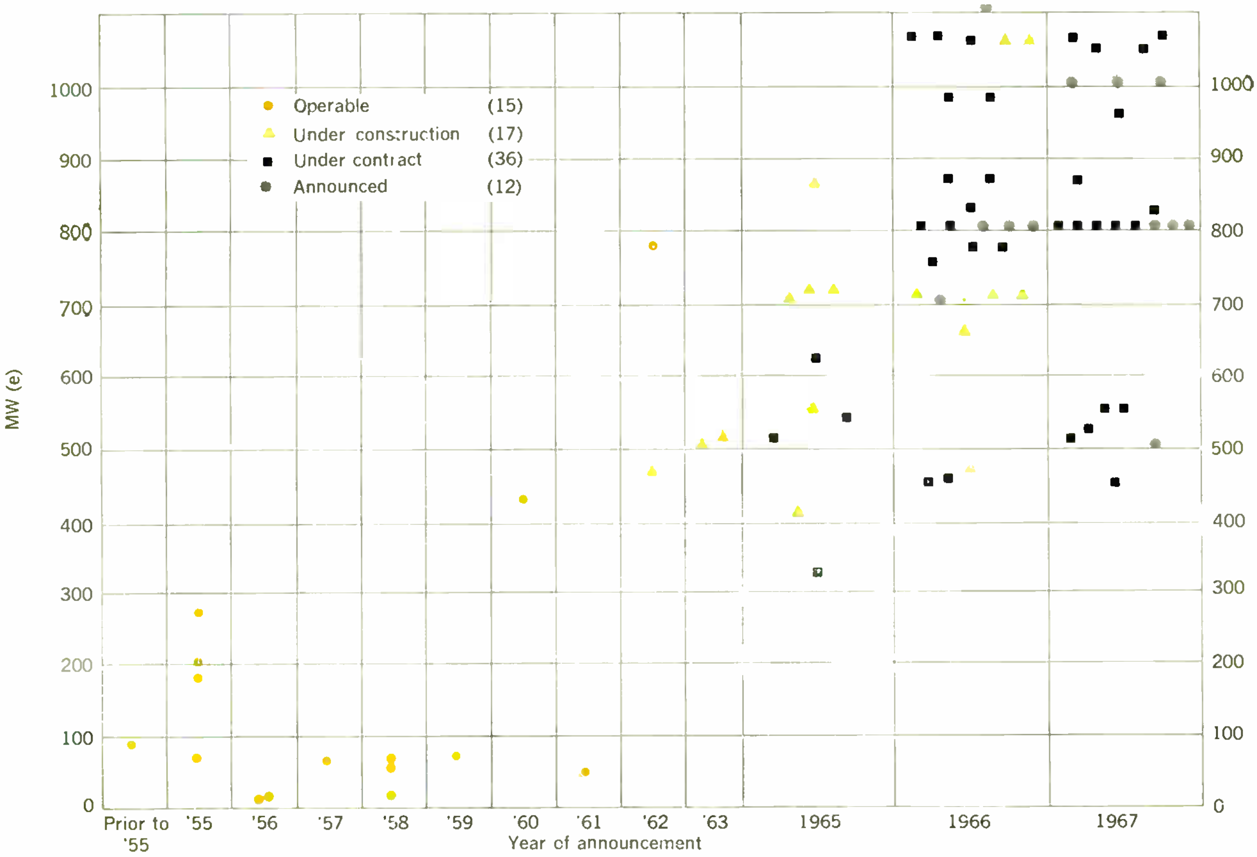
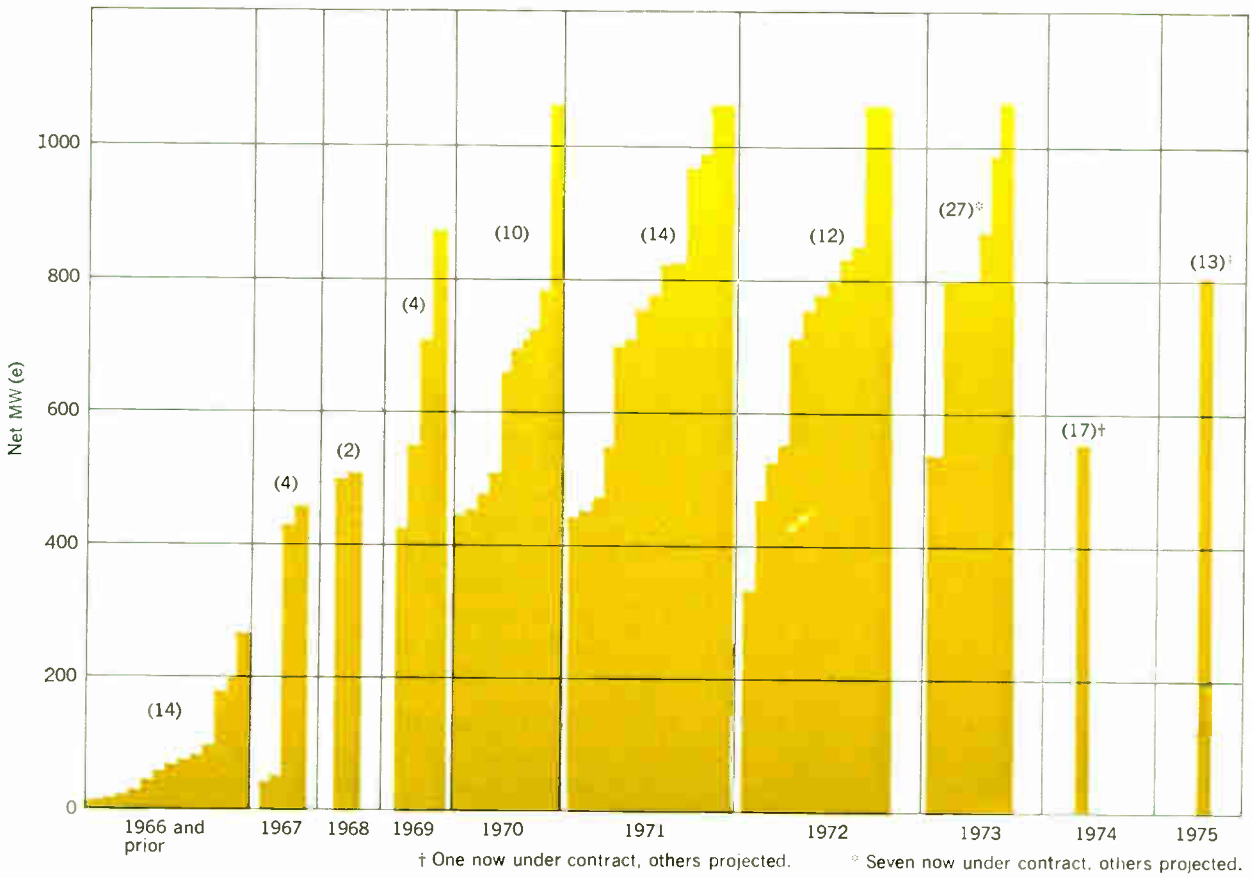


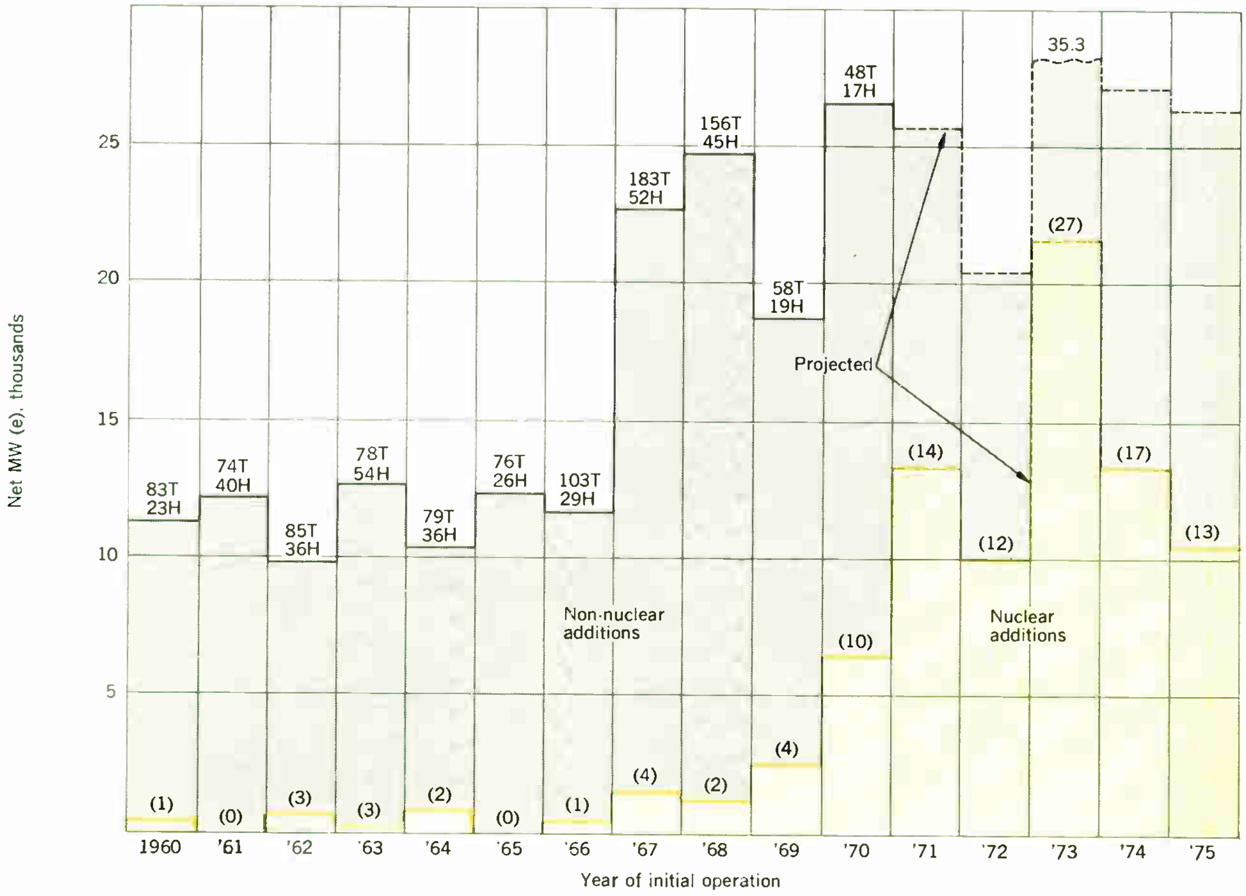
FIGURE 3. Trends in reactor size.



† One now under contract, others projected. * Seven now under contract, others projected.

FIGURE 4. Nuclear power reactors built, being built, or planned in the United States (by year of initial power operation).

FIGURE 5. Annual additions to U.S. electric generating capacity. Figures in parentheses indicate number of units.



II. Annual additions to U.S. generating capacity

Year of Initial Operation	Number of Nuclear Generating Units Added	Added Nuclear Generating Capacity, 1000 MW (e)	Number of Added Non-nuclear Generating Units		Added Non-nuclear Generating Capacity, 1000 MW (e)	Total Net Generating Capacity, 1000 MW (e)	Nuclear Additions, percent of generating capacity
			Hydro	Fossil			
1960	1	0.2	23	82	10.9	11.1	1.8
1961	0	0	40	74	12.2	12.2	0
1962	3	0.4	36	82	9.6	10.0	4.0
1963	3	0.1	54	75	12.6	12.7	8.0
1964	2	0.08	36	77	10.2	10.3	0.8
1965	0	0	26	76	12.6	12.6	0
1966	1	0.04	29	102	11.7	11.7	0.3
1967	4	1.4	52	179	21.3	22.7	6.2
1968	2	1.2	45	154	23.7	24.9	4.8
1969	4	2.6	19	53	16.0	18.6	14.0
1970	10	6.5	17	48	20.1	26.6	24.4
1971	14	13.5			12.2	25.7	52.5
1972	12	10.0			10.3	20.3	49.2
1973	27	21.6			13.7	35.3	61.1
1974	17	13.6			13.6	27.2	50.0
1975	13	10.4			15.7	26.1	39.8
Total	113	82.7			226.4	308.0	23.6

port on the civilian power program and prospects as they appeared at the start of 1967. This report, identified as the 1967 Supplement to the AEC's 1962 Report to the President, takes note in some detail of the changes that have taken place since 1962 in the technical, economic, and resources picture. It is intended to provide an updated background for those further intensive studies currently under way.

The Supplement reaffirmed the specific objectives summarized in the 1962 Report to the President:

1. The demonstration of economic nuclear power by assuring the construction of the plants incorporating the presently most competitive reactor types.
2. The early establishment of a self-sufficient and growing nuclear power industry that will assume an increasing share of the development costs.
3. The development of improved converter and, later, breeder reactors to convert the fertile isotopes to fissionable ones, thus making available the full potential of the nuclear fuels.
4. The maintenance of U.S. technological leadership in the world by means of a vigorous domestic nuclear power program and appropriate cooperation with, and assistance to, our friends abroad.

The 1967 Supplement also noted that the Commission intends to continue to exercise positive and vigorous leadership in achieving the technical goals and in assuring growing participation by the nuclear industry as nuclear power becomes economic.

We are pleased with the tremendous response from the government, industrial, and utility organizations in terms of the talented technical, analytical, and management personnel assigned to work with us on these assessment efforts and with their awareness that these efforts must be primarily oriented to the technical areas. Nevertheless, although we believe we have some of the best U.S. talent working in this area, we realize that many of the assumptions we are now making may be subject to significant changes. New information (in areas such as reactor plant sales, initial operating experience with the new large plants, identification of new ore reserves, progress in safety assessments and licensing procedures, and results from development programs for advanced reactor concepts) may significantly change our present approach.

The assessments now in progress are not delaying reactor development work, at different levels of support, on a number of advanced reactor concepts and programs, including the high-temperature gas-cooled reactor, the heavy-water-moderated reactor, the seed-blanket light-water thermal breeder, the molten-salt thermal breeder, and the liquid-metal-, steam-, and gas-cooled high-gain fast breeders. In addition, increased emphasis has been placed on supporting technology programs, including fuels and materials, physics, safety, coolant technology,

> Advanced converters are actively studied and developed in the European Community for exactly the same reasons that have prompted their development elsewhere; namely,
>
> 1. They should reach a lower cost of electricity than the proven-type reactors.
> 2. They should make a more efficient use of uranium, and build up more quickly than the proven-type reactors the plutonium amounts needed for the fast reactors.
> 3. They should be technically ready in the early future, and be assured of one to two generations of useful life before the breeders reach the point of self-sufficiency, achieving both low cost and doubling time shorter than the doubling time of installed nuclear power.
>
> *Jules Gueron*
> *European Atomic Energy Community*

> The special distinction of the Canadian position is that what most people understand by "planned advanced reactor concepts" seem very much like the reactors we now have under construction. Even when we look a hundred years ahead we see natural-uranium heavy-water-moderated reactors right there in the competition. I do not think this is because we underrate the competition. We see the fuel cost lying in the range of 0.3 to 0.6 mill/kWh in large generating stations of several millions of kilowatts capacity. We cannot tell whether steam will still be the favored working substance—but that takes us beyond the present topic, for nothing other than steam is yet planned.
>
> *W. Bennett Lewis*
> *Atomic Energy of Canada, Ltd.*

instrumentation and control, components, and fuel recycle.

Fast breeder reactors

Although the ongoing assessments may have an impact on timing, there is no question but that our objective is to permit the tapping, in an economical manner, of the unlimited amount of latent energy available in our uranium resources and to use most efficiently and economically the plutonium that will be supplied by the light-water-reactor complex. This can be done only by high-gain fast breeders. The development of these breeders will increase the fuel utilization from the 1 to 2 percent now available to over 50 percent. These high-gain breeders will be capable of providing an excess of burnable plutonium, which will be used to fuel new fast breeders in an expanding power system.

For this reason, the sodium-cooled liquid-metal fast-breeder reactor (LMFBR) has been established as the highest-priority phase of our civilian power program. Primary considerations leading to this selection were potential economy and fuel utilization, reactor manufacturer and utility interest, and technological experience gained in the United States and abroad. In addition, worldwide interest is concentrated on the sodium-cooled breeder.

Based on the AEC assessment that the LMFBR program must be given high priority, formal intensive and comprehensive reviews are continuing for each technical element of the program to determine the current state of the art, to identify the objectives, and to prepare detailed program plans to help achieve these objectives in a timely manner. These assessments are (and have been) based upon the proposition that fast breeder reactors of the future will be built by U.S. industry and will operate in a utility environment.

Much of the assessment and planning effort is being carried out by the LMFBR Program Office at the Argonne National Laboratory (ANL) in Illinois. The Program Office is an organization of over 30 full-time senior scientists and engineers who assist the Division of Reactor Development and Technology (RDT) in the detailed planning and evaluation of the LMFBR program. The Program Office staff is seeking the advice of each of the participants in the LMFBR program in the preparation and review of the program plans in order to take maximum advantage of the talent and experience available in the United States and abroad.

The utilities and utility-sponsored organizations are participating in many of the planning phases. In addition, we are encouraging potential reactor plant contractors to bring the utilities into their planning and to encourage the utilities to participate with them in studies for future demonstration plants and for the larger, commercially economic plants.

