VOLUME I ☐ NUMBER 4 APRIL 1964

# IEEE spectrum

## features



## departments



THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

## the cover

The assemblage of components shown is a small portion of an IBM thin-film panel that has been developed for a Navy airborne computer. The entire panel consists of 32 NOR logic circuits in a space 2.5 by 1.865 inches. Film circuits such as this permit the use of circuit elements with close tolerances and the integration of a large number of passive components and their interconnections. In this issue, a three-part series on thin films begins. The first article beginning on page 72 deals mainly with the two deposition techniques of cathode sputtering and vacuum evaporation and with their use in the fabrication of film resistors, capacitors, and *R-C* networks. Subsequent articles will be concerned with thin-film transistors and cryogenic thin films.




**IEEE SPECTRUM** is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: Box A, Lenox Hill Station, New York, N.Y. 10021. Cable address: ITRIPLEE. Telephone: 212—PL 2-6800. Published at 20th and Northampton Sts., Easton, Pa. **Change of address** requires thirty days notice to IEEE Headquarters. **Annual subscription:** IEEE members, first subscription included in dues, one additional subscription $13.50. College and public libraries, $13.50 in U.S.A. and Canada, $14.50 elsewhere. Others, $18.00 in U.S.A. and Canada, $19.00 elsewhere. Single copies: members, one additional copy $1.00, college and public libraries, $1.50 in U.S.A. and Canada, $1.75 elsewhere. Others, $2.00 in U.S.A. and Canada, $2.25 elsewhere. **Editorial correspondence** should be addressed to IEEE SPECTRUM at IEEE Headquarters. **Advertising correspondence** should be addressed to IEEE Advertising Department, 72 W. 45 St., New York, N.Y. 10036. Telephone: 212—MU 2-6606.

**This twentieth century is undoubtedly an electronics era, rich in rapid evolutionary development. It occurred to us that perhaps there was some electronic activity in the nineteenth century or earlier. Jack Ramsay found that much of electronics was initiated prior to 1900 in rudimentary form.**

# The Electronics Spectrum Before 1900

Our Figure 1 shows how applied electronics was born. It is well-known that the "Edison effect" was the discovery of electron emission from a lamp filament, and that Edison made a diode vacuum tube, to establish that a current could cross a vacuum. It is characteristic of Edison that he immediately sought a practical application for his phenomenon. In 1884 he exhibited a vacuum tube voltmeter used to monitor the supply voltage of his illumination circuits. He had observed that the diode plate current was a highly sensitive indication of the filament voltage and, in a sense, he anticipated Richardson's equation. The patent is a classic.

Chunder Bose's microwave detector using steel springs is shown in Figure 2, for both microwaves and point-contact rectifiers ("coherers") existed before 1900. In Bose's rectifier, the springs were subject to a light pressure and a small polarizing voltage. The sensitivity was phenomenal. Bose regarded his point-contact devices as organisms that could be elated or depressed by drugs. He later became a physiologist. It is curious that modern impurity doping uses precisely the conception introduced by Bose. He actually treated his contacts with chemicals and measured the response to a "flash of radiation" (pulse) using a homemade recorder. All this modern sophistication was carried out in the late 1890's without vacuum tubes or modern theory. There is also evidence that Bose discovered negative resistance devices.

Bose was the most advanced microwave experimenter of his time. One of the followers of Hertz, he succeeded in generating microwaves down to 5 mm and produced a fantastic variety of devices and techniques. Waveguides, horn antennas, "cut-off" gratings, dielectric lenses, microwave reflectors, the double-prism directional coupler, polarimeters, interferometers, dielectrometers—the list is like a modern laboratory catalog. No modern microwave engineer should be ignorant of Sir Jagadis Chunder Bose's accomplishments in microwave technique.

Figure 3 shows the circular waveguide antennas excited by spark generators originated by Oliver J. Lodge in 1894. Lodge was in the microwave business before Bose but did not develop it as extensively. He originated coherers, waveguides, and lenses. Extensive microwave research was carried on between 1888 and 1900. Hertz's fundamental experiments establishing the identity of light waves with microwaves were copied and extended by many investigators, both in the U.S.A. and Europe. Hull built a 9.1 cm

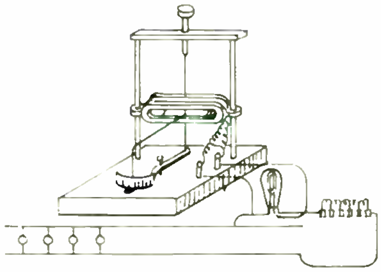


FIGURE 1. EDISON'S VACUUM-TUBE VOLTMETER (1884)

interferometer in Chicago in 1897. Much was done in Europe. Righi at Bologna even produced a textbook on microwave optics, which is still a masterpiece. From Righi's work Marconi developed radio communication in 1896, though Edison had a radio system in 1885, also using spark generation and elevated antennas. Pressure of other interests prevented Edison and Lodge from developing communication without wires.

The application of radio waves proposed before 1900 includes the following: Ship detection and collision avoidance (Edison, Branly), direction-finding (Brown), aeronautical communications (Vallot), radio relay (Guaraini), medical (D'Arsonval), countermeasures (Tommasini, Marconi). Tesla visualized radio guidance, time signals, radio surveying, storm detection, latitude and longitude determination, rf power transmission and so on. Most striking in anticipating modern optical communications was Graham Bell's successful voice-modulation of a light beam by a microphone, detected at 500 yards with a selenium rectifier in a paraboloid reflector (1880).