The government's role

On the basis that the proper role of government is to take the lead in developing and demonstrating the technology in such ways that natural economic forces will promote industrial applications and lead to a self-sustaining and growing nuclear power industry, the Commission intends to proceed under arrangements similar to those made in the past. The Commission will, generally speaking, participate in the development of technology directly, working through its own laboratories and with industrial contractors. Cooperative programs will be used, if possible, for demonstration reactors, bringing to bear both the developing industrial capability and the utility interest through direct participation and sharing of the required investment and associated risks. Experience has shown that a number of demonstration plants will be necessary before reliable and economic power plants for the utility environment are achieved.

Since the utilities are the ultimate customers for LMFBR plants, it is well to examine the requirements that must be met before the LMFBR plant can be introduced into the utility environment. The incentives must be attractive. They must include cheaper electricity and a market for plutonium from water reactors. Acceptance also will be contingent on developed technology, on the existence of a competitive and self-sustaining industry, and on a minimum investment of risk capital.

We are taking many positive steps to prepare the detailed research and development plans and to work with industry and the national laboratories to insure that the national plan of action is meaningful. The LMFBR program plans must be implemented by the requisite research and development necessary to resolve the technical problems, to develop the LMFBR technology, and to obtain solutions to the many engineering and application problems and questions on safety. This work includes the conduct of systems studies and engineering studies and the development and implementation of proper codes, standards, and specifications. In this regard, there has been during recent years a growing recognition of a matter crucial to the successful development of nuclear power plants: the achievement of demonstrated reliability through strong engineering and through increased experience gained from component, system, and plant testing and operations. The coupling between these factors and reactor safety must be recognized and the implications clearly understood. Delays, increased costs, and in some cases cancellation of experimental and demonstration reactor projects have been at least partly the result of inadequate depth in the bases for reactor design and engineering. Such inadequacies can result in an inability to

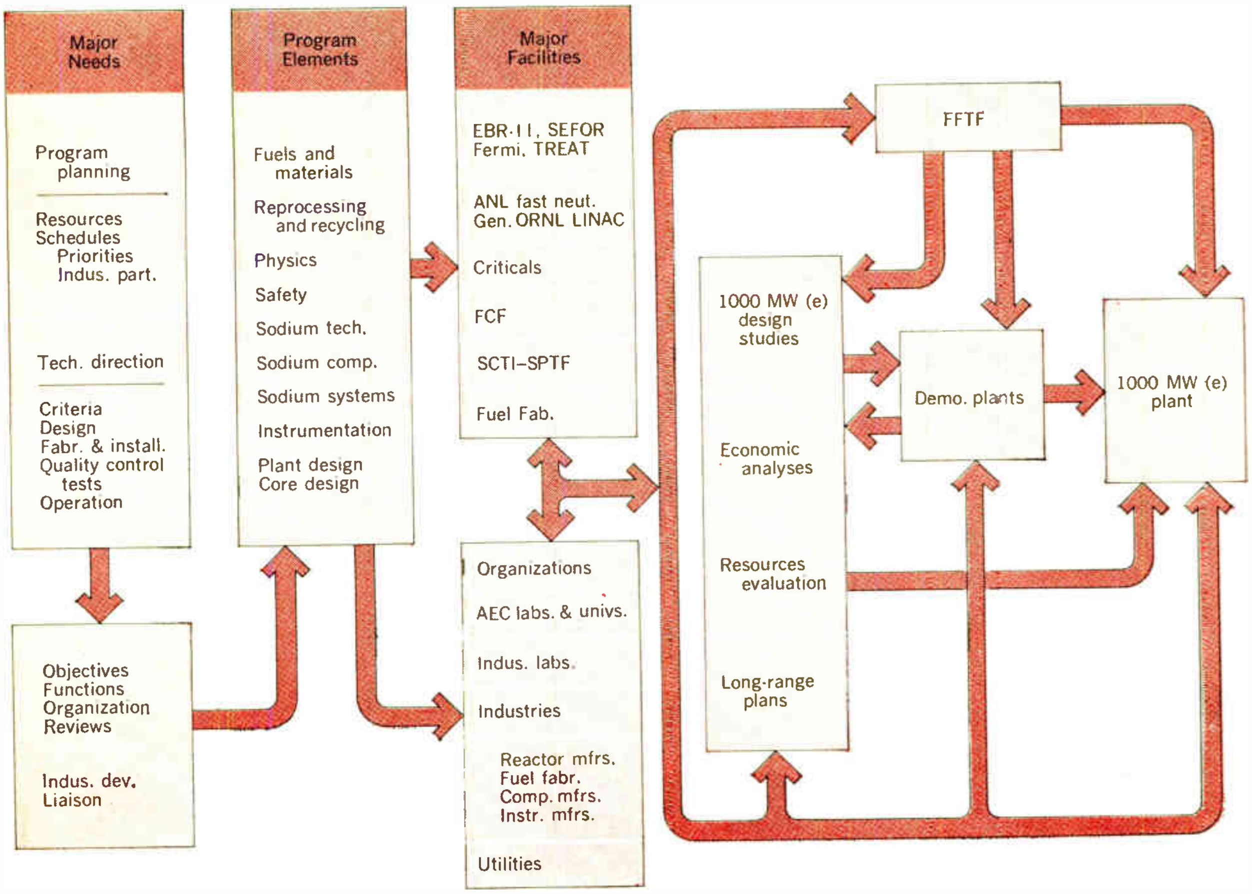


FIGURE 6. The path to the LMFBR objective.

demonstrate the worth of the concept under study, and we are taking rather strong steps to minimize such a likelihood in the future on the Commission-sponsored advanced reactor programs.

Thus, in exercising the role of positive and vigorous leadership recommended in the 1962 Report to the President and reaffirmed in the 1967 Supplement, the Commission intends to place particular emphasis, in cooperation with industry, on the development and application of a continually improving standard of design, construction, and operating practice. Such upgrading applied to the Commission's own programs and experimental facilities, to cooperative programs, and throughout the industry, should become an important factor in insuring meaningful results from experimental and demonstration programs, in improving the economics, in providing the high degree of reliability and availability necessary for utility application, and in continuing with increasing assurance the nuclear industry's excellent record of safety.

Long-term stability

As has been so evident with the light-water reactors, once reasonably economical technical and engineering solutions are available, the key to long-term stability in the LMFBR power-plant business is a competitive self-sustaining industry. The development of such an industry requires that the AEC, in addition to promulgating a national development program plan, take the lead in encouraging interest and participation by top management of industrial organizations and in assisting these organizations to build and strengthen their LMFBR capabilities. This includes aiding industry in identifying potential strengths, weaknesses, and priorities; guiding program-oriented research and development and production contracts to industry; providing training opportunities in AEC facilities; and participating with reactor plant contractors in the design, construction, and operation of demonstration plants.

Finally, the AEC intends to insure a minimum investment of risk capital by the utilities. This can be done by adequate appraisals of LMFBR engineering and technology with respect to operation in the utility environment, proceeding with demonstration plants large enough to demonstrate key LMFBR plant features, yet small enough for acceptable risks; by encouragement of guarantees to utilities by reactor manufacturers; and by financial participation of the AEC in the development of the demonstration plants.

The path to the LMFBR objective is illustrated in Fig. 6. The needs for the program planning and technical direction have been discussed. While each of the program elements is being examined and redirected as necessary, major emphasis is being placed on building test facilities to support adequately the necessary research, development, and test programs. These government-owned facilities will be available to all laboratory and industrial organizations in the LMFBR community, and their importance will be even more evident in the next few years. The availability of these vital LMFBR facilities, particularly the Fast Flux Test Facility (FFTF) for testing fuels and materials, will be an essential key to a viable, self-sustaining LMFBR industry of the future. The most important, most complex, and most difficult of these facilities is the FFTF to be built at Richland, Wash.

The primary objective of the FFTF is to provide a fast flux testing capability for the LMFBR program and for other AEC programs. Secondary objectives include contributing to the development of the following: competitive industrial capability; systematic methods for safe and economical plant design, construction, and operation; adequate codes, standards, and specifications for LMFBR plants; and sodium systems and components for LMFBR plants. The facilities at the National Reactor Testing Station (NRTS) in Idaho, and at the Liquid Metal Engineering Center (LMEC) in Santa Susana, Calif., will play increasingly important roles in the development of advanced reactors, particularly in the LMFBR program.

The overall verification of the R & D program will be accomplished in the first demonstration plants. As did the light-water demonstration plants, these first LMFBR plants will provide the necessary confidence to proceed with the design and construction of commercial LMFBR plants of 1000 MW or more of electrical output.

A ten-year forecast

The next decade will be a busy one for the nuclear community. Much of the basic R & D to establish a sound technological base for a competitive, self-sustaining nuclear industry must be accomplished during the next ten years. The results of our current studies should lead to further concentration on the more promising concepts while narrowing the range of concepts and technologies under active consideration. There should be considerable progress made toward expanding the applications of nuclear power from electric power generation to the desalting of seawater to help meet our growing requirements for fresh water, and to other process applications, such as the production of chemicals and metals, the synthesis of ammonia for fertilizers, the electrolytic production of aluminum, and the production of magnesium from seawater by electrolysis.

In the next decade, considerable effort will be applied in the areas of reactor safety and engineered safety features and in radioactive waste management systems for advanced reactor concepts as well as for existing commitments. Emphasis will be placed on research that will provide the knowledge necessary for determining the safety problems associated with all reactors and also will provide the base for developments needed to resolve these problems. The developments may include, where necessary, the demonstration of the causes and consequences of abnormal conditions, along with the development of countermeasures for these conditions.

Programmatic and project efforts for specific reactor concepts must have the support of applied R & D activities needed to generate the technological base underlying the initiation and accomplishment of specific reactor programs and projects, to resolve critical problems, and to assess continually the ongoing programs and projects. The Commission's general reactor technology programs conducted in industrial, university, and Commission laboratories provide this support and, in addition, open up new areas of technology upon which future advances may be based.

Key facilities, such as the FFTF, will be designed and constructed. Within the next decade, they are expected to be contributing vital design information and experimental results for use by the government, its laboratories, and industry. Reactor plant manufacturers and utilities will be working together with the AEC in the design, development, construction, and operation of the first demonstration breeder power plants.

In ten years we should be looking forward to the early commitment of second-generation breeder demonstration plants, and of the commercial plants, 1000 MW (e) or larger, that will be built in the 1980s. These must be a product of a viable reactor industry, accepted by the utilities as a necessary step in the economic and yet efficient utilization of our uranium resources. We are encouraging the reactor manufacturers and the utilities to work together from the early planning stages on through to the economic power plants. The AEC will continue to encourage their participation in reactor development programs on a sound technical and management basis, which includes a recognition and understanding that careful and high standards of design, construction, inspection, and operation of nuclear power plants are basic to the reliable and safe performance needed for utility operation.

Conclusion

In summary, within the next decade, the advanced reactor programs in the United States will provide a complex, exciting, and challenging task for the Commission's staff and for all other participants. The next decade, replete with tremendous developmental and reactor plant commitments, should culminate in successful achievements based on the confidence reached by mutual cooperation between reactor plant manufacturers, component suppliers, industrial and national laboratories, the utilities, and the AEC. This confidence is needed to achieve a viable, competitive, and self-sustaining nuclear industry.

All major nuclear countries are working in the field of fast reactors, but the U.S.S.R. and the U.K. are unique in that they already have under construction prototype models. The reasons for proceeding with construction in the U.K. of a prototype fast reactor now are:

1. The anticipated reduction in generating costs.
2. The clear need and benefit in the long term to offset rising uranium prices.
3. The satisfactory demonstration of fuel behavior and analysis of key safety features.
4. The long time scale for the development of alternative systems and the marginal benefits at present anticipated from them.
5. The advantage which accrues from being in the field first (so long as one is successful).

Peter W. Mummery
United Kingdom Atomic Energy Authority

Supervisory control of remote manipulation

Man is expanding his range of research to encompass a new universe that extends from the confines of an atom to the fringes of outer space. Exploring such a universe requires a new breed of manipulators to perform tasks accurately and intelligently in areas inaccessible to man

William R. Ferrell, Thomas B. Sheridan
Massachusetts Institute of Technology

The relatively short distances to be spanned in using the manipulators presently available permit real-time control systems, frequently employing direct linkages. The advent of space flight necessitates the design of manipulation systems that can perform complicated tasks, on the moon and beyond, upon command from earth. Such a design must compensate for the communications time delay due to the distances involved and also for a difference in environment not directly observable by the human operator. The answer to the problem seems to lie in a computer-controlled remote unit, capable of making limited decisions of its own but supervised from home base.

Remote manipulation involves the use of a mechanical device to perform a task that, if it were more accessible to the operator, would normally be done by hand. A task may be remote either in distance or in scale. The protective barriers of "hot cells" for containing radioactive or toxic materials, the reaches of space, and depths of the ocean can all be spanned by remote manipulators. These manipulators can also bridge the gap between the capabilities of the human hand and those necessary to move large quantities of earth or structural steel, or to perform surgery on single living cells. Just as instruments have extended man's senses in range and resolution, so will they extend his ability to manipulate objects within environments never before opened to him.

The most promising approach to rapid and precise manipulation over short distances and within relatively accessible areas is to link the remote devices as tightly and as compatibly as possible with the operator. The objective is for his hands to feel and perform as if they were, in fact, at the remote site. Representative of such devices are the AEC Argonne Laboratory series of "master–slave" manipulators, a pioneering development of R. C. Goertz begun in the early 1940s.[1] The human operator controls a "master" arm, with articulation resembling a man's except that the elbow telescopes instead of bending. A "slave" arm duplicates motion in all degrees of freedom, and feeds back to the master any resistance it encounters. The slave may be coupled to the master directly, by mechanical linkages or steel tapes, or the connection may be electrical or hydraulic, as in manipulators built by R. S. Mosher[2] of the General Electric Company. His "Handyman" is a classic example, demonstrating the advantages of close coupling between the operator's arm and hands and those of the remote device.

There are, however, circumstances in which a human operator will not be able to act as if he were on the spot doing the manipulating himself, and performance will suffer correspondingly: (1) The operator may be handicapped, or other tasks may share his attention. (2) Signal transmission may be affected by noise or power limitations. (3) If the distance to the remote site is very great, there will be a long transmission delay.

Two alternative approaches have been taken to obtain

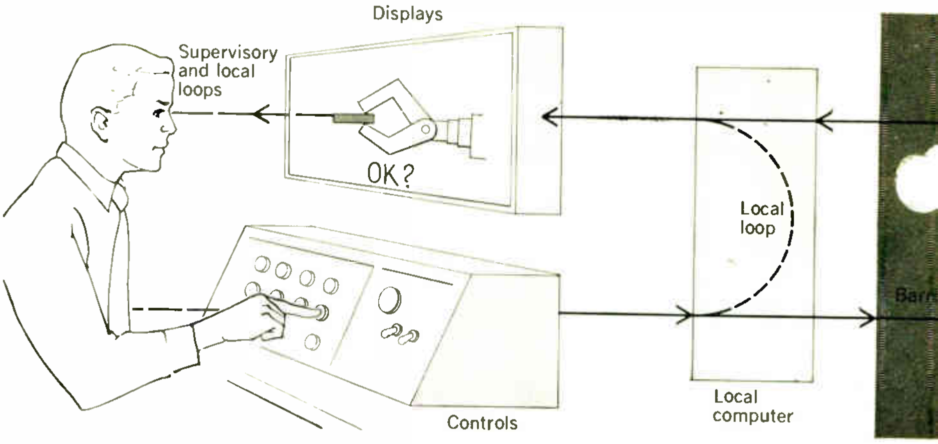


FIGURE 1. Schematic diagram of supervisory control of remote manipulation.

versatile remote manipulation when it is impossible to get good results with tight, closed-loop control of remote actions by the operator. The first approach is to build a robot that can perform the necessary tasks independently. It might be remotely programmed occasionally, perhaps, but would be relied upon, for the most part, to cope with all contingencies. The second approach is to design a system having two separate levels of control:

1. Direct control of the manipulator by a computer at the remote site, which processes feedback and makes relatively simple, routine decisions.

2. Supervision by an operator who occasionally sets sequences of subgoals for the remote device, modifies its parameters, perhaps, and compensates for its limited decision-making capability.

There is substantial similarity between the robot and supervisory control approaches since they both require a degree of intelligent behavior from the remote system. But, although the emphasis in the former case is on the independent action of the machine, in the latter it is on mutual cooperation with the operator.

Automatic manipulation was first attempted by Ernst.[3] He developed a computer-controlled mechanical hand with a sense of touch that could "feel" about to find blocks then stack them or put them in a box, thus coping with a variety of disturbances and uncertainties, much as would a blind child. Further work in automatic manipulation has been conducted by Minsky and his colleagues of the M.I.T. Artificial Intelligence Group.[4] Recognizing the importance of a "visual" input to a manipulator, they have concentrated on optical recognition of objects and spatial patterns.

A computer-run manipulator supervised by an operator was first realized by Reswick and Mergler and their students[5] at the Case Institute of Technology; they made a device that would enable a quadraplegic to handle various objects such as tableware. Subsequent work at Case has dealt with rate-controlled industrial manipulators, and the use of digital computers to simplify the operator's task and reduce the required operating time.[6]

In the authors' view, the long-term problems most critical to remote manipulation concern the ways in which human and artificial intelligence should interact. Their approach to the supervisory control of manipulation has grown out of a general concern with human response and decision-making abilities in control tasks, plus a specific interest in the problem of long transmission delay affecting an operator's use of a remote manipulator.