Many physical "effects" discovered before 1900 have come to be foundation stones of modern electronic technology. We have space to mention only a few: Seebeck effect (1821), Peltier effect (1834), Faraday effect (1845), Kerr effect (1875), Hall effect (1879), piezoelectric effect (1880), photo-electric effect (1887), X-rays (1896), and radioactivity (1897).

The most significant electrical engineering before 1900 was, surprisingly, telegraphy. Starting around 1830 (Henry) telegraphy was dominant for 50 years. Power engineering arose later, largely because Edison acquired generators for his electric light illumination from 1882 onward. Edison's love for DC had to give way to AC techniques developed by Tesla, Westinghouse, and others; the three-wire transmission system is, however, due to Edison.

The idea of a civilization based on the applications of electricity will be found in the writings of none other than that great American, Benjamin Franklin.

The illustrations are taken from the following reference with acknowledgements:

Figure 1

W. C. White, "Electronics . . Its Start From The 'Edison Effect' Sixty Years Ago," General Electric Review, p 540, October 1943.

Figure 2

"Collected Physical Papers of J. C. Bose," Longmans Green & Co., 1927.

Figure 3

"Signalling Through Space Without Wires," The Electrician Printing and Publishing Co., Ltd., London, England, 1898.

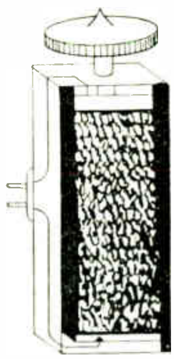


FIGURE 2. BOSE'S SEMICONDUCTOR RECTIFIER (1895)

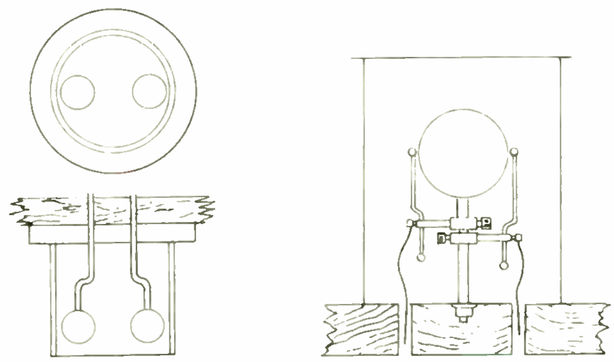


FIGURE 3. LODGE'S WAVEGUIDE ANTENNAS (1894)

# The ART of engineering

**A synthetic ruby crystal observed in cross section** by laser interferometer techniques shows high contrast of optical fringes, also visible in the background. The 6-inch-long rod has a ½-inch diameter; it was inspected with a Perkin-Elmer instrument consisting of a gas laser that emits continuous coherent light at 6328 Å and a modified Twyman-Green interferometer. The instrument has enabled inspection of optical materials with double or more the accuracy previously obtainable, and in some cases it permits studies never before undertaken. The laser's beam can produce 3000 times more useful light energy than an arc light.

# Reflections

## 75 years ago

**Magnetism.** "It has been found that the secondary current wave of a closed iron circuit induction coil or transformer, whose primary circuit receives alternating current, is lagged from its theoretical position of 90 degrees behind the primary wave, an additional 90 degrees, so that the phases of the two currents are directly opposed; or the secondary current, working lamps only in its circuit, is one-half a wave length behind the primary, instead of only a quarter wave length, as might have been expected.

"But when it is understood that the iron core polarized in one direction by the primary impulse does not begin to lose its magnetism when that impulse

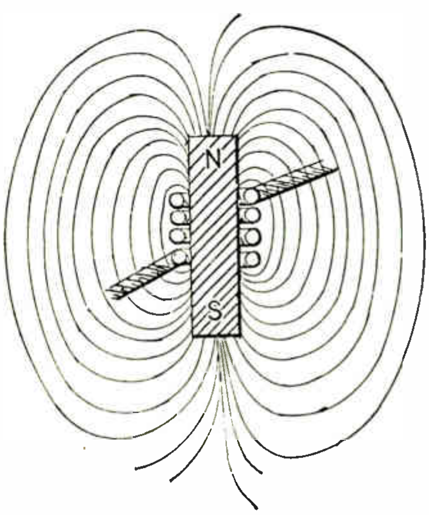



Fig. 7

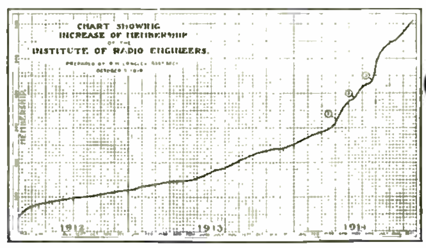
simply weakens, but waits until an actual reversal of current has taken place, it will be seen that the secondary current, which can only be produced when magnetic lines are leaving the core and cutting the secondary coil, or when the lines are being evolved and passing into the core from the primary coil, will have a beginning at the moment the primary reverses, will continue during the flow of that impulse, and will end at substantially the same time with the primary impulse, provided the work of the secondary current is not expended in overcoming self-induction, which would introduce a further lag. Moreover, the direction of the secondary current will be opposite to that of the primary, because the magnetic circuits which are opened up by the primary current in magnetizing the core, or which are closed or collapsed by it in demagnetizing the core, will always cut the secondary coil in the direction proper for this result. Transformers of the straight core type with very soft iron in the cores and with not too high rates of alternation should approximate more nearly the theoretical relation of primary and secondary waves, because the magnetic changes in the core are capable of taking place almost simultaneously with the changes of strength of the primary current. This fact also has other important practical and theoretical bearings.