If there is a delay in the loop, operators of manipulators tend to adopt a strategy of performing tasks by a sequence of open-loop moves, each one followed by a wait of one loop delay time, for correct feedback.[7] In this way, high accuracy and reliability can be obtained, but completion time is approximately linear with delay. When the delay is very long, the time it takes to get a job done is correspondingly great. The average loop delay of 2.6 seconds, obtained when operating a manipulator on the moon from earth, might be tolerable; but the minimum delay of 188 seconds with one on Mars would certainly be excessive. An increase in the amount of task the operator can command without feedback and have the machine successfully perform would accelerate the work, since the time taken depends directly on the number of open-loop moves, each costing a delay. Such an increase can be accomplished by giving the remote device a degree of autonomy, so that commands for, and feedback from, every elementary action need not cross the barrier separating the operator and the remote site.

Characteristics of supervisory control

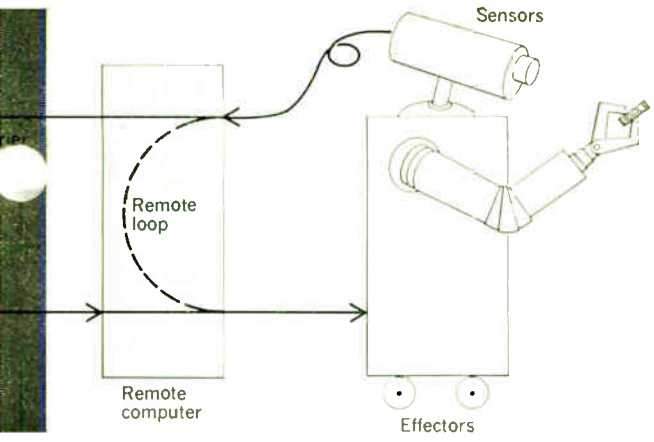
The fundamental components of a supervisory control system for manipulation are shown in Fig. 1. The usual features of a man–machine system diagram are present, but three loops are shown, each representing a type of function:

1. The remote loop, closed through the computer at the distant site, represents the remote device acting as an automaton with a short-range goal.

2. The supervisory loop, closed through the man, represents his functions of intermittently setting goals for the remote device, adjusting its characteristics, and, if needed, even guiding the effectors directly.

3. The local loop is independent of the remote device, and signifies the operator's use of his computer to model the remote system so that he may predict its behavior and hence improve his supervision. Thus, through mimicry, the local device can provide the operator with "quasi-feedback" not subject to transmission delay.

The local computer, in addition to serving as an assistant, translates the outgoing and incoming information between a form appropriate for transmission and machine use and a form suited to the man's displays and



controls. In this way we avoid using more expensive remote computing capacity for this purpose.

In the design of such a supervisory system, two interrelated problems arise: setting up a method of communication between the operator and the machine; and enabling the remote device to achieve short-range goals independently.

Man-machine communication

In order to supervise a manipulator, an operator will have to divide the task into a sequence of subgoals that the system can achieve on its own. If the system has the capacity to handle quite large chunks of a task, this may not be a difficult process; but if the goals it requires are of such short range that the operator normally doesn't consciously think about them, he is likely to find them hard to express explicitly in a formal symbolic language. This difficulty suggests that two control modes are necessary: *analogic control*, in which there is a degree of geometrical or dynamic similarity between the operator's controlling movements and the machine's interpretation of them; and *symbolic control*, in which commands are expressed in a formal language.

In the latter mode, the symbols must be easy for the human operator to string together into meaningful sequences (such as words in a sentence or instructions in computer-compiler language), be quickly communicable (as typing alphanumeric symbols), and be unambiguously interpretable by the computer. Analogic commands can be used for supervisory purposes, as well as for direct control. The kind of interpretation to be given to movement patterns should be selected by the operator, using symbolic code. Just as different kinds of control suit the operator for different kinds of commands to the machine, different displays are appropriate for the reverse communication. For example, an operator can interpret a geometrical situation more quickly when it is presented in the form of a picture or graph rather than as a string of alphanumeric symbols. Similarly, a successful maneuver need not be shown in detail; "OK" may be a sufficient response from the machine.

Complementary modes of control and display have been found useful in other contexts, such as space vehicles, in which astronauts use both hand controls and a computer keyboard for various positioning maneuvers. Computer-aided design systems also are analogous to supervisory manipulation. It is common, for example, to use the light pen to indicate a line segment and a keyset to specify the class of constraints, such as whether a line is to be interpreted as straight or circular, as joined to other lines, etc.

Remote feedback processing and decision making

In order for the remote device to carry on some degree of independent activity directed toward a goal, it must have, in addition to effectors, the capacity to obtain and process feedback, and to make decisions.

It was pointed out by Ernst[3] that an autonomous manipulator must have a picture of the world, and a notion of its own location and state, in order to predict the consequences of its actions, and hence to generate alternative courses of action, and decide among them. An internal representation of the environment has two fundamental characteristics that determine the amount of information it contains about the world: the number of relevant dimensions of the environment; and the resolution along any single dimension. The former, although dictated largely by the nature of the environment itself, may also be a function of the task. In most ordinary manipulations the purely mechanical dynamics can be neglected, since the mechanical friction and inertias are small and motor speeds and forces pose no limitations. The necessary resolution along any dimension depends partly upon how crowded the environment is, and partly upon the availability of feedback. For example, to touch an object gently, one either must know its position quite exactly or have a tactile sense to indicate when contact with it has been made.

Simply because predictions often are not borne out and environments change, such feedback as there is must be able to modify the internal representation of the environment. In at least this elementary sense, the system must be able to learn from experience.

The form of the internal representation is determined largely by the necessity for the machine to

1. Generate sequences of realizable actions that will result in a desired state.
2. Evaluate the effects of each sequence.
3. Decide among the alternative sequences.
4. Be able to use the same procedures for determining how to go from any state to any other.

The first two requirements mean that all possible sequences of elementary actions must be distinguishable in the internal representation, and assigned a measure of desirability or cost. Of these, the most desirable or the least costly is chosen. Clearly, if every possibility is to be examined, the tree of sequences of elementary actions will be very large, especially if the environment imposes few constraints. The magnitude of the problem is such that for complex behavior, heuristic methods, rather than complete algorithms, will eventually have to be used; and it will always be up to the operator to provide subgoals close enough to the current situation that the remote device can find ways of reaching them. Recursive procedures, though not absolutely necessary, are especially desirable in such circumstances. Without them, every case is a special case, and an enormous repertory of programs would be necessary to have a versatile system in a complex environment.

The frequency and kind of communication between the operator and the machine, as well as the balance between human and machine decision making, will depend upon the task, and upon the level of machine intelligence. If the remote computer-plus-manipulator is reliable at understanding and achieving complex subgoals, there may be little need for allocating computation capacity and transmission power to sense and feed back to the human a complete and precise picture of the situation. However, when things go wrong with a complex system, a considerable amount of intercommunication may be necessary to set them right. This suggests that, for a given remote computing capacity, there will be an optimum allocation to diagnostic and trouble-shooting routines.

Experiments in the laboratory

During the past two years several projects in the Man–Machine Systems Laboratory at M.I.T. have been concerned with supervisory manipulation. Corresponding to the problem areas described in the foregoing there has been emphasis on both man–machine interaction and machine manipulation.

Computer simulation of the manipulator. To obtain basic data on the way an operator uses a rate-controlled manipulator under conditions of intermittent and delayed feedback, S. G. McCandlish[8] simulated an entire manipulator, and its environment, on a PDP-1 computer. As part of his experimental work, he investigated the use of automatic subroutines for performing parts of the manipulator task. The operator, viewing a CRT display, performed a task that mimicked lifting a block from one hole and putting it in another, as shown in Fig. 2. The operator could control the vertical and horizontal velocities of a pair of manipulator jaws by rotating knobs, and could initiate their opening and closing with switches. The constraints closely approximated the real world. Motor lags and gravity, as well as tipping, bouncing, and breaking of the object, were all simulated, so that their effects were evident in the control and display of the task.

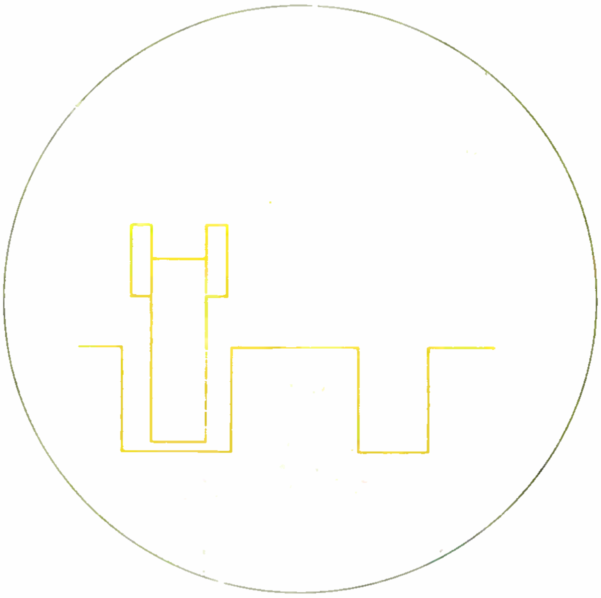


FIGURE 2. Oscilloscope display of a simulated remote manipulator and its assigned task.

After many hours of practice, operators were given the option of using subroutines to perform parts of the task automatically, such as centering the object over the hole. The display was turned off during automatic operations, but could be requested by the operator. The objective was to use a minimum total of motor commands and calls for subroutines. The performance measure was indeed found ultimately to decrease when the subroutines were used. But two interesting qualifications were observed: The use of subroutines did not, at first, reduce the total number of commands; the subjects performed less well on the portion of the task that could not be automated, using the new assistance to ease the difficulty of the task rather than improve the criterion score. This points up the need to assess the work load imposed on an operator, and the way he distributes it among elements of a task, if performance measures are to be properly interpreted. For example, he may tend to reject powerful subroutines that require a high level of strategy and foresight in their application in favor of less effective but also less demanding methods.

It was also observed that if a subroutine could not be counted on to perform its function with complete reliability, as when "bugs" were in the programs, the operators frequently sought reassurance that the system was still working by calling up the display. For the operator to have confidence in such a complex system, it will be important to incorporate sufficient error traps and diagnostics, features common to on-line computer-programming languages.

Laboratory experiments with actual manipulators. To encompass a wider variety of tasks and more realistic constraints than are possible with a wholly computer simulation, a workable system was made in which an AMF Model 8 master–slave manipulator, of the type commonly used in nuclear laboratories, is controlled through a small PDP-8 computer. Each degree of freedom of the manipulator is operated by a stepping motor. The computer supplies command pulses to the motors, and maintains a running count of the input to each one. This obviates the need for position feedback, provided the manipulator is not overloaded. The discrete position increments are quite small, at worst about 7.94 mm per pulse, and a series of them approximates a continuous motion.

Combined analogic and symbolic control has been developed through closely related projects by W. Verplank,[9] T. D. Rarich,[10] and D. Manalan.[11] The basic on–off controller is an analog of the manipulator arm, as shown in Fig. 3. It is worn by the operator on his shoulder so that he can move it to maintain its orientation relative to the remote arm. The correspondence between the degrees of freedom of the arm and the controller is not exact, however. The modifications are evident in the illustration, and were made to allow for the difference between the working positions of the man's and the manipulator's hands in relation to the wrist. Movements of the controller relative to the shoulder harness close switches, providing the computer with forward–stop–reverse commands, corresponding to the appropriate degrees of freedom.

Symbolic input through separate controls on the harness determines the way in which the computer interprets the commands. By means of these controls the operator can cause the manipulator to respond with either a constant velocity or an incremental change in position, and can vary the magnitude of the velocity or the size of the in-

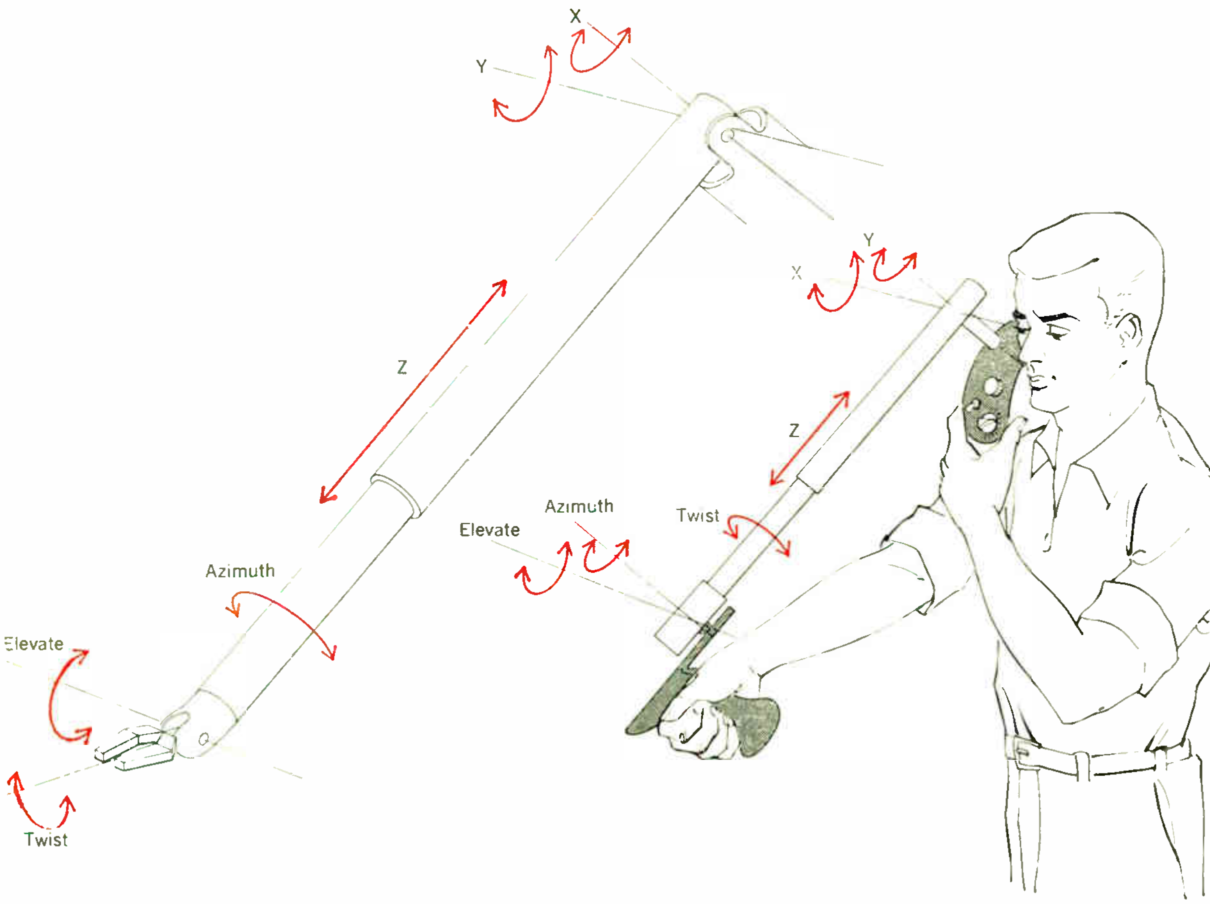


FIGURE 3. Analog controller for a computer-run manipulator.

crement. The operator has, at the same time, the option of requesting the remote hand to stop should it encounter an object.

It is far easier and more natural to use this device than to input the same commands through a teletypewriter; although, compared with a direct master–slave servo link between the controller and manipulator, the device is slow and clumsy. With a more extended supervisory system, "analog" information, such as that derived from a controller of this kind, could be interpreted by the computer as setting the values of variable arguments in subroutines. For example, using a parenthetical phrase to denote an appropriate movement of the analog arm, symbolic inputs might be of the form:

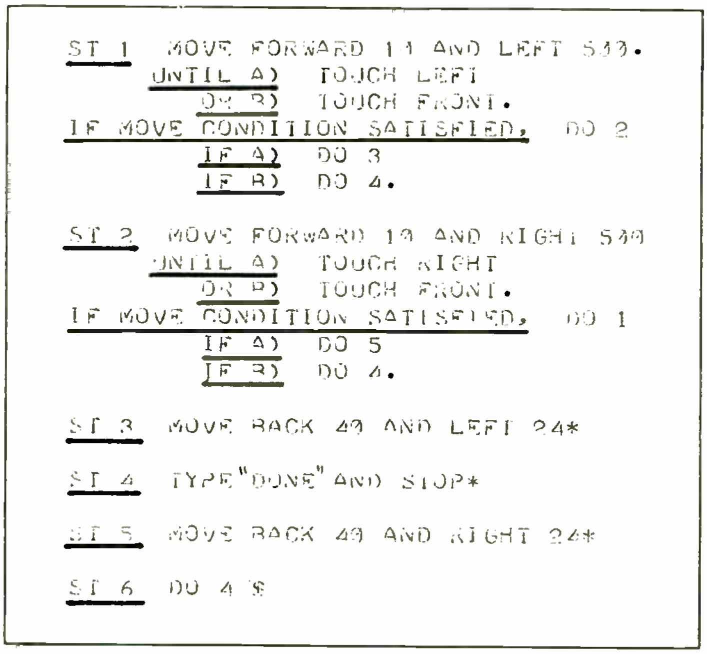
"GET (that object)"
"SEARCH (with this pattern in this direction)"

Supervision of a computer-run manipulator can be thought of as programming, and one approach toward efficient supervision is the development of a manipulation-oriented programming language. In order to shorten the time required by the operator to formulate instructions and communicate them to the machine, the language must have commands at least at the level of generality at which one would instruct a very small child, such as: "*Get object a.*" "*Put a on b.*"

Complex commands can be built up from a sufficiently powerful and flexible set at a more primitive level, a level which includes instructions for positioning and orienting the remote hand, for testing the state of the sensors, and also for conditional branching within the program itself. Subroutines written with these commands can then be used as basic elements at the next higher level. A compiler of this kind has been developed by D. Barber[12] for use with the computer-run manipulator described in the foregoing. At the primitive level, the language is constructed around a basic action: a movement in a given direction terminated on the achievement of specified sensor states, and/or a given distance moved. The variety of terminal conditions allows branching and the calling of subroutines. Objects, positions, and routines can be named and used in other programs. Insofar as possible, the computer handles the bookkeeping, checks for consistency, and provides the format in which the operator's commands are to be written. The objective is to make active programming possible during a manipulation task. In the program example shown in Fig. 4, the computer's typing has been underlined. Actually, the first letters of the major words are sufficient input ("F 10 L 500" for the first line).

Toward a theory of manipulation control

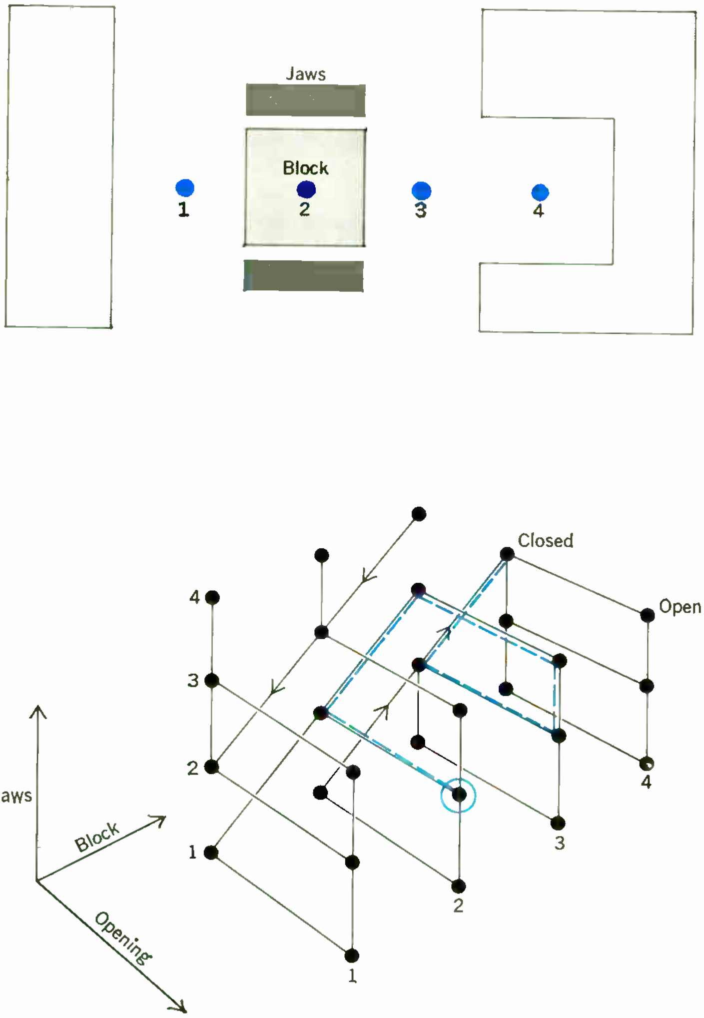
For orderly development of supervisory control of manipulation, means must be found to describe a manipulation task quantitatively, and to relate the task description to a specific sequence of elementary operations that will bring about the desired result. Although manipulation is qualitatively different from what engineers usually consider as control, D. Whitney[13] has been ex-



```
ST 1   MOVE FORWARD 10 AND LEFT 540.
     UNTIL A)   TOUCH LEFT
          OR B)   TOUCH FRONT.
IF MOVE CONDITION SATISFIED,   DO 2
          IF A)   DO 3
          IF B)   DO 4.

ST 2   MOVE FORWARD 10 AND RIGHT 540
     UNTIL A)   TOUCH RIGHT
          OR B)   TOUCH FRONT.
IF MOVE CONDITION SATISFIED,   DO 1
          IF A)   DO 5
          IF B)   DO 4.

ST 3   MOVE BACK 40 AND LEFT 24*

ST 4   TYPE"DONE" AND STOP*

ST 5   MOVE BACK 40 AND RIGHT 24*

ST 6   DO 4 $
```

FIGURE 4. Example of the use of Mantran, a manipulation-oriented computer programming language.

FIGURE 5. A simple task and its state graph.



amining ways in which a manipulation problem can be represented in terms of modern optimal control theory, and "solved" by use of mathematical programming techniques.

The configuration of the manipulation environment, which includes the manipulator itself, is represented as a state space, or graph, having one dimension for each degree of freedom of interest, and a finite set of states, or nodes, in each dimension. A simple example is given in Fig. 5 for a pair of jaws and a block. The jaws may be open, as shown, or closed; and both the pair of jaws and the block may translate in a single degree of freedom to one of four positions. Lines connecting adjacent nodes in the graph represent transitions between states, such as incremental displacements of the manipulator *or* manipulated object, or both, in the given degree of freedom. Some transitions either are not possible or can be accomplished only by means of one of a specified set of commands; for example, the block cannot change state at the same time the jaws are open. Some actions, such as pushing, are not always reversible, thus resulting in a directed graph.

Only the basic commands of open, close, move right, and move left are needed. The structure of the graph, however, explicitly represents the effective commands for push and carry by the diagonal branches. This important feature allows assignment of different costs to carrying out the same command under different circumstances. If pushing the block is risky enough, then putting it in the hole from the position illustrated should involve carrying it first rather than pushing it all the way, and the path shown in color, Fig. 5, will be optimal.

Any manipulation task is represented by specifying all states of interest and all transitions between them, the cost of transition, the initial state, and the terminal or desired state. Any sequence of elementary commands that forms a path between the initial and desired states is a possible way to perform the task. The problem is to determine whether a path exists and, if there are two or more, to choose the one having least cost.

The procedure by which one can find an optimal path through the manipulation state space graph is most easily seen by considering a simple example. The graph of Fig. 6 has seven states represented by nodes. Each of the possible transitions between states is shown, and is labeled with the cost of making it. The first step is to associate with each state a cost index (shown in the circle at each node). Initially set at a dummy value, this index will be progressively adjusted until it represents the minimum cost of getting to each state from the starting state.

Naturally, the cost of getting to the initial state from itself is zero, and its index can be set at once. To start with, each of the other states is assigned a cost index with the dummy value of ∞. Next, the cost indexes are adjusted by making them satisfy the requirement that any portion of an optimal path is also an optimal path. This is done by taking each state in turn and setting its index equal to the minimum of the sums of the index at, and the transition cost from, each of the adjacent states.

Two orders of considering the states for evaluating their cost indexes are shown in Fig. 6 in color. In the evaluation for each state, its adjacent states are also treated in order—in this case, counterclockwise from 12 o'clock. Alternative equal-cost paths are eliminated by accepting the first minimum obtained. With each evaluation, the state is

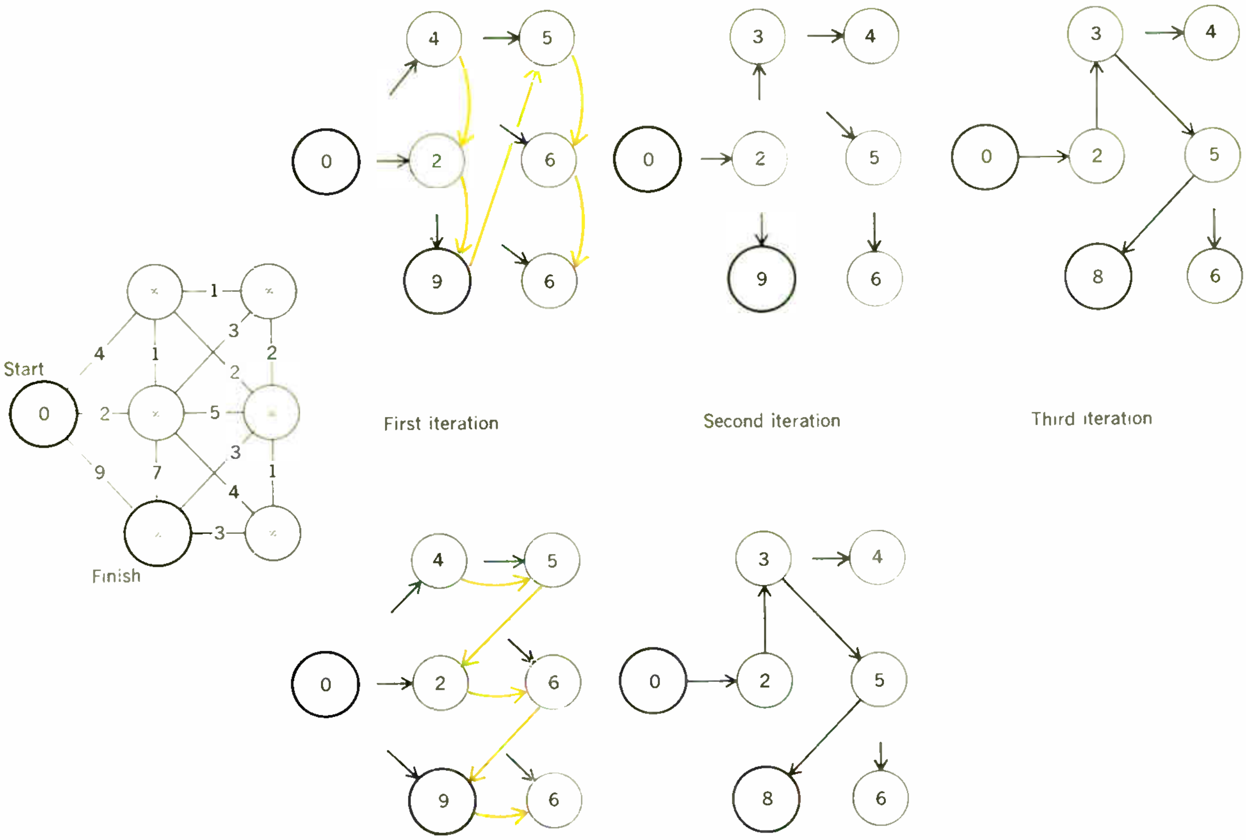


FIGURE 6. Iterative relaxation of a state graph to obtain an optimal path.

tagged with the direction—or command—by which it is arrived at to give the current cost index. This allows the path, from the start to any state, to be backfigured.

Using the order shown in color in the first row, Fig. 6, three iterations are necessary to reach the equilibrium in which all indexes are the minimum costs. Of course, a fourth iteration was needed to verify that the indexes didn't change. The alternative order shown in the second row requires two iterations, plus a third for verification.

Two provable points are worth noting: The optimal path from the start to all reachable points is automatically obtained; and the number of iterations will never be greater than the number of nodes in the graph, but may be far less, depending on the complexity of the optimal paths relative to the order in which the states are evaluated.

The procedure just described is a relaxation algorithm, attributable to L. R. Ford.[14] When it is known that every trajectory must pass through one of a first set of states, then one of a second set of states, and so on, the cost of reaching points in the state space may be evaluated according to this ordering relation, and only one iteration is necessary. This restriction on the Ford algorithm, the equivalent of R. Bellman's dynamic programming, is not possible in the manipulation state space, because the optimal path may detour, back up, or otherwise vary.

A major advantage of having an operator in a supervisory capacity can be illustrated by reference to Fig. 7, a manipulation problem in a plane requiring control of the jaws to reposition block *A* to *A** without colliding with either the wall or another block, *B*.

If the jaws may be open or closed, and there are ten discrete positions on each degree of freedom for each object and for the jaw assembly, then a "brute force" optimization procedure for the whole task would require a graph of (2 × 10 × 10) × (10 × 10) × (10 × 10) = 2 × 10^6 states. However, if a human supervisor partitions the task into two subtasks, "move block *B* to *B**" and "move block *A* to *A**," then even the brute force approach is broken into two optimization computations with a graph of 2 × 10^4 states, 1/100th of the previous requirement. With more subgoals or more clever assignment of variables, the computation could be reduced still further.

Whitney has successfully implemented the state space approach by using a PDP-8 computer and a simple three-degree-of-freedom manipulator equipped with touch sensors. Moving in a plane with a 10 × 10 grid, the jaws follow optimal paths to grasp or place objects, while keeping track of other objects and boundaries and allowing for their own inability to rotate.

Conclusions

Supervisory control will be necessary for effective remote manipulation when there are long time delays and when the task is not defined in detail beforehand. The graph of Fig. 8 is a qualitative representation of the effect of increasing delay on the performance of a complex task by systems having different proportions of human and machine decision making. Completion time is depicted as indeterminate but large when the task is left entirely to the machine, because an unspecifiable degree of versatility is desired. At that limit, the machine would have to be a robot, with the same goals and ability as the system

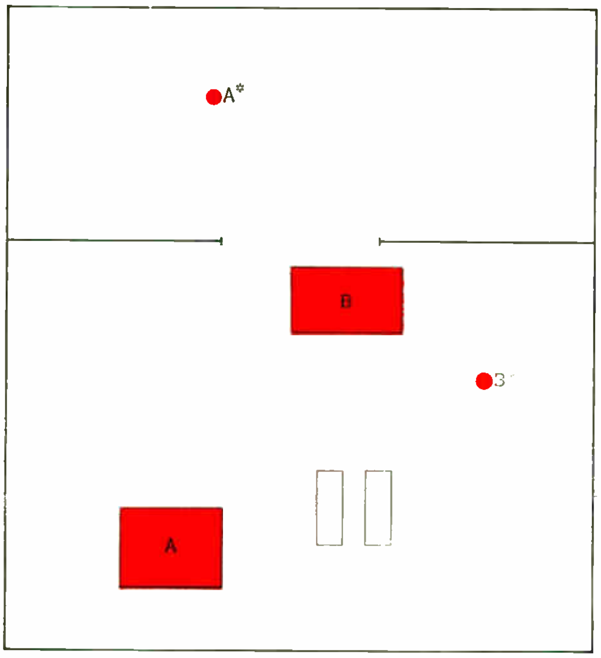
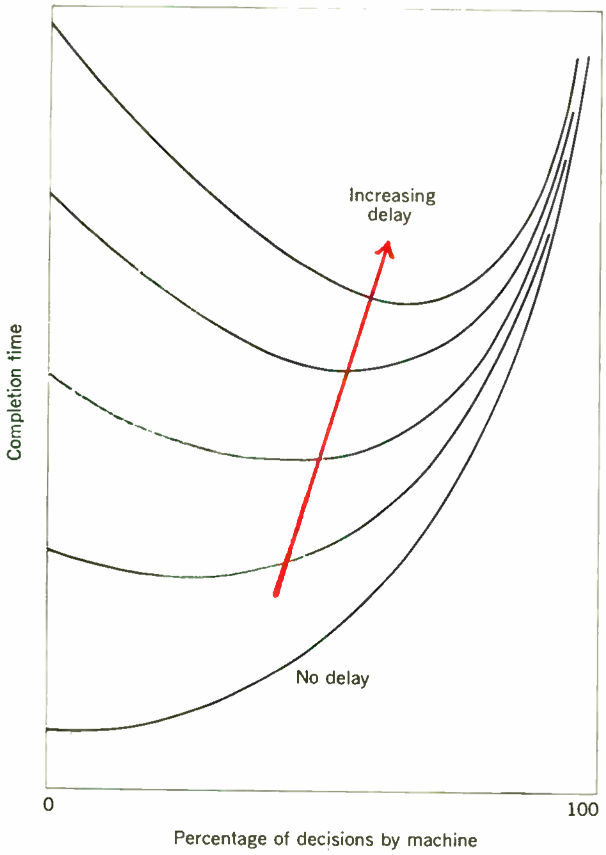


FIGURE 7. Movement of A to A* is greatly simplified if the operator specifies movement of B to B* as a subgoal. The manipulation problem requires control of the jaws to reposition block A without colliding with either the wall or another block, B.

FIGURE 8. Effect of delay on task completion time. Systems have different proportions of human and machine decision making. At one extreme, the decisions are entirely by machine, with a minimal completion time. At the other extreme, supervision is entirely human; and the completion time is longer because of necessary move-and-wait strategy.



supervisor. However, a compromise involving a small but finite amount of supervision is possible, and would be characterized by major reliance on the machine. At the opposite extreme, the operator is in direct control of the remote device, but must sacrifice completion time by using a move-and-wait strategy in order to maintain precision. In the region of combined man-and-machine control, the operator is able to extend his open-loop "moves" so that he gives fewer but more comprehensive commands. With fewer commands there are correspondingly fewer waits for correct feedback. When the delay is short, the operator can perform the task most quickly by using direct control; but with greater delay, the minimum completion time occurs when progressively more reliance is placed on the machine.