"Let us assume a plain iron core, Fig. 7, magnetized as indicated, so that its poles N S complete their magnetic circuits by what is called free field or lines in space around it. Let a coil of wire be wound thereon as indicated. Now assume that the magnetism is to be lost or cease, either suddenly or slowly. An electric potential will be set up in the coil, and if it has a circuit, work or energy will be produced or given out in that circuit, and in any other inductively related to it. Hence the magnetic field represents work or potential energy. But to develop potential in the wire the lines must cut the wire. This they can do by collapsing or closing on themselves. The bar seems, therefore, to lose its magnetism by gaining it all, and in doing so all the external lines of force moving inward cut the wire. The magnetic circuits shorten and short circuit themselves in the bar, perhaps as innumerable molecular magnetic circuits interior to the iron medium. To remagnetize the bar, we may pass an electric current through the coil. The small closed circuits are again distended, the free field appears, and the lines moving outward cut across the wire coil opposite to the former direction, and produce a counter potential in the wire, and consequent absorption of the energy represented in the free field produced. As before studied, the magnetism cannot disappear without giving out the energy it represents, even though the wire coil be on open circuit, and be therefore unable to discharge that energy. The coil open circuited is static, not dynamic. In such assumed case the lines in closing cut the core and heat it. Let us, however, laminate the core or subdivide it as far as possible, and we appear to have cut off this escape foı the energy. This is not really so, however. We have simply increased the possible rate of speed of closure, or movement of the lines, and so have increased for the divided core the intensity of the actions of magnetic friction and local currents in the core, the latter still receiving the energy of the magnetic circuit. This reasoning is based on the possibility in this case of cutting off the current in the magnetizing coil and retaining the magnetic field. This is of itself probably impossible with soft iron. That the core receives the energy when the coil cannot, is shown in the well-known fact that in some dynamos with armatures of bobbins on iron cores, the running of the armature coils on open circuit gives rise to dangerous heating of the cores, and that under normal work the heating is less. In the former case the core accumulates the energy represented in the magnetic changes. In the latter the external circuit of the machine and its wire coils take the larger part of the energy which is expended in doing the work in the circuit. In this case, also, the current in the coils causes a retardation of the speed of change and extent of change of magnetism in the iron cores, which keep down the intensity of the magnetic reaction. In fact, this retardation or lag and reduction of range of magnetic

change may in some machines be made so great by closing the circuit of the armature coils themselves or short circuiting them, that the total heat developed in the cores is much less than under normal load." (Elihu Thomson, "Magnetism in its Relation to Induced Electromotive Force and Current," *Trans. AIEE*, vol. VI, Nov. 1888–Nov. 1889, pp. 280–282.)

## 50 years ago

**Growth Stock.** "Thru the kindness of Mr. Ralph H. Langley, Assistant Secretary of the Institute, the Editor is enabled to present to the membership the accompanying chart. This interesting curve shows the increase in membership of the Institute of Radio Engineers between May, 1912 and September, 1914. Since the preparation of this chart there has been a further considerable increase in membership. Attention is called to the accelerated rate of increase in membership directly above each of the points 1, 2, and 3 of the chart. At each of the




points indicated by these figures, systematic efforts to increase the membership were made by the Board of Direction. The results were gratifying, and lead the Board to expect still greater increases in membership so soon as the advantages of participation in the Institute activities become even better known to the radio fraternity." (*Proc. IRE*, June 1914, p. 129.)

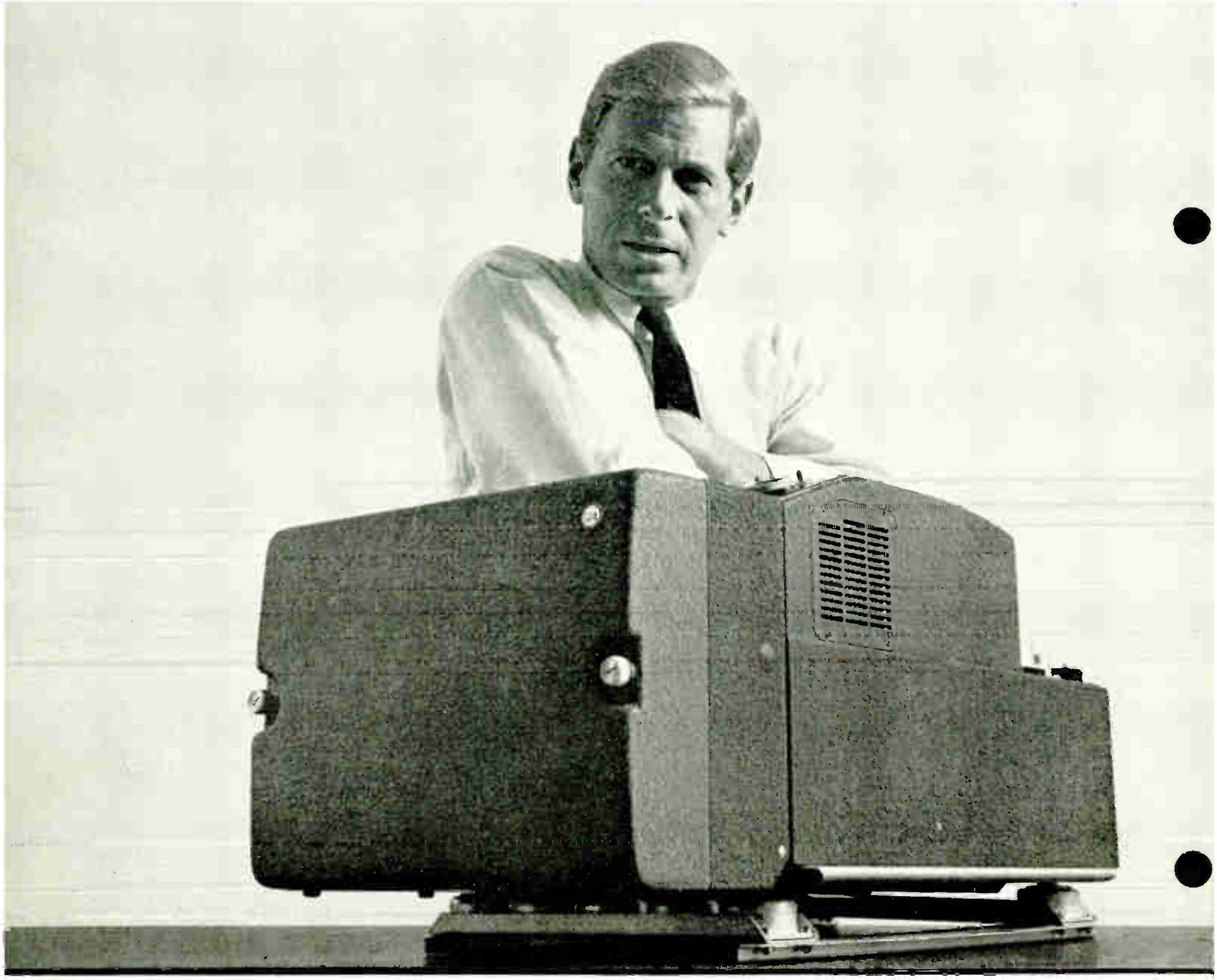
**Mine Duty Controllers.** "The application of the electric motor to bituminous coal mines has developed so rapidly that there is no section of the industry in which the motor has not been installed. Not only is the application universal, but the use of electricity in coal mines is steadily increasing, principally due to the increased cost of producing steam at the collieries.