Within the range of moderate delays, the potential advantages of supervisory control will depend strongly upon both the efficiency of the communication between operator and machine and the existence of powerful programs to give the remote device a degree of autonomy. The subtle problems in these two areas are yielding to efforts in a number of laboratories. As implementation in hardware is not a major difficulty, practical supervisory manipulators will be possible in the not too distant future. Although they are especially suited for use in space, their relative independence from the operator recommends them for undersea, industrial, or laboratory applications, whenever central control or feedback information is severely limited. Perhaps even more important, the studies leading to aided remote manipulation may also lead to a fuller understanding of the general problem of man–machine cooperation.

REFERENCES

1. Goertz, R., "Some work on manipulator systems at ANL, past, present, and a look at the future," *Proc. 1964 Seminars on Remotely Operated Special Equipment*, U.S. Atomic Energy Commission Rept. CONF-640508.

2. Mosher, R., "Industrial manipulators," *Sci. Am.*, vol. 211, pp. 88–96, Oct. 1964.

3. Ernst, H. A., "MH1—a computer operated mechanical hand," Sc.D. dissertation, Massachusetts Institute of Technology, 1961.

4. Minsky, M. L., "An autonomous manipulator system," Proj. Rept. III, Project MAC, July 1965–July 1966, Massachusetts Institute of Technology.

5. "Medical Engineering," Prog. Rept. 1, Case Institute of Technology, 1962.

6. Taylor, R. J., "A digital interface for the computer control of a remote manipulator," Contractor Rept. CR 80843, National Aeronautics & Space Administration, 1966.

7. Ferrell, W. R., "Remote manipulation with transmission delay," *IEEE Trans. Human Factors in Electronics*, vol. HFE-6, pp. 24–32, Sept. 1965.

8. McCandlish, S. G., "A computer simulation experiment on supervisory control of remote manipulation," S.M. thesis, Massachusetts Institute of Technology, 1966.

9. Verplank, W., "Symbolic and analogic command hardware for computer-aided manipulation," S.M. thesis, Massachusetts Institute of Technology, 1967.

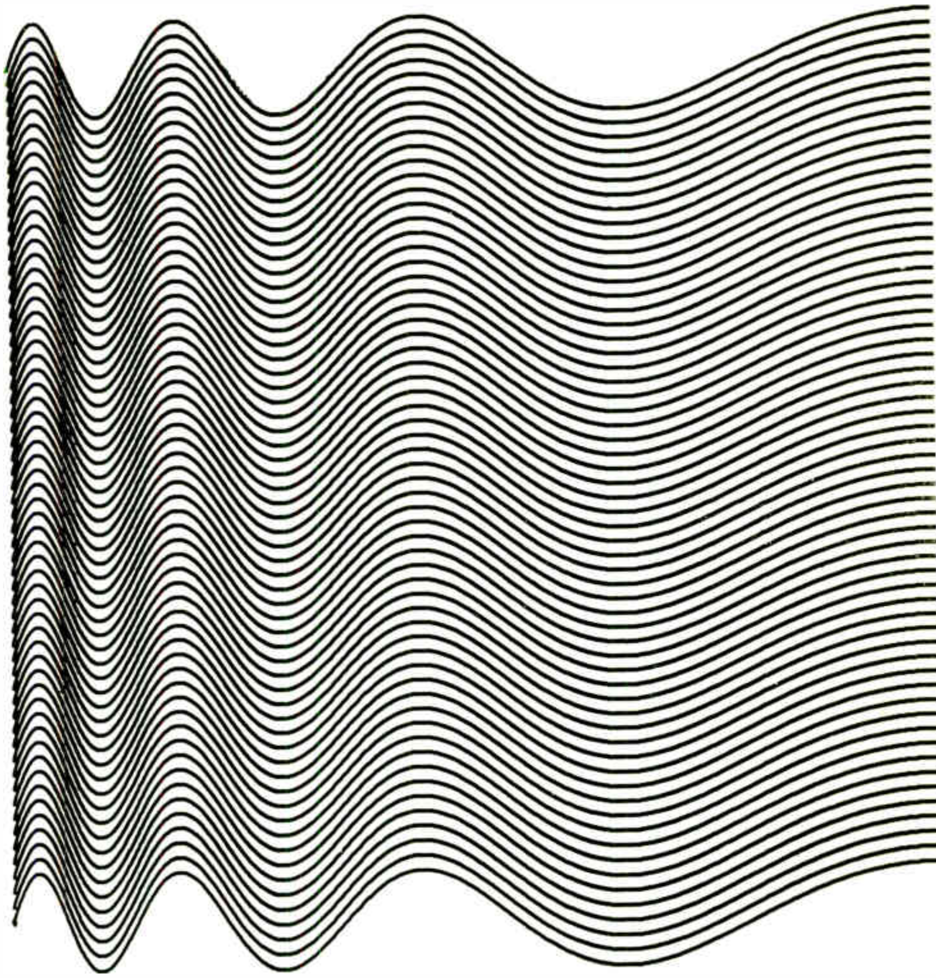
10. Rarich, T. D., "Development of SCM-1—a system for investigating the performance of a man–computer supervisory controlled manipulator," S.M. thesis, Massachusetts Institute of Technology, 1966.

11. Manalan, D. A., "A remote manipulator with industrial applications," S.B. thesis, Massachusetts Institute of Technology, 1965.

12. Barber, D. J., "MANTRAN: a symbolic language for supervisory control of a remote manipulator," S.M. thesis, Massachusetts Institute of Technology, 1967.

13. Whitney, D., Massachusetts Institute of Technology, doctoral dissertation in preparation.

14. Ford, L. R., "Network flow theory," Internal memo., Rand Corp., Aug. 1954, p. 923.



The digital computer as a creative medium

In the computer, man has created not just an inanimate tool but an intellectual and active creative partner that; when fully exploited, could be used to produce wholly new art forms and possibly new aesthetic experiences

A. Michael Noll *Bell Telephone Laboratories*

Digital computers are now being used to produce musical sounds and to generate artistic visual images. The artist or composer interacts directly with the computer through a console. This article explores the possibilities of the computer as an artistic medium and makes some predictions about the art of the future.

The notion of creating art works through the medium of machines may seem a little strange. Most people who have heard about the experimental use of digital computers in creative endeavors have probably shrugged them off as being of no consequence. On the one hand, creativity has universally been regarded as the personal and somewhat mysterious domain of man; and, on the one hand, as every engineer knows, the computer can only do what it has been programmed to do—which hardly anyone would be generous enough to call creative.

Nonetheless, artists have usually been responsive to experimenting with and even adopting certain concepts and devices resulting from new scientific and technological developments. Computers are no exception. Composers, film animators, and graphic artists have become interested in the application of computers in their creative endeavors. Moreover, recent artistic experiments with computers have produced results that should make us reexamine our preconceptions about creativity and machines. Some of the experiments, described in this article, suggest, in fact, that a tight interaction between artist and computer constitutes a totally new, active, and exciting artistic medium.

How does an artist work?

There is an anecdote attributed to Henri Matisse about how to approach the creative act of painting. You take a blank white canvas, the French artist said, and after gazing at it for a while, you paint on it a bright red disk. Thereafter, you do nothing further until something occurs to you that will be just as exciting as the original red disk. You proceed in this way, always sustaining, through each new gambit with the paint and brush, the initial high visual excitement of the red disk.

The anecdote is a somewhat simplified version of Matisse's idea, but even if we take it lightly, it can do a number of things for us. For one thing, it dispels some of the sense of mystery that hovers over the procedures of the creative person. It tells us something concrete and easily visualized about the creative process while emphasizing the role of the unexpected ideas for which the artist lies in wait and for which he sets a formal "trap" in his medium.

Even a relatively "passive" medium—paint, brushes, canvas—will suggest new ideas to the artist as he becomes engaged. The resistance of the canvas or its elastic give to the paint-loaded brush, the visual shock of real color and line, the smell of the paint, will all work on the artist's sensibilities. The running of the paint, or seemingly "random" strokes of the brush, may be accepted by him as corporate elements of the finished work. So it is that an artist explores, discovers, and masters the possibilities of the medium. His art work is a form of play, but it is serious play.

Most of all, the Matisse anecdote suggests that the artistic process involves some form of "program," one certainly more complex than the anecdote admits, but a definite program of step-by-step action. Without doing too much violence to our sense of what is appropriate, we might compare it to a computational hill-climbing technique in which the artist is trying to optimize or stabilize at a high level the parameter "excitement."

Once we have swallowed this metaphor, it becomes less improbable to imagine that computers might be used, in varying depths of engagement, as active partners in the artistic process. But computers are a *new* medium. They do not have the characteristics of paints, brushes, and canvas. Nor are the "statements" that grow out of the artist's engagement with them likely to be similar to the statements of, for example, oil paintings. An interesting question to explore, then, is how computers might be used as a creative medium. What kinds of artistic potentials can be evolved through the use of computers, which themselves are continually being evolved to possess more sophisticated and intelligent characteristics?

The character of the computer medium

In the present state of computer usage, artists are certainly having their problems in understanding engineering descriptions and in learning how to program computers in order to explore what might be done with them. However, they *are* learning, and they have already used digital computers and associated equipment to produce musical sounds and artistic visual images.[1-4]

The visual images are generated by an automatic plotter under the control of the digital computer. The plotter consists of a cathode-ray tube and a camera for photographing the images "drawn" on the tube face by deflections of the electron beam. The digital computer produces the instructions for operating the automatic plotter so that the picture-drawing capability is under program control. Musical sounds are produced by the computer by means of a digital sampled version of the sounds that must then be converted to analog form by a conventional digital-to-analog converter.

For both of these artistic applications, a challenging problem is the composition of special-purpose programming languages and subroutines so that the artist can communicate with the computer by using terminology reasonably similar to his particular art. For example, a special music compiler has been written so that the composer can specify complex algorithms for producing a single sound and then pyramid these basic sounds into a whole composition. A similar philosophy has been used in a special language developed for computer animation called Beflix.[5] Both applications share the drawback that the artist must wait a number of hours between the actual running of the computer program and the final generation of pictorial output or musical sounds when he can see or hear the results.

Since the scientific community currently is the biggest user of computers, most descriptions and ideas about the artistic possibilities for computers have been understandably written by scientists and engineers. This situation will undoubtedly change as computers become more accessible to artists who obviously are more qualified to explore and evolve the artistic potentials of the computer medium. Unfortunately, scientists and engineers are usually all too familiar with the inner working of computers, and this knowledge has a tendency to produce very conservative ideas about the possibilities for computers in the arts. Most certainly the computer is an electronic device capable of performing only those operations that it has been explicitly instructed to perform. And this usually leads to the portrayal of the computer as a powerful tool but one incapable of any true creativity. However, if creativity is restricted to mean the production of the unconventional or the unpredicted, then the computer should instead be portrayed as a creative medium—an active and creative collaborator with the artist.

Computers and creativity

Digital computers are constructed from a myriad of electronic components whose purpose is to switch minute electric currents nearly instantaneously. The innermost workings of the computer are controlled by a set of instructions called a program. Although computers must be explicitly instructed to perform each operation, higher-level programming languages enable pyramiding of programming statements that are later expanded into the basic computer instructions by special compiler pro-

This frame is selected from a movie produced by Stan Vanderbeek using a special animation programming language devised by Ken Knowlton. Each frame consists of a fine mosaic of dots that are combined to make desired shapes and forms. Intriguing and unusual "dissolves" and "stretches" that are easily done using the computer would be tedious if not impossible to execute by conventional hand animation techniques.

grams. These programming languages are usually designed so that the human user can write his computer program using words and symbols similar to those of his own particular field. All of this leads to the portrayal of the computer as a tool capable of performing tasks exactly as programmed.

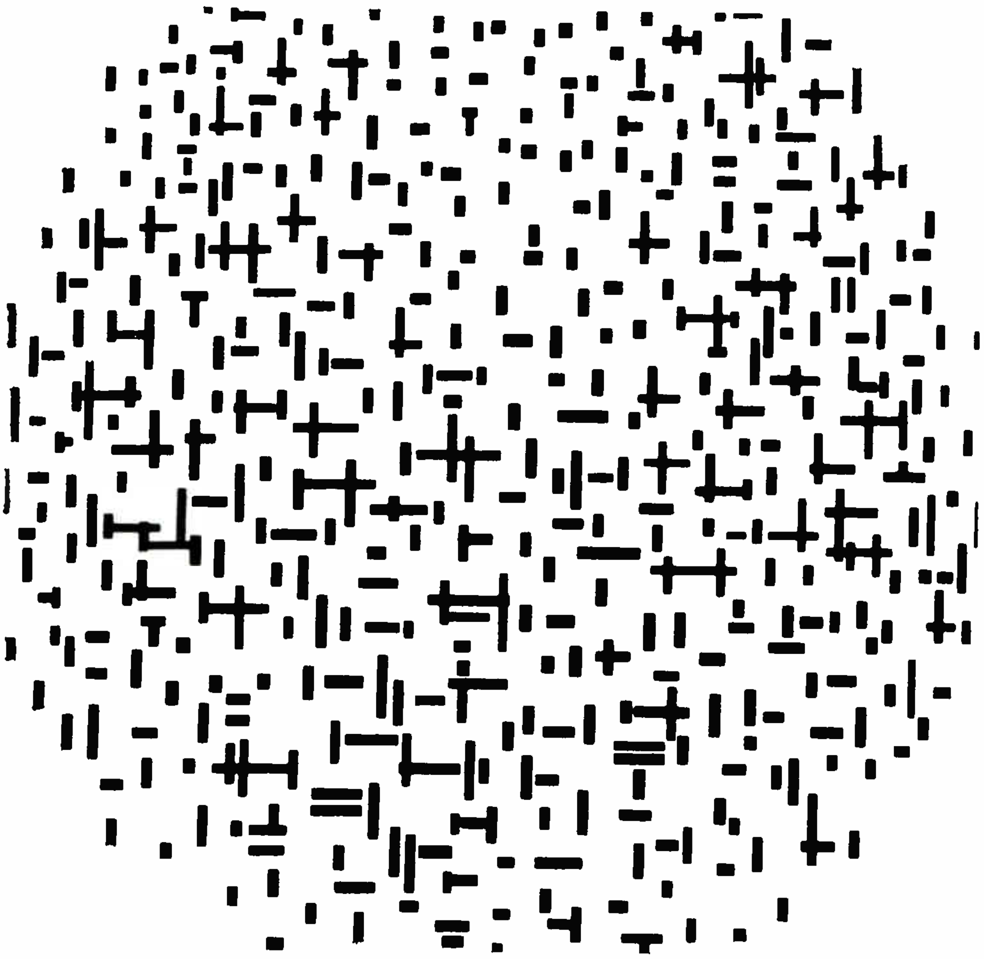
However, the computer is such an extremely powerful tool that artistic effects can sometimes be easily accomplished that would be virtually impossible by conventional artistic techniques. For example, by calculating and drawing on the automatic plotter the perspective projections from two slightly different directions of some three-dimensional object, the computer can generate three-dimensional movies of novel shapes and forms. Such three-dimensional animation, or kinetic sculpture, is far too tedious to perform by any other method. The computer's ability to handle small details has made possible intriguing dissolves and stretches, such as those executed by Stan Vanderbeek, without the tedium of conventional hand animation. Mathematical equations with certain specified variables under the control of the artist have also been used by John Whitney to achieve completely new animation effects. Much of "op art" uses repetitive patterns that usually can be expressed very simply in mathematical terms. The waveforms reproduced on page 89, which are like Bridget Riley's painting "Currents," were generated as parallel sinusoids with linearly increasing period. Thus, computer and automatic plotter can eliminate the tedious part of producing "op" effects.

Computers most certainly are only machines, but they are capable of performing millions of operations in a fraction of a second and with incredible accuracy. They can be programmed to weigh carefully, according to specified criteria, the results of different alternatives and act accordingly; thus, in a rudimentary sense, computers can appear to show intelligence.[6] They might assess the results of past actions and modify their programmed algorithms to improve previous results; computers potentially could be programmed to learn. And series of numbers can be calculated by the computer that are so complicatedly related that they appear to us as random.

Of course, everything the machine does must be programmed, but because of the computer's great speed, freedom from error, and vast abilities for assessment and subsequent modification of programs, it appears to us to act unpredictably and to produce the unexpected. In this sense, the computer actively takes over some of the artist's creative search. It suggests to him syntheses that he may or may not accept. It possesses at least some of the external attributes of creativity.

The Mondrian experiment

How reasonable is it to attribute even these rudimentary qualities of creativity to an inanimate machine? Is creativity something that should only be associated with the products of humans? Not long ago, in 1950, A. M. Turing expressed the belief that at the end of the century "one will be able to speak of machines thinking



The picture on the left, "Composition With Lines" (© Rykmuseum-Kröller-Müller), is a reproduction of a work by the Dutch painter Piet Mondrian. The picture on the right was generated by a digital computer using pseudorandom numbers with statistics approximating the Mondrian painting (© A. Michael Noll 1965). When xerographic reproductions of both pictures were shown to 100 subjects, the computer-generated picture was preferred by 59 of them. Only 28 subjects identified the Mondrian painting. Apparently, many of the observers associated randomness with human creativity and were therefore led astray in making the picture identifications.

without expecting to be contradicted.'"[7] Turing proposed the now well-known experiment consisting of an interrogator, a man, and a machine, in which the interrogator had to identify the man by asking the man and the machine to answer questions or to perform simple tasks.