"The design of satisfactory motor and control equipment for coal mines is a considerable problem, because of the severe operating conditions encountered. The class of attendants usually employed by the mines to maintain electric motors and their accessories, is unskilled, principally because of the location of the mines and the conditions of work. Many changes in help take place for the same reasons, and it is, therefore, difficult to train the attendants and make experts of them. These conditions make it essential that rugged and reliable electrical equipment be installed.

"As the use of electricity in the mines has increased, the electric motor for operating mining machinery has been developed much more rapidly than suitable controlling devices. Only within the past few years have those concerned realized that the selection of a proper controller is as important as the selection of a suitable motor. Too often is the controller given secondary consideration with the result that a good motor may appear to disadvantage, simply because it is not operated with the right control equipment.

"At coal mines, motors are used for haulage, hoisting, ventilating, pumping, coal-cutting, tipple or breaker power, drilling, washing, machine shop and blacksmith shop. In the design of electrical equipment for mines, it is necessary to consider carefully the following: explosive dust and gases, continuity of operation, voltage fluctuations and dampness. It is, of course, advisable to install the control apparatus in a locality which is unaffected by sparks or short circuits tending to some derangement of the apparatus. The design, however, must be made so as to minimize leaks to ground, short circuits, etc., which may cause ignition of explosives, of mine gases or of coal dust, with disastrous results. In gaseous mines, arc producers such as circuit breakers, switches and sliding contacts of rheostats must be properly protected, either by breaking the arcs in oil, or by providing explosion-proof cases.

"On account of the unskilled attendants in the mines, the electrical equipment receives less than the usual amount of intelligent attention, while on account of the conditions of operation, it should receive more. Much is therefore left to the designing engineer to solve, but on the other hand, much could be done to improve conditions by the employment of a supervising electrical engineer at an attractive salary, whose duty it would be to *see* that better attention is paid to the care and maintenance of the electrical apparatus." (H. P. Reed, "Mine Duty Control-



# We gave them our best – but they wanted more

**We had given our customers the world's finest oscillograph. It was accurate, rugged and reliable beyond compare – and it could simultaneously produce 26 different channels of test data. But progress was demanding a new standard.**

The race for space was making the field of dynamic measurement increasingly complicated. Every day brought new temperature, pressure, velocity and acceleration tests. An oscillograph was urgently needed that could do everything CEC's "Old Faithful" could do – and do it with *twice* as many recording channels.

But how do you double the recording capacity of an oscillograph – design it to accept all photographic media (conventional paper, film and print-out-paper) – and still keep it small enough to be practical? Our engineers had an immediate answer: "We aren't sure." After many months of constant development, trial and error – they still weren't sure.

Then came their final answer: "We've got it." That was more than eight years ago. Now the world's first "universal" oscillograph is a trusted friend to virtually every installation where advanced oscillography is required.

This is a "graphic" example of how customer demands are setting new standards – and will continue to set them. And it explains why CEC has remained the leader in the development and production of all phases of data recording.

It's also the reason why CEC maintains a network of 22 sales and service offices throughout the nation. To expedite application engineering; and to help train customer personnel in the use of advanced instrumentation.

CEC

**CONSOLIDATED ELECTRODYNAMICS**

A SUBSIDIARY OF **BELL & HOWELL**/PASADENA, CALIF. 91109
INTERNATIONAL SUBSIDIARIES: WOKING, SURREY, ENGLAND AND FRANKFURT/MAIN, GERMANY

lers," *Trans. AIEE*, vol. XXXIII, part I, Jan. 1914 to June 24, 1914, pp. 367–368.)

## 25 years ago

**Electromagnetic Horns.** " 'Wave-Guide Radiators and Electromagnetic Horns' was the subject of a demonstration-lecture given by G. C. Southworth and A. P. King of the Bell Telephone Laboratories.

"This was a continuation of a similar lecture given a year previously on the use of wave guides as transmission lines. A series of experiments showed the characteristics of wave guides employed as radiators. The electromagnetic horn was considered a special form of such a radiator. These horns may be used either singly or in arrays, to produce by directivity effective power ratios of several hundreds or possibly thousands. The directional characteristics of such horns were demonstrated. Most of the demonstrations employed frequencies of about 3000 megacycles corresponding to wave lengths shorter than ten centimeters."

February 1, 1939—President Heising, presiding. (Report of New York Section meeting, *Proc. IRE*, April 1939, p. 290.)

**The Engineering Profession.** "But our principal concern here is the engineering profession, and we inquire, what is the engineering profession; is it a profession at all; and if it is, will it develop into the full stature to which the importance of its works entitles it to aspire?

"It is relatively young. The military engineer appeared in the first steps of the mechanization of warfare, when forts began to take shape. His counterpart in peaceful affairs was called a civil engineer. With the industrial revolution, and especially with the spread of mechanization from the factory into every walk of life, engineering became exceedingly diversified. Applying science in an economic manner to the needs of mankind is its broad field. Its disciplines are spread over all the sciences as they become thus applied, and embrace also portions of economics, law, and business practice, which are integral parts of the process of application. It is somewhat loosely organized as professions go. To a minor extent, only, it limits its numbers; but the very strictness of its essential disciplines provides some selection of its neophytes. Until recently it has done very little in an organized fashion to inculcate in its younger members the philosophy of the profession, leaving this largely to those of its individual who are also members of the teaching profession. That branch which represents the consultant, and others to a degree, have drawn codes; but there is no body of codified principles which is accepted and applied by the profession as a whole. It has no highly distinct language or jargon, for it must continuously work with laymen.

"These are, however, incidentals. The important point is this: Does it have a central theme of ministering to the people? Most certainly it serves the public in myriad ways, but are its individual members activated primarily by the professional spirit of dignified and authoritative counsel and guidance?" (Vannevar Bush, "The Qualities of a Profession," *Electrical Engineering*, vol. 58, April 1939, pp. 157–158.)

**Educational Broadcasting.** "Proceedings of the Second National Conference on Educational Broadcasting. Held in Chicago, Illinois, on November 29, 30, December 1, 1937. Edited by C. S. Marsh, Published, 1938, by the University of Chicago Press, Chicago, Illinois. 387 pages. Price $3.00.

"This volume, appearing many months after the conference was held of which it is a record, is still of interest, largely because of the continuing question of the desirability of control or regulation of broadcasting in America.

"The principal objective of the conference was 'to provide a national forum where interests concerned with education by radio could come together to exchange ideas and experiences.' The general plan for the two and one-half days of its duration included four general sessions with prepared addresses representing the industry, the audience, and education. There were two afternoons of discussion sections and a banquet with speeches on the general topic.

"The conference was attended by nearly five hundred registrants, with a broad national representation—eleven college presidents and about two hundred other representatives largely interested in some phase of education—about seventy-five from other welfare, social, or national organizations, fifty-five representatives of the radio industry, and scattered groups of journalists, publishers, editors, advertisers, and federal and state government officials.

"The volume is not of primary direct interest to the radio engineer as such. The advanced state of his art and his technical excellence seem to be granted by all speakers. Nor does the volume particularly concern the artist who appears in radio programs. His great success in entertaining and in holding interest and again his technical excellence is granted. This is a volume which records the growing pains of the new public utility, the broadcast industry, and the inevitable conflict between the entrepeneurs who had the foresight to develop the industry and who built their own system and others who now feel shut out of the necessarily closed system.

"The speeches delivered in the four general sessions and constituting the material of the present volume were in general carefully prepared, sometimes very witty and, to judge from the remarks of the chairmen, were in some cases delivered by persons particularly well studied in the art of talking before a microphone. It was, apparently, the opinion of one spokesman for informal lobby and dining room discussions that most of the prepared speeches were largely favorable to this American system and that the smoothness of the conference did not reflect a general satisfaction with the system as operated." (Book Review by K. S. Van Dyke, *Proc. IRE*, April 1939, p. 296.)

**Engineering Experience.** "Broadly, character of experience that is satisfactory is that which indicates competency to practice professional engineering, and the 'ability to do.' Specifically, the following items are pertinent.

"The school of experience should be a laboratory for the application of principles to the solving of problems. Knowledge and experience should become integrated so that they interact and produce something far beyond the arithmetical sum of textbook information and the doing of routine tasks.

"On the same job one may exercise judgment in directing the work, while another may simply follow plans and specifications to the letter. Again, in design one may use judgment in choice of standards while another may make slavish use of them.

"Qualifying experience is not derived from work that does not require or involve the application of engineering principles and theories, although the work itself may require skill and may be necessary in professional activity, for example, ordinary surveying, drafting, mechanical calculations, or routine cost accounting." (C. F. Scott, "What Is Engineering Experience?" *Electrical Engineering*, vol. 58, April 1939, p. 153.)

# Spectral lines

**How Reduce the Damping?** In the past two months I have used this space to develop a rationale for the many changes in engineering education that have occurred in the last 25 years. This month it is my intent to close out the trilogy by discussing the work remaining ahead.

It was almost trite last month to say that we were witnessing a technological explosion; explosion seems an inaccurate metaphor, since we usually assume an explosive event to end, after which we take on the job of restoring the status quo, provided that the explosion left us with any status. The technical changes of recent years are also referred to as a technological revolution, but these words also seem to carry an implication of eventual return to a steady state of merely differing characteristics. The past state or a steady state is not to be expected today.