A crude approximation to Turing's experiment was performed using Piet Mondrian's "Composition With Lines" (1917) and a computer-generated picture composed of pseudorandom elements but similar in overall composition to the Mondrian painting.[8] Although Mondrian apparently placed the vertical and horizontal bars in his painting in a careful and orderly manner, the bars in the computer-generated picture were placed according to a pseudorandom number generator with statistics chosen to approximate the bar density, lengths, and widths in the Mondrian painting. Xerographic copies of the two pictures were presented, side by side, to 100 subjects with educations ranging from high school to postdoctoral; the subjects represented a reasonably good sampling of the population at a large scientific research laboratory. They were asked which picture they preferred and also which picture of the pair they thought was produced by Mondrian. Fifty-nine percent of the subjects preferred the computer-generated picture; only 28 percent were able to identify correctly the picture produced by Mondrian.

In general, these people seemed to associate the randomness of the computer-generated picture with human creativity whereas the orderly bar placement of the Mondrian painting seemed to them machinelike. This finding does not, of course, detract from Mondrian's artistic abilities. His painting was, after all, the inspiration for the algorithms used to produce the computer-generated picture, and since computers were nonexistent 50 years ago, Mondrian could not have had a computer at his disposal. Furthermore, we must admit that the reduction in size of the original painting and its xerographic reproduction degrades its unique aesthetic qualities. Nevertheless, the results of the experiment in light of Turing's proposed experiment do raise questions on the meaning of creativity and the role of randomness in artistic creation. In a sense, the computer with its program could be considered creative, although it can be argued that human creativity was involved in the original

program with the computer performing only as an obedient tool.

These questions should perhaps be examined more deeply by more ambitious psychological experiments using computer-generated pictures as stimuli.

Toward real-time interaction

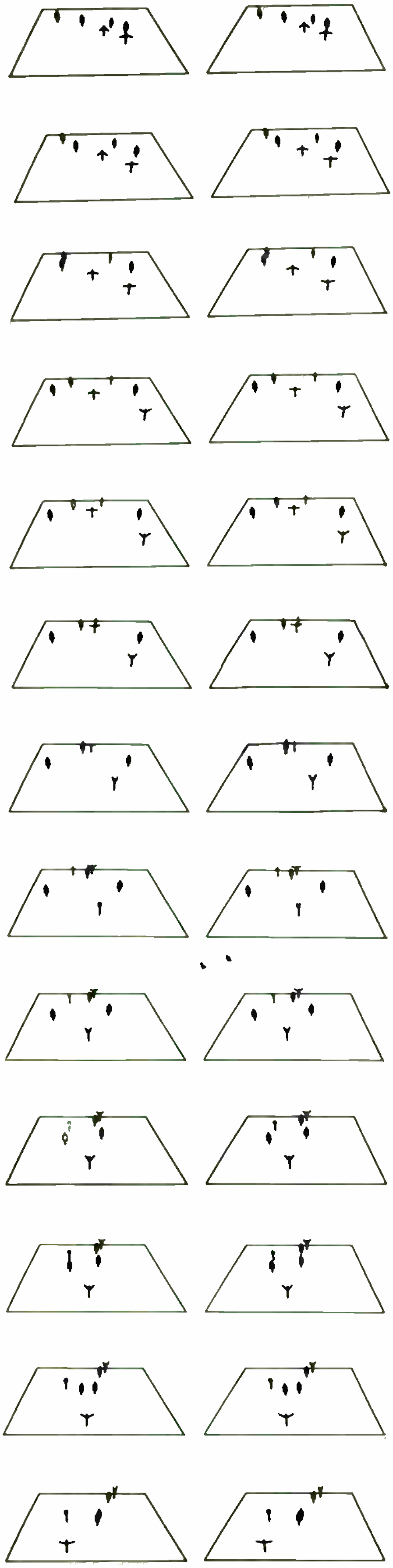
Although the experiments described show that the computer has creative potentialities beyond those of just a simple tool, the computer medium is still restrictive in that there is a rather long time delay between the running of the computer program and the production of the final graphical or acoustic output. However, recent technological developments have greatly reduced this time delay through special interactive hardware facilities and programming languages. This tightening of the man–machine feedback loop is particularly important for the artist who needs a nearly instantaneous response.

For example, in the field of music an electronic graphic console has been used to specify pictorially sequences of sounds that were then synthesized by the computer.[9] Functions for amplitude, frequency, and duration of a sequence of notes were drawn on the face of a cathode-ray tube with a light pen. If desired, the computer combined specified functions according to transparently simple algorithms. Thus, the fine details of the composition were calculated by the computer and the overall structure was precisely specified by the graphical score. The feedback loop was completed by the computer-generated sounds heard almost immediately by the composer, who could then make any desired changes in the score.

A similar man–machine interactive system has been proposed for choreography.[10] In this system, the choreographer would be shown a computer-generated three-dimensional display of complicated stick figures moving about on a stage, as shown on page 94. The choreographer interacts with the computer by indicating the spatial trajectories and movements of the figures. Random and mathematical algorithms might be introduced by the computer to fill in certain fine details, or even to give the choreographer new ideas to evaluate and explore.

A new active medium

The beginnings of a new creative partnership and collaboration between the artist and the computer clearly emerge from these most recent efforts and proposals. Their common denominator is the close man–machine interaction using the computer to generate either musical sounds or visual displays. The computer acquires a creative role by introducing randomness or by using mathematical algorithms to control certain aspects of the artistic creation. The overall control and direction of the creative process is very definitely the artist's task. Thus, the computer is used as a medium by the artist, but the great technical powers and creative potentialities of the computer result in a totally new kind of creative medium. This is an *active* medium with which the artist can interact on a new level, freed from many of the physical limitations of all other previous media. The artistic potentialities of such a creative medium as a collaborator with an artist are truly exciting and challenging.



The choreographer of the future might sit at the console of a computer and see a display of human figures stylized by simple stick figures as shown in these frames from a computer-generated movie. By interacting with the computer, the choreographer might create his dance composition, perhaps leaving certain movements to be suggested by pseudorandom and mathematical algorithms within the computer.

Interactive aesthetic experiences

In the previous examples the artist sat at the console of the computer and indicated his desires to the computer by manually using push buttons or by drawing patterns on an electronic visual display. These are probably efficient ways of communicating certain types of instructions to the computer; however, the communication of the actual subconscious emotional state of the artist could lead to a new aesthetic experience. Although this might seem somewhat exotic and conjectural, the artist's emotional state might conceivably be determined by computer processing of physical and electrical signals from the artist (for example, pulse rate and electrical activity of the brain). Then, by changing the artist's environment through such external stimuli as sound, color, and visual patterns, the computer would seek to optimize the aesthetic effect of all these stimuli upon the artist according to some specified criterion.

This interactive feedback situation with controlled environment would be completely dynamic. The emotional reaction of the artist would continually change, and the computer would react accordingly either to stabilize the artist's emotional state or to steer it through some preprogrammed course. Here then is a completely new aesthetic experience utilizing man–machine communication on the highest (or lowest, if you will) subconscious levels and computer processing and optimization of emotional responses. Only a digital computer could perform all the information processing and generate the sights and sounds of the controlled environment required for such a scheme. One is strongly tempted to describe these ideas as a consciousness-expanding experience in association with a psychedelic computer!

Although such an artistic feedback scheme is still far in the future, current technological and psychological investigations would seem to aim in such a direction. For example, three-dimensional computer-generated color displays that seem to surround the individual are certainly already within the state of the art. Electroencephalograms are being scrutinized and studied in great detail, using advanced signal analysis techniques; it is not inconceivable that some day their relation to emotional state might be determined.

Artistic consequences

Predictions of the future are risky in that they may be really nothing more than what the person predicting would like to see occur. Although the particulars should be viewed skeptically, they actually might be unimportant; if the art of the future follows the directions outlined here, then some general conclusions and statements can be made that should be independent of the actual particulars.

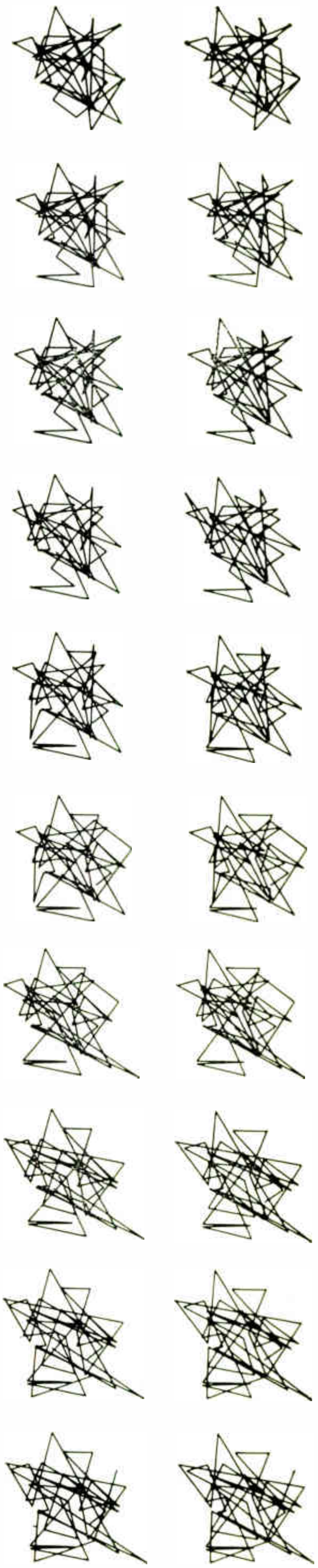
The aesthetic experience will be highly individualistic, involving only the individual artist and his interactions with the computer. This type of participation in the creative and aesthetic experience can be experienced by artist and nonartist alike. Because of the great technical and creative power of the computer, both the artist and nonartist are freed from the necessity of strong technical competence in the use of different media. The artist's "ideas" and not his technical ability in manipulating media could be the important factor in determining artistic merit. Conceivably, a form of "citizen-artist" could emerge, as envisioned by Allon Schoener.[11] The interactive aesthetic experience with computers might fill a substantial portion of that great leisure time predicted for the man of the future.

The artist's role as master creator will remain, however, because even though the physical limitations of the medium will be different from traditional media, his training, devotion, and visualization will give him a higher degree of control of the artistic experience. As an example, the artist's particular interactions with the computer might be recorded and played back by the public on their own computers. Specified amounts of interaction and modification might be introduced by the individual, but the overall course of the interactive experience would still follow the artist's model. In this way, and for the first time, the artist would be able to specify and control with certainty the emotional state of each individual participant. Only those aspects deliberately specified by the artist might be left to chance or to the whims of the participant. All this would be possible because the computer could monitor the participant's emotional state and change it according to the artist's specifications. The artist's interaction with the computer would be of a new order because the physical restrictions of the older media would be eliminated.

This is not to say that the traditional artistic media will be swept away; but they will undoubtedly be influenced by this new active medium. The introduction of photography—the new medium of the last century—helped to drive painting away from representation, but it did not drive out painting. What the new creative computer medium will do to all of the art forms—painting, writing, dance, music, movies—should be exciting to observe. We might even be tempted to say that the current developments and devices in the field of man-machine communication, which were primarily intended to give insight into scientific problems, might in the end prove to be far more fruitful, or at least equally fruitful, in the arts.

REFERENCES

1. Mathews, M. V., "The digital computer as a musical instrument," *Sci.*, vol. 142, no. 3592, pp. 553–557, Nov. 1, 1963.
2. Rockman, A., and Mezei, L., "The electronic computer as an artist," *Canadian Art*, vol. XXI, pp. 365–367, Nov./Dec. 1964.
3. Noll, A. M., "Computers and the visual arts," *Design Quarterly*, no. 66/67, pp. 65–71, 1967.
4. Zajac, E. E., "Computer animation: a new scientific and educational tool," *J. Soc. Motion Picture and Television Eng.*, vol. 74, pp. 1006–1008, Nov. 1965.
5. Knowlton, K. C., "A computer technique for producing animated movies," *Am. Fed. Infor. Proc. Soc. (AFIPS) Conf. Proc.*, vol. 25, pp. 67–87, 1964.
6. Minsky, M. L., "Artificial intelligence," *Sci. Am.*, vol. 215, pp. 246–260, Sept. 1966.
7. Turing, A. M., "Computing machinery and intelligence," *Mind. A quarterly review of psychology and philosophy*, vol. 59, N.S., pp. 433–460, Oct. 1950.
8. Noll, A. M., "Human or machine: a subjective comparison of Piet Mondrian's 'Composition with Lines' (1917) and a computer-generated picture," *The Psychological Rec.*, vol. 16, pp. 1–10, Jan. 1966.
9. Mathews, M. V., "A graphical language for composing and playing sounds and music," presented at the 31st Conv. of the Audio Engineering Society, Oct. 1966 (preprint no. 477).
10. Noll, A. M., "Choreography and computers," *Dance Magazine*, vol. XXXXI, pp. 43–45, Jan. 1967.
11. Schoener, Allon, "2066 and all that," *Art in America*, vol. 54, pp. 40–43, Mar.–Apr. 1966.



These frames are the left and right images of a 3D pair from a computer-generated three-dimensional movie. The object randomly changes its shape in what might be called a form of kinetic sculpture. To view the 3D effect, place a sheet of paper on edge between a stereo pair. Position your head so that each eye sees only one image. With a bit of adjustment, the images should seem to converge and appear three-dimensional.

The scanning electron microscope

Though not a new device, the scanning electron microscope is only beginning to come into its own as a useful and practical scientific instrument, particularly for the direct examination of various kinds of surfaces

R. F. W. Pease *University of California, Berkeley*

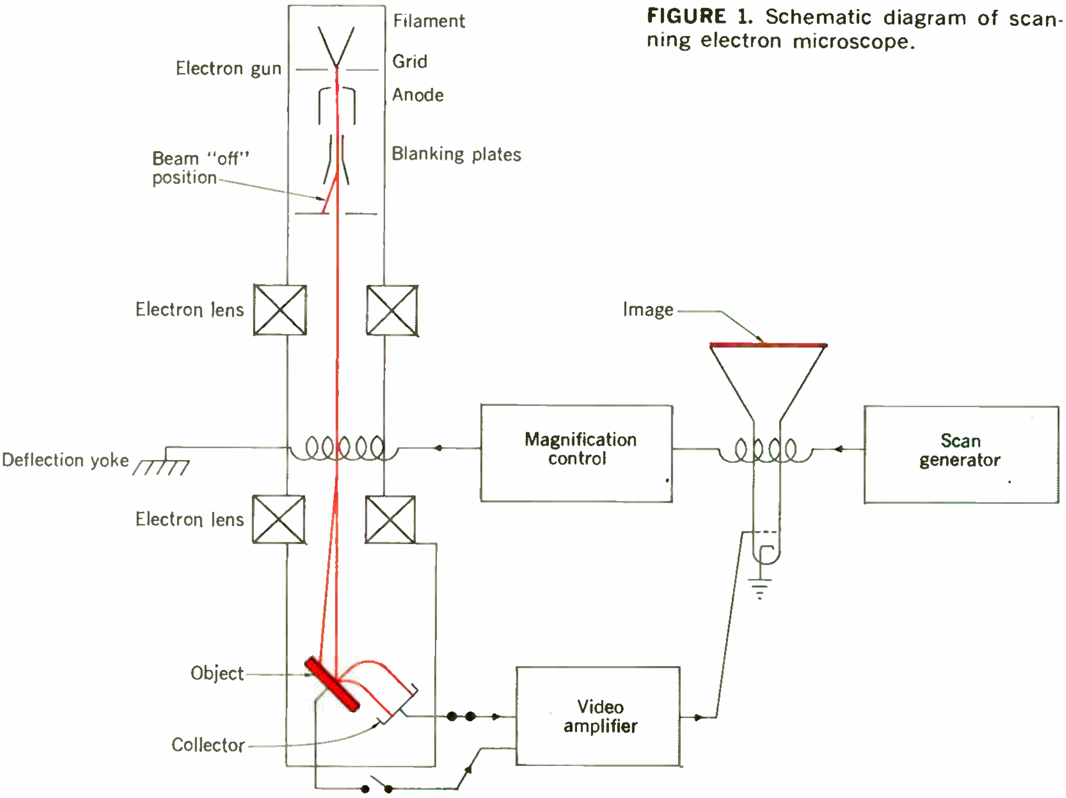


FIGURE 1. Schematic diagram of scanning electron microscope.

The scanning electron microscope combines the techniques of the cathode-ray tube and the conventional electron microscope—both considered indispensable to modern technology. The SEM, which presents a picture having a distinct three-dimensional appearance, is finding application in the examination of wood fibers in connection with paper manufacture, of surfaces undergoing ionic bombardment, and of corrosion. At the present time, one of the most pressing problems is to reduce its total cost.

The microscope and the cathode-ray oscilloscope are probably two of the most productive instruments in science. It is strange therefore that the scanning electron microscope (SEM), which employs the techniques of both these instruments, should still be relatively unknown. Historically, the SEM is as old as the conventional electron microscope[1,2]; but since the main impetus to electron microscope development was to improve resolution per se, the scanning instrument suffered a loss of interest because its greater complexity offered no improvement in resolution.

Recently, the advantages of the SEM have become significant in a number of different applications; most of these applications have involved the direct examination of surfaces.

Operation of the instrument

The principle of the SEM is shown in Fig. 1. The electron optical column demagnifies the electron source to focus a fine beam of electrons on the specimen surface. This beam is scanned across the specimen surface in a rectangular raster in synchronism with the spot of a cathode-ray tube. The signal resulting from the interaction of the beam with the specimen modulates the CRT brightness, and therefore a picture of the specimen is built up on the cathode-ray-tube face. In many applications the secondary electrons are used to build up the picture. Since the secondary-electron yield depends on the angle between the specimen surface and the electron beam, the picture obtained represents surface topography.[3–5] If the signal used to build up the picture is derived from the characteristic X-ray production, then the picture obtained represents elemental distribution and the instrument is known as a scanning X-ray microanalyser (in England) or an electron microprobe (in the United States). This particular mode of operation has been described extensively in the literature and will not be discussed here.

The magnification is simply the ratio of the scan amplitude on the CRT face to that on the specimen, and hence can be controlled by an attenuator between the two sets of scan coils. No adjustment of the electron lenses is necessary when changing magnification.

The resolving power can be defined as the diameter of the electron beam at the specimen; if the current density of the beam is Gaussian with respect to radius, then d is generally taken to be that diameter within which 80 percent of the current is contained.[4] The resolution actually observed on a micrograph, however, will also be affected by the penetration of the primary electron into the specimen.

Several factors set a lower limit to the value of d; there are also limitations to the maximum current that can be focused into a beam of given diameter. These limits have been described in previous publications[3,4,6] and only an outline is presented here. If the beam emerging from the final lens has a semiangle of convergence α, then spherical and chromatic aberrations will give rise to two disks of confusion whose respective least diameters are given by

$$d_s = \tfrac{1}{2} C_s \alpha^3 \qquad d_c = \frac{\Delta V}{V} C_c \alpha$$

where C_s and C_c are the coefficients of spherical and chromatic aberration of the final lens, and $\Delta V/V$ is the fractional spread in the energy of electrons passing through the final lens.

Diffraction of the electrons will give rise to an Airy disk having a diameter d_f in angstroms, given by

$$d_f = \frac{1.5}{\alpha} V^{-1/2}$$

where V is the electron energy in electronvolts.

The minimum value of d can be found by combining these disks. But if the beam is to have a given current i, then a further term, d_0, should be added. This term can be derived from the Abbe sine condition[7] or the Liouville theorem[8] in the following form:

$$d_0^2 = \frac{6i}{\pi J_c} \frac{kT}{eV} \cdot \alpha^{-2}$$

where d_0 is the diameter of the Gaussian image of the electron source (that is, the diameter obtained simply from the source diameter and the focal lengths and positions of the lenses), J_c is the emission current density at the cathode e in the electronic charge, k is Boltzmann's constant, and T is the absolute temperature of the cathode. The value of the beam diameter d is usually estimated by adding in quadrature the component diameters to obtain an expression of the form

$$d^2 = P\alpha^{-2} + Q\alpha^2 + R\alpha^6$$

where

$$P = \frac{6kT}{\pi e V} \frac{i}{J_c} + (1.22\lambda)^2$$

$$Q = \left(C_c \frac{\Delta V}{V} \right)^2 \qquad R = \left(\frac{C_s}{2} \right)^2$$

If i is negligible, then the expression for d reduces to that for the resolution of a conventional electron microscope (less than 5 Å); however, d is generally larger than 5 Å for a variety of reasons. One reason is that in order to collect secondary electrons (of about 4 eV energy) the specimen must be in a region of low magnetic field (about 240 amperes per meter); this places restrictions on the design of the final lens that result in larger aberrations than in the conventional electron microscope.[9] Another reason is that V is generally lower than in a conventional electron microscope in order for penetration to be minimized. The result is increased chromatic aberration and a larger diffraction effect.

The minimum useful value of i is set by the desired picture quality, the contrast present in the video signal, the time available to record a micrograph (because of power supply instability or inconvenience), and the efficiency of video-signal production. When secondary-electron collection is used, the secondary current is of the same order as the primary current and the detection using a scintillator photomultiplier system is very efficient[10]; in this mode a primary electron current of at least 10^{-12} ampere (6×10^6 electrons per second) is required to build up a picture of "reasonable quality" (2.5×10^5 picture points, 10 distinct density levels) in a "reasonable time" (about 3 minutes).[3]

A microscope with C_s of 1 cm and C_c of 0.8 cm should be able to focus a current of 10^{-12} ampere into a spot of diameter 50 Å for $V = 15$ kV and of diameter 30 Å for $V = 30$ kV. Experimental results on such a microscope gave the respective diameters as 75 Å and 50 Å; however, the best point-to-point resolution observed with this microscope was about 100 Å, and for most specimens the resolution was about 200 Å.[6,9] The reason for these disappointing values might be explained by the effect of penetration of the primary beam into the specimen. The secondary electrons are of low energy (about 4 eV) and very low depth of origin (<50 Å),[11] and can be considered to consist of two classes: those emitted by the action of

FIGURE 2. Pure iron after oxidation in air at 630°C for five minutes.

the beam entering the specimen and those emitted by the action of the backscattered beam emerging from the specimen. Electrons in the first class would emerge from an area little larger than the primary beam and hence would always contribute to the useful signal. Electrons in the second class comprise, for metallic specimens, more than half the total signal,[11,12] yet they emerge over an area much larger than the primary beam[13] and hence should be considered as noise at high magnification. Indeed, scanning electron micrographs taken of the same specimen appear increasingly noisy as the magnification is increased; at the highest magnifications (about 100 000) focusing is extremely difficult because of the noise.[6]

Applications

The applications of the SEM in the secondary electron mode usually involve the examination of surfaces that cannot be replicated.

Obviously, very rough surfaces are best examined with the SEM. Considerable work has been done on wood fibers in connection with the manufacture of paper.[14] Metallurgical applications have included the examination of rough fracture surfaces[15] and a study of corrosion.[16] Figure 2 shows the surface of iron undergoing oxidation in the SEM. Replication of this surface would destroy the surface for meaningful subsequent examination. Other similar applications include the examination of cathodes undergoing activation,[17] of crystals decomposing,[18] and of surfaces undergoing ionic bombardment.[19–21]

One application that is receiving much attention is the examination of semiconductor devices and integrated circuits.[22–24] In the secondary-electron mode it is possible to distinguish between regions of different voltage. The trajectory of a secondary electron can be made to depend on the potential of the point of origin; hence, by selecting only those electrons falling on a certain area, a variation in signal level with surface potential can be obtained.[5] Voltage differences of 0.25 volt can be detected in this way.

Figure 3(A) shows a transistor in the secondary mode. The steps in the surface level at the edges of the different regions are clearly visible. The longer-range contrast is due to the different voltages of the different regions. The collector regions and tub region are at −2 volts, the emitter region is grounded, and the base region is slightly positive. When examining a micrograph in the voltage-contrast mode it is often convenient to have a comparison micrograph of the same region taken in a mode that does not show voltage contrast (for example, backscattered electron collection).

A more sensitive method for detecting applied voltages has recently been described. Here, the voltage is applied at a known frequency and the video amplifier is tuned to this frequency; thus, only the voltage-dependent signal is amplified and the signal resulting from topography is rejected. In this way, voltage differences as small as 5 mV across p-n junctions can be detected.[25]

In many applications involving the examination of semiconductors, the video signal is derived from currents induced across p-n junctions caused by electron bombardment, which creates free carriers in the junction region.

Figure 3(B) shows the same transistor as in Fig. 3(A); in this micrograph the video signal is derived from the collector current that results when the electron beam strikes the junction regions. The reference gray level is shown on the left of the picture; a positive signal (bright) results when the collector–n tub region is bombarded and a negative (dark) signal results when the emitter–base region is bombarded. The aluminum leads and variations in surface topography also show up because they affect the total energy of the beam that penetrates to the junctions.

In many cases, a more satisfactory way of displaying these variations is to differentiate the induced currents; this can sometimes be accomplished simply by not grounding the specimen. Not only is the surface detail more prominent than when the induced current is used as a measure of brightness, but the picture has a more three-dimensional appearance, which many people find more informative; see Fig. 3(C).

Internal structure can be revealed in a specimen having a shallow junction and using as a video signal the junction current caused by the diffusion of excess carriers (induced by the electron beam) from the surface to the junction. In this manner, spatial variations in carrier lifetime give rise to contrast; hence, dislocation networks and other defects can be revealed.[26]

A further variation involves displaying the video signal as a *y* perturbation of the display CRT spot. Since the eye is more discriminating with respect to position than to brightness, such a picture is more informative in terms of signal magnitudes. This mode is therefore particularly useful, in any application, when the contrast is low.[27, 28]

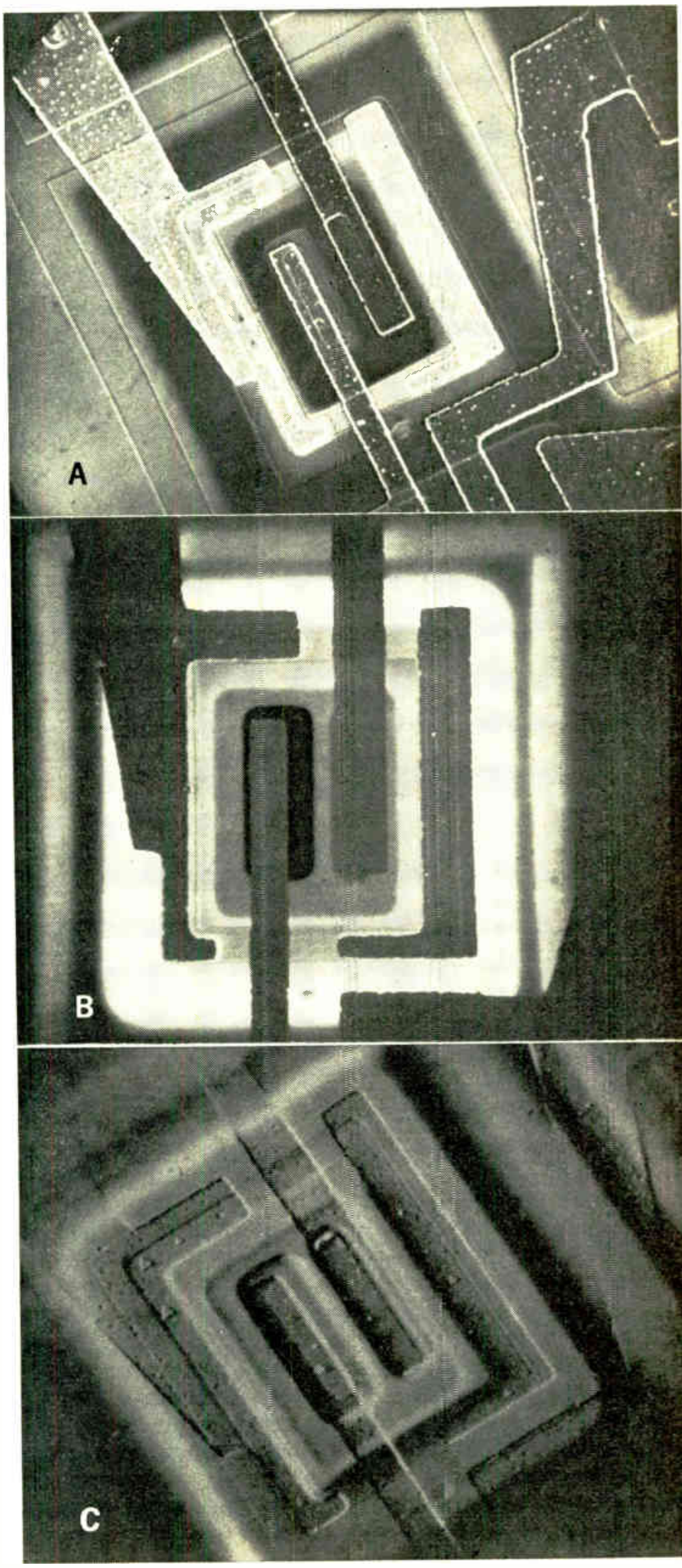
Another type of video signal that is used to build up the scanning picture is derived from the light emitted as a result of electron bombardment of the specimen.

Infrared light resulting from recombination in GaAs has been used to study defects and p-n junctions in these materials (Fig. 4).[29-33] Some work using this mode (in the visible-wavelength range) to study phosphors has also been reported.[34,35] Certain biologic tissues have been examined in this way; in this application the regions to be examined can often be stained with cathodoluminescent material to reveal the regions of interest.[36]

Effect of the beam on the specimen

In conventional electron microscopy the beam generally has a deleterious effect on the specimen. This can be attributed to heating, to electrostatic charging, or to chemical rearrangement, of which one example is contamination. This last effect is the formation of a film over the regions bombarded by the electron beam as a result of polymerization of residual hydrocarbons in the vacuum. In the SEM the beam current is many orders of magnitude lower than in the conventional electron

FIGURE 3. Planar p-n-p transistor as seen in the SEM. The field of view is 200 by 280 μm. A—Secondary electrons are used to build up the picture. The surface topography and variations in electric potential are visible. Collector and connecting electrode are at −2 volts and show up brightly. The emitter is grounded and the base is positive and appears dark. B—The collector current is the video signal and the junction regions are now clearly visible. C—The video signal is now the time derivative of the collector current (the spot is scanned from left to right) and the transistor takes on a three-dimensional appearance, where the height represents collector current.



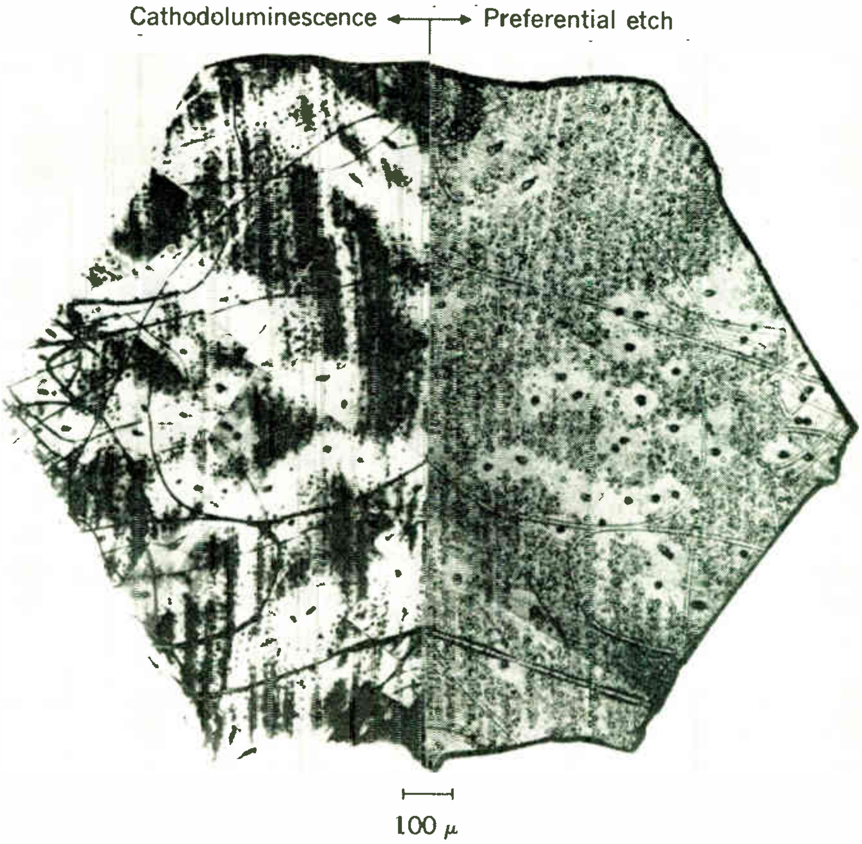


FIGURE 4. Gallium arsenide crystal. On the left is a mosaic of a number of scanning electron micrographs obtained using infrared recombination radiation as the video signal. On the right is a light micrograph of the same crystal etched to reveal dislocation. By comparing the pictures it can be inferred that the contrast mechanisms are largely similar.

microscope (CEM); therefore, heating is usually negligible and contamination is reduced. Electrostatic charging can present a problem in the viewing of thick insulators, but through operation at low primary beam energy (about 1 keV) the specimen surface will stabilize to that of its surroundings.[37]

In some applications the effects of the beam are put to positive advantage. These applications are not microscopy in the strict sense, but since much of the experimental work has been carried out on scanning electron microscopes, a very brief description is included. The heating effect is used for thermal micromachining, welding, or zone melting when the beam current is raised to about 10 μA or above.[10] Electrostatic charging is used for information storage; the use of thermoplastic film is perhaps the best-known target material in this respect.[38] However, a new type of electron beam charge storage system[39] has recently been described in which the charge induced in the oxide of a metal-oxide semiconductor transistor affects the threshold voltage of the device when the charged region extends across the channel width. No change in threshold voltage occurs when the charged region extends over only part of the channel width. Information is stored by charging certain regions extending halfway across the channel. To read the information, the beam bombards regions extending across the other half of the channel. With the correct potential applied across the channel, drain-current pulses will then result only when the beam bombards regions adjacent to previously bombarded areas. The oxide film can be selectively discharged by electron bombardment with a negative gate-electrode potential. The write sensitivity is about 10^{-6} C/cm^2 and the read sensitivity about 10^{-7} C/cm^2; the read process can be either destructive or nondestructive. The packing density depends on integrated-circuit techniques, but a value about 3×10^6 bits/cm^2 appears quite feasible. In this context, electron beams themselves are being used to expose photoresist and form masks for subsequent etching.[40] Transistors with connecting electrodes less than one micrometer wide have been made in this way.[41]