Furthering the thought through technical language, it could be said that the changes of the last quarter century should not be looked upon as a random event, but rather as a series of forcing functions which will produce transient responses in the system to the end of time. The best title that I have yet found to describe our situation seems to be "technical evolution." Such a title implies change; it implies selection of appropriate means and devices; it suggests competition, and conveys the thought of continuing and unending progress under the control of the environment and the forces existing at the time.

If we are in an evolutionary condition, we must accept completely a philosophy of growth, expansion, and new challenge. Any return to past days of ease, to days of stable technology, is both unthinkable and contrary to the purpose of the engineer—a person who must be dedicated to change and progress if he is to improve the lot of man. We must realize that engineering education is for a changing world in which engineers will answer questions by methods now unknown.

A system under forcing can be studied by analysis of its responses. One observation to be made of our engineering manpower system is that, in its response to change, its characteristics show heavy overdamping. This should be no surprise, because it is said that of all engineers ever educated, at least 90 per cent are living today. Major changes have appeared in our curricula in only the past 15 years, and my guess is that less than 15 per cent of our estimated 850000 engineers can be assumed to have received a modern engineering education. Those of age 35 or over graduated from college before most of our new disciplines were visualized, and many years before they were regularly taught. They have not had the opportunity to acquire new skills and knowledge.

To profit from the inherent abilities and experience of these older engineers, continued education seems to be required. Formal programs embracing the new knowledge will allow them to turn the technical corners around which their fields have moved, then to progress to positions of responsibility.

For many years some colleges of engineering have conducted programs in extension or continuing education, but these are not the programs of which I write. In fact these programs all too often pointed in the other direction—toward upgrading a high school graduate to do more adequately the work of a technician. Short courses for such practitioners, review for professional examinations, or undergraduate work on the nondegree level were usual, but little was done for the engineering graduate to allow him to maintain his place in a fast-moving technology. Today we are beginning to recognize the need for both short and long courses, directed not toward graduate degrees but providing modern technology in concise and direct form to the men now in supervising or managerial positions. The Engineers Joint Council is discussing methods, and some schools are already offering seminar series and courses of varying lengths, but much greater availability is needed. Here, perhaps, using closed-circuit television, we can extend the abilities of our thin faculty ranks, and make available experts from the laboratories.

Not all industry is aware of the need for re-education—although our electrical industry is largely on board. One large corporation depending upon research for new products has stated that it must undertake such re-education; a second company oriented to production believes it inadvisable to give up the time of the men concerned. Industry must supply free time for the men to be benefited; employers must provide part of the motivation to speed the response of the system to technological needs.

Many engineers work with companies too small to undertake their own programs; others work in locations remote from our educational centers. Provision of needed educational media for such men then becomes a responsibility of their technical societies. Engineering journals must become more than technical news magazines; they must pull their members always toward increased technical competence. This is a major tenet of IEEE publications policy; we see incomplete acceptance of the view in other societies.

For the recent graduate there must be opportunity for technical advance through graduate work. Graduate credit programs are already available in our major cities. and are contributing to the employment of new graduates.

We should make good use of short annual seminars for the practicing engineer. At present our job is to reduce the backlog, and to assure that our colleges provide a solid base for today's graduates. Engineering education must drop the concept that it is teaching to earn, and adopt a new philosophy that it is educating for future learning.

*J. D. Ryder*

# Developments in broadband antennas

*A survey is presented for the purpose of providing the nonspecialist with a basic understanding of the remarkable advances that have taken place over the past decade in the field of broadband antennas*

*E. C. Jordan, G. A. Deschamps, J. D. Dyson, R. E. Mayes* University of Illinois

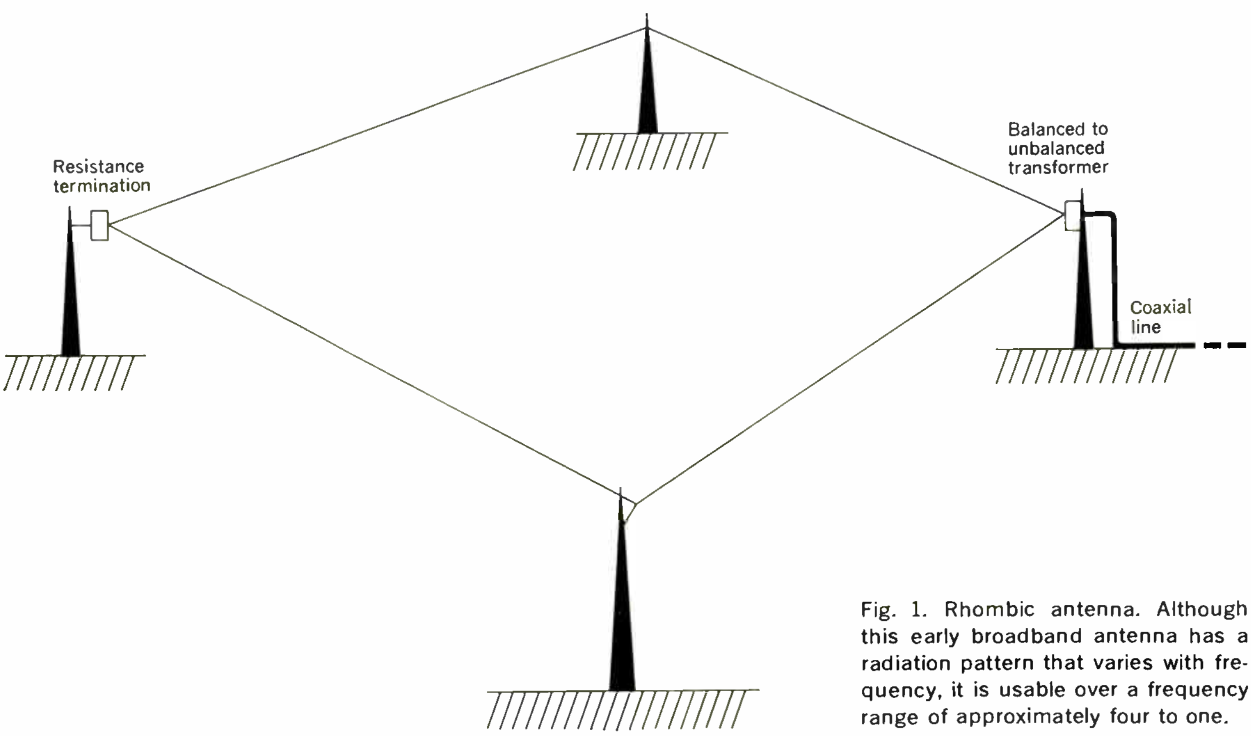



Fig. 1. Rhombic antenna. Although this early broadband antenna has a radiation pattern that varies with frequency, it is usable over a frequency range of approximately four to one.

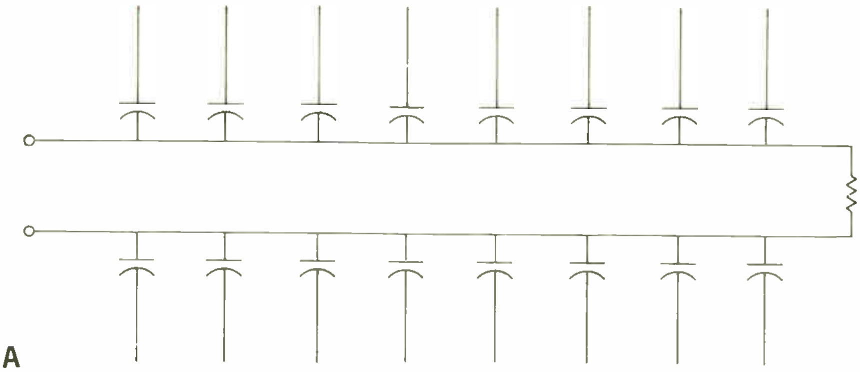
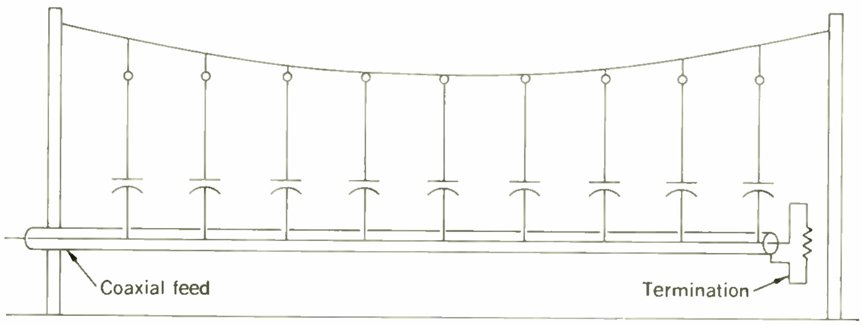



Fig. 2. The fishbone (A) and comb (B) antennas, which have a pattern and impedance that are relatively constant over a two-to-one frequency band.




Until a few years ago the ultimate limiting factor on the bandwidth of any communication system using radiated-wave propagation was most frequently the antenna. Since the antenna performs the dual functions of an impedance-matching device and a directional radiator, the characteristics of major importance are its impedance and its directional properties. Depending upon the application, one or the other (or both) of these characteristics may limit the useful bandwidth. The required or desired bandwidth also varies markedly with application; for example, the required bandwidth may vary from a few cycles per second for a single-channel VLF system up to about 6 Mc/s for a single television channel. For multichannel operation, the radio broadcast band covers a three-to-one bandwidth, the VHF television band covers a four-to-one bandwidth, and the high-frequency communication band covers a ten-to-one bandwidth, from 3 to 30 Mc/s. Finally, for countermeasures work it is usual to state that the desired frequency range extends from "dc to light."

Because of the wide variety of operational requirements that exist, there is no unique definition of antenna bandwidth. For our purposes, a broadband antenna will be considered to be one which retains certain desired or specified radiation pattern, polarization, or impedance characteristics over more than an octave (that is, a two-to-one frequency range).

## Early broadband antennas

Some of the earliest broadband antennas were long-wire types designed to operate in the high-frequency (short-wave) band or in the low-frequency band. For the most part, they were broadband only in the sense that impedance remained relatively constant over the useful range; in general, no attempt was made to achieve a constant pattern. Among these antennas the well-known rhombic antenna has held a dominant place since the early days of radio. This antenna (Fig. 1) is essentially a resistance-terminated transmission line that has been opened out to form the four sides of a rhombus. Because of the traveling-wave current distribution along the terminated line, the main beam is in the forward direction (toward the termination) at an elevation angle that depends, in a complicated fashion, on the included angle of the rhombus and the lengths of the sides in wavelengths. Fortunately the beam is quite broad in the vertical plane and the angle above ground of the maximum increases as the frequency decreases. This change of angle with frequency is in the correct direction for transmission or reception of ionospherically reflected waves, so a rhombic antenna of fixed dimensions is usable over a wide frequency range (of the order of four to one) in the short-wave band.

The wave antenna, consisting of a long, elevated wire parallel to the ground and resistance-terminated at both ends, is another traveling-wave-type antenna. In contrast with most antennas, which operate best over a highly conducting ground, the wave antenna depends for its operation upon the finite conductivity of the earth beneath it. An incident radio wave traveling along the surface of a finitely conducting earth has a forward tilt and a horizontal component of electric field intensity. It is this horizontal component of electric field, produced by the finite earth conductivity, that induces a traveling wave of voltage in the horizontal wire and the resulting antenna action. Because the antenna has an impedance that is nearly independent of frequency, it is known as an aperiodic antenna. However, the radiation pattern does vary with the length of the wire in wavelengths, and hence with frequency. Wave antennas are used for long-wave or low-frequency reception.

The fishbone receiving antenna consists of a long resistance-terminated transmission line loosely coupled by capacitors to an array of closely spaced (less than $\lambda/4$), untuned, horizontal dipoles; see Fig. 2(A). For vertically polarized signals, one half of a fishbone antenna is erected vertically and fed against ground to form a comb antenna; see Fig. 2(B). Tapering the coupling capacitors to larger values towards the termination equalizes the antenna currents and reduces resonance effects. Because the capacitive coupling of the elements to the transmission line is tighter at the higher frequencies, fewer of the elements are strongly excited. Hence, the effective length of the array varies inversely with frequency in such a manner as to maintain a fairly constant pattern, gain, and impedance, over the useful bandwidth of more than two to one. The fishbone and comb antennas have been described chiefly for comparison with the log-periodic dipole and log-periodic monopole types to be described later.

In contrast to the terminated wire and loaded transmission line types just described, there is a class of antennas that owes its broadband properties to broad, specially shaped surfaces. It was recognized quite early that a fat dipole had a much lower antiresonant (full-wavelength) impedance than a thin one, and that in general, fat antennas had smaller impedance variations than thin ones. The importance of broad surfaces was emphasized by Schelkunoff in the treatment of the biconical antenna, and many broad-surfaced specially shaped antennas found early application in television transmitting antennas and countermeasures antennas. A very successful broadband antenna was the discone[1] (a cone fed against a disk), which maintained good impedance and pattern characteristics over a four-to-one bandwidth (Fig. 3). For countermeasures work many antennas having surfaces of various shapes were developed,[2] some of which had remarkably wide impedance bandwidths and usable pattern bandwidths of the order of five to one. It must be admitted, however, that most of these early designs were arrived at by an intuitive or cut-and-try approach.

Another group of antennas, some of which display fairly wide bandwidths, consists of various helical and spiral shapes. When the circumference of a helical antenna is of the order of a free-space wavelength, the antenna radiates in the axial mode—that is, with the maximum radiation along the axis of the helix. In this mode, the helical antenna has desirable impedance, pattern, and circular polarization properties over nearly an octave.[3] By expanding the diameter of the helix along its length to form a conical monofilar helix fed from the base end, Springer[4] showed that the bandwidth could be increased. His observation that there appeared to be an effective aperture that moved toward the smaller end of the cone as the operating wavelength decreased was perhaps the first indication of things to come. Later, Chatterjee[5,6] also considered monofilar helical antennas formed on a conical surface and fed against a ground plane. He demonstrated that they could be excited from either end, and obtained usable bandwidths of approximately four to one. At about the same time, Turner[7] proposed a balanced antenna constructed in the form of an Archimedes spiral. This planar antenna, constructed with narrow constant-width arms and radiating a broad lobe on each side of the structure, gave promise of being usable over the then remarkable bandwidths of between seven and eight to one.

## Frequency-independent antennas

In 1954, Rumsey[8] put forth the idea that a structure entirely definable by angles, without any characteristic length dimension, should have properties that are independent of the frequency of operation. However, all such angle structures extend to infinity, so the key question was which of such structures retained these frequency-independent characteristics when truncated to a finite length. It should be noted that the well-known biconical structure is an angle structure that is *not* frequency independent when it is truncated to form a practical antenna. Both impedance and pattern vary with frequency for any finite length.

Rumsey proposed that an equiangular spiral structure, which satisfies the angle requirement, might have the desired properties, and Dyson[9-11] undertook a comprehensive experimental study of an antenna based on the equiangular spiral geometry shown in Fig. 4. The equiangular or logarithmic spiral* is defined by

$$\rho = e^{a(\phi-\delta)} \quad \text{or} \quad \phi - \delta = \frac{1}{a}\ln \rho$$

where $\rho$ and $\phi$ are conventional polar coordinates, and $a$ and $\delta$ are constants. In Fig. 4 the edges of the metallic arms are defined by

$$\rho_1 = ke^{a\phi} \quad \text{and} \quad \rho_2 = ke^{a(\phi-\delta)}$$

for one arm, and by

$$\rho_3 = ke^{a(\phi-\pi)} \quad \text{and} \quad \rho_4 = ke^{a(\phi-\pi-\delta)}$$

for the other arm, where the constants $a$, $k$, and $\delta$ determine the rate of spiral, size of the terminal region, and arm width, respectively. With this particular spiral the angle between the radius vector and the spiral remains the same for all points on the curve—hence the term "equiangular spiral." Experimental investigation established that this particular geometry did indeed retain its frequency-independent properties after truncation, and this design was the basis for a large class of successful frequency-independent antennas.

When this angular structure is excited in a balanced manner at the origin, the current flows outward with small attenuation along the spiral arms until a region of given size in wavelengths is reached. In this region (the active or radiating region) essentially all of the incident energy transmitted along the spiral arms is radiated, and somewhat beyond this region the presence or absence of the arms is of no consequence. Because the radiating region is of constant size in wavelengths, it moves toward the origin as the wavelength of operation