The contaminating effect of the beam can also be used as a high-resolution resist to form a mask for subsequent etching.[42] Figure 5(A) shows a surface of a 2500 Å thick Au film overlaying a Si substrate. A grid of lines of contamination has been deposited by the scanning electron beam. If the etching process is dry, it can be monitored with the SEM [Fig. 5(B)] until etching is complete [Fig. 5(C)]. Ridges of Au on Si, 700 Å wide and 2500 Å high, are formed in this way [Fig. 5(C)]. Where there is no substrate, structures as small as 100 Å wide can be made.[43-45]

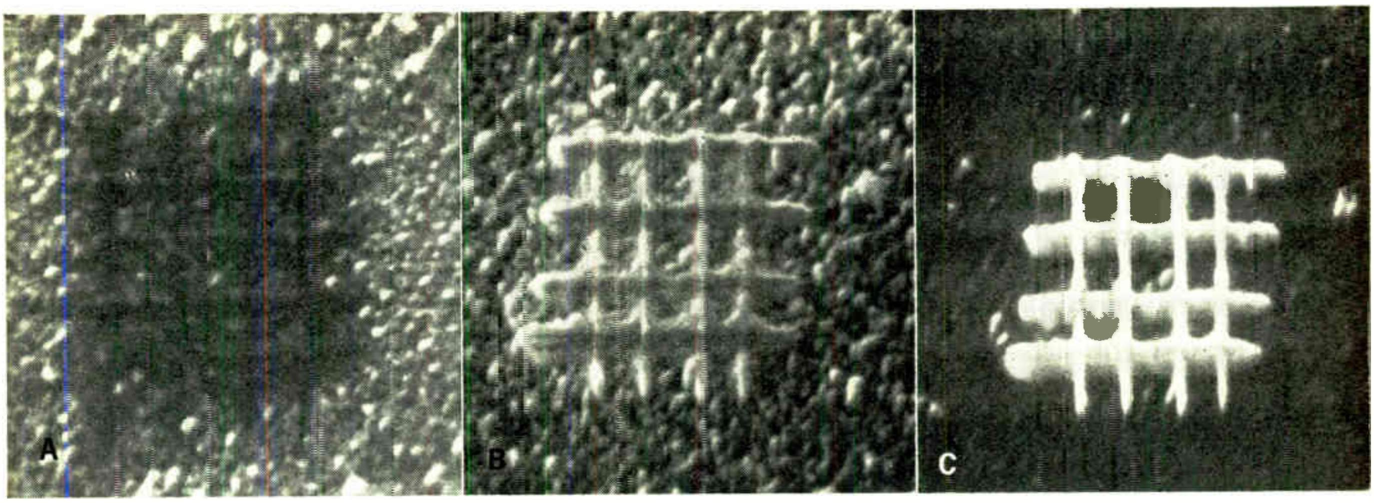


FIGURE 5. A series of micrographs taken with a scanning electron microscope showing micromachining with a 5-keV argon-ion beam. The field of view is 3.6 by 5.2 μm. A—Here the machining has only just commenced and the dark grid lines of the electron-beam-induced polymer can be seen. B—The protective action of the machining has caused the appearance of ridges underneath the grid lines. C—The ion beam has etched away the gold film to expose the (dark) surface, except under the grid lines. The difference in widths of the vertical and horizontal ridges is due to the difference between the viewing and etching directions.

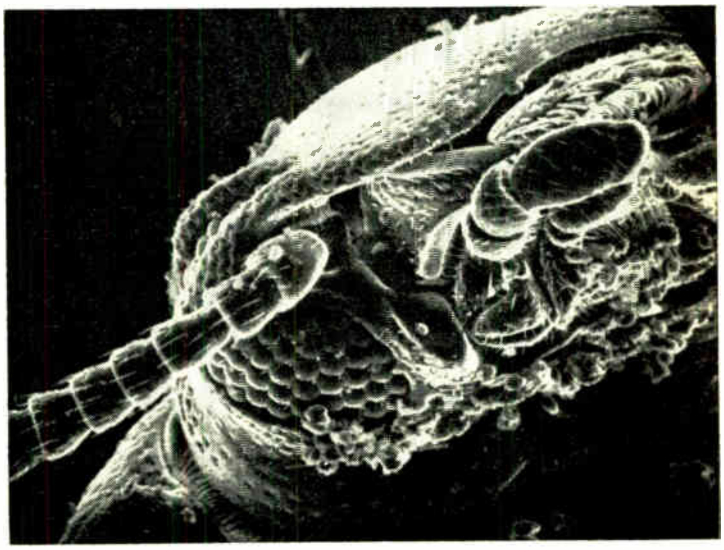


FIGURE 6. A low-magnification scanning electron micrograph of the head of a living Tribolium beetle. The field of view is 1.1 by 0.8 mm.

One very recent development exploits the low beam power of the SEM to view certain living creatures—for example, the Tribolium beetle, shown in Fig. 6. All developmental stages of the Tribolium beetle appear to be resistant to the vacuum and also to the low level of irradiation.[46] Not only can living specimens be examined at much higher resolution than is possible with the light microscope but very small selected areas (less than a micrometer in diameter) can be preferentially irradiated with the electron beam to investigate the effect of such radiation on the subsequent development of the specimen and its offspring.

Possible future developments

There seems to be no obvious way of improving the resolution of the SEM in the reflection mode of operation because of the limit placed by penetration. Significant improvements in picture quality may well be achieved if the field-emission cathode can be developed as a reliable source of electrons. Preliminary studies suggest that it is possible to focus much more current into a probe of 200 Å when using a cathode employing field emission rather than thermionic emission.[47] Not only could pictures be recorded in a shorter time, but this increased current should enable greater contrast expansion to be used to reveal detail on specimens of low contrast.

One recent development is the scanning transmission electron microscope.[48] At first sight it may seem curious to use the rather cumbersome scanning technique when the CEM is both cheaper and quicker. The advantage of the scanning technique lies in the wider choice of contrast mechanisms. The CEM relies chiefly on the absorption and scattering of the electrons to show contrast. The SEM can use any measurable interaction as a contrast mechanism. One such mechanism is to use the characteristic energy loss suffered by electrons on transmission through certain elemental films; in this way the local elemental content could be revealed. The resolution of this technique is not limited by the electron penetration and could be as good as in a CEM. The energy resolution could be as good as 6 mV in 30 kV,[49] which is better than the 0.5 volt resolution actually observed in a CEM modified to image electrons in certain energy ranges.[50]

On the commercial side, the SEM costs considerably more than the CEM and one development must be to reduce this cost. An alternative would be to offer an electron microscope that can be used either as a CEM or SEM. Such a microscope (which is not available commercially) has been in operation for some years at the Pulp and Paper Research Institute at Point Claire, Quebec.[51] As far as the writer is aware, however, this very logical development has not yet been applied to a modern high-performance electron microscope.

The author is indebted to Dr. O. C. Wells for the use of his bibliography on scanning microscopy.[52] Figure 3 is reproduced with the permission of C. J. Varker and his coauthors and R. Bakish. Figure 4 is reproduced with permission of H. G. Casey, Jr., and the editor of the *Journal of the Electrochemical Society*. Figure 5 is reproduced with the permission of Dr. A. N. Broers and the editor of *Microelectronics and Reliability*. Figure 6 is reproduced with the permission of the editor of *Science*. The author also acknowledges the support in part of the Air Force Avionics Laboratory and the Joint Services Electronics Program.

REFERENCES

1. Knoll, M., "Static potential and secondary emission of bodies under electron irradiation," *Z. Tech. Phys.*, vol. 11, pp. 467–475, 1935.

2. Zworykin, V. K., Hillier, J., and Snyder, R. L., "A scanning electron microscope," *ASTM Bull.*, no. 117, pp. 15–23, 1942.

3. Oatley, C. W., Nixon, W. C., and Pease, R. F. W., "Scanning electron microscopy," *Advances in Electronics and Electron Physics*, L. Marton, ed. New York: Academic Press, 1965, vol. 247, pp. 181–247.

4. Smith, K. C. A., and Oatley, C. W., "The scanning electron microscope and its fields of application," *Brit. J. Appl. Phys.*, vol. 6, pp. 391–399, 1955.

5. Everhart, T. E., Wells, O. C., and Oatley, C. W., "Factors affecting contrast and resolution in the scanning electron microscope," *J. Electron. Control*, vol. 7, pp. 97–111, 1959.

6. Pease, R. F. W., "High resolution scanning electron microscopy," Ph.D. dissertation, Cambridge University, 1963.

7. Langmuir, D. B., "Theoretical limitations of cathode-ray tubes," *Proc. IRE*, vol. 25, pp. 977–991, Aug. 1937.

8. Pierce, J. R., *Theory and Design of Electron Beams.* Princeton, N.J.: Van Nostrand, 1953, p. 118 *et seq.*

9. Pease, R. F. W., and Nixon, W. C., "High resolution scanning electron microscopy," *J. Sci. Instr.*, vol. 42, pp. 81–85, 1965.

10. Steigerwald, K. H., "Electron beam milling," *Proc. 3rd Symp. Elec. Tech.*, pp. 262–290, 1961.

11. Bronshtein, I. W., and Fraiman, B. S., "The secondary emission properties of metals and semiconductors," *Sov. Phys.-Solid-State*, vol. 3, pp. 2337–2339, 1962.

12. Kanter, H., "Contributions of backscattered electrons to secondary electron emission," *Phys. Rev.*, vol. 121, pp. 681–684, 1961.

13. Pease, R. F. W., "The spatial distribution of secondary electrons," ERL Tech. Memo. M-136, University of California, Berkeley, 1965.

14. Smith, K. C. A., "Scanning electron microscopy in pulp and paper research," *Pulp Paper Mag. Can.*, vol. 60, pp. 1366–1371, 1959.

15. Tipper, C. F., Dagg, D. I., and Wells, O. C., "Surface fracture markings on alpha iron crystals," *J. Iron Steel Inst.*, vol. 193, pp. 133–141, 1959.

16. Pease, R. F. W., and Ploc, R. A., "Scanning electron microscopy as applied to the oxidation of ion," *Trans. Met. Soc. AIME*, vol. 233, pp. 1949–1954, 1965.

17. Beck, A. H. W., and Ahmed, H., "The activation process in dispenser cathodes," *J. Electron. Control*, vol. 14, pp. 623–627, 1963.

18. McAuslan, J. H. L., and Smith, K. C. A., "The direct observation in the scanning microscope of chemical reactions," *Proc. European Regional Conf. Electron Microscopy*, Stockholm, 1956, pp. 343–345.

19. Stewart, A. D. G., and Boyde, A., "Ion etching of dental tissue in a scanning electron microscope," *Nature*, vol. 196, pp. 81–82, 1962.

20. Broers, A. N., "Selective ion beam etching in the scanning electron microscope," Ph.D. dissertation, Cambridge University, 1965.

21. Pease, R. F. W., Broers, A. N., and Ploc, R. A., "Scanning electron microscopy of the growth and subsequent sectioning by sputtering of iron oxide films," *Proc. 4th European Regional Conf. Electron Microscopy*, Prague, 1964, pp. 389–390.

22. Thornton, P. R., Culpin, M. J., and Drummond, I. W., "Direct observation of the high-field regions in GaAs," *Solid-State Electron.*, vol. 6, pp. 532–533, 1963.

23. Everhart, T. E., Wells, O. C., and Matta, R. K., "Evaluation of passivated integrated circuits using the scanning electron microscope," *J. Electrochem. Soc.*, vol. 111, pp. 929–936, 1964.

24. Everhart, T. E., Wells, O. C., and Matta, R. K., "A novel method of semiconductor device measurements," *Proc. IEEE*, vol. 52, pp. 1642–1647, Dec. 1964.

25. Saparin, G. A., Spivak, G. V., and Stepanov, S. S., *Proc. 6th Internat'l Congr. Electron Microscopy*, pp. 609–610, 1966.

26. Czaja, W., and Patel, J. R., "Observations of individual dislocations and oxygen precipitates in silicon with a scanning electron beam method," *J. Appl. Phys.*, vol. 36, pp. 1476–1482, 1965.

27. Everhart, T. E., "Deflection-modulation CRT display," *Proc. IEEE* (*Letters*), vol. 54, pp. 1480–1482, Oct. 1966.

28. Chang, T. H. P., Ph.D. dissertation, Cambridge University, 1966.

29. Wittry, D. B., and Kyser, D. F., "Use of electron probes in the study of recombination radiation," *J. Appl. Phys.*, vol. 35, pp. 2439–2442, 1964.

30. Wittry, D. B., and Kyser, D. F., "Cathodoluminescence at p-n junctions in GaAs," *J. Appl. Phys.*, vol. 37, pp. 1387–1389, 1965.

31. Casey, H. G., Jr., and Kaiser, R. H., "Analysis of n-type GaAs with electron beam excited radiative recombination," *J. Electrochem. Soc.*, vol. 114, pp. 149–153, 1967.

32. Casey, H. G., Jr., "Investigation of inhomogeneities in GaAs by electron beam excitation," *J. Electrochem. Soc.*, vol. 114, pp. 153–157, 1967.

33. Wittry, D. B., and Kyser, D. F., "Measurement of diffusion lengths in direct-gap semiconductors by electron beam excitation," *J. Appl. Phys.*, vol. 38, pp. 375–382, 1967.

34. Davoine, F., and Pinard, P., "Study of cathodoluminescence by scanning electron microscopy," *Proc. 4th European Regional Conf. Electron Microscopy*, Prague, 1964, pp. 113–114.

35. Davey, J. P., "Electron microprobe scanning system for the study of luminescence," Ph.D. dissertation, Cambridge University, 1965.

36. Pease, R. F. W., and Hayes, T. L., "Scanning electron microscopy of biological material," *Nature*, vol. 210, p. 1049, 1966.

37. Thornley, R. F. M., "New applications of the scanning electron microscope," Ph.D. dissertation, Cambridge University, 1960.

38. Everhart, T. E., "Nonthermal electron beam recording," *Proc. 1st Internat'l Conf. Electron and Ion Beam Sci. Technol.*, pp. 341–351, 1965.

39. MacDonald, N. C., and Everhart, T. E., "An electron beam activated switch," presented at IEEE Electron Devices Meeting, Washington, D.C., 1966; also at IEEE Symp. on Electron Ion and Laser Beam Technol., Berkeley, Calif., 1967 (to be published).

40. Wells, O. C., Everhart, T. E., and Matta, R. K., "Automatic positioning of device electrodes using the scanning electron microscope," *Extended Abstr. Electrochem. Soc.*, vol. 12, no. 2, pp. 5–10, 1963.

41. Thornley, R. F. M., and Hatzakis, M., "Electron optical fabrication of solid-state devices," presented at IEEE Electron Devices Meeting, Washington, D.C., 1966; also at IEEE Symp. on Electron Ion and Laser Beam Technol., Berkeley, Calif., 1967 (to be published).

42. Broers, A. N., "Micromachining by sputtering through a mask of contamination laid down by an electron beam," *Proc. 1st Internat'l Conf. Electron and Ion Beam Sci. Technol.*, pp. 191–204, 1965.

43. Broers, A. N., "Combined electron and ion beam processes for microelectronics," *Microelectron. Reliability*, vol. 4, pp. 103–104, 1965.

44. Mollenstedt, G., and Speidel, R., "Newer development in microminiaturization," *Proc. 3rd Symp. Electron Beam Technol.*, pp. 340–357, 1961.

45. Pease, R. F. W., and Nixon, W. C., "Microformation of filaments," *Proc. 1st Internat'l Conf. Electron and Ion Beam Sci. Technol.*, pp. 220–230, 1965.

46. Pease, R. F. W., and Hayes, T. L., "Electron microscopy of living insects," *Science*, vol. 154, pp. 1185–1186, 1966.

47. Crewe, A. V., "Experiments with a field emission gun," *Proc. 6th Internat'l Congr. Electron Microscopy*, pp. 629–630, 1966.

48. Crewe, A. V., "A transmission scanning microscope," *Proc. 6th Internat'l Congr. Electron Microscopy*, pp. 631–632, 1966.

49. Boersch, H., "Inelastic scattering of electrons in solids," *Proc. 6th Internat'l Congr. Electron Microscopy*, pp. 9–10, 1966.

50. Castaing, R., El Hili, A., and Henry, L., "Microanalyse qualitative par images filtrées," *Proc. 6th Internat'l Congr. Electron Microscopy*, pp. 93–94, 1966.

51. Smith, K. C. A., "A versatile scanning electron microscope," *Proc. European Regional Conf. Electron Microscopy*, Delft, 1960, pp. 177–180.

52. Wells, O. C., "A bibliography of scanning electron microscopy," *Proc. 9th Ann. Symp. Electron Ion and Laser Beam Technol.*, 1967 (to be published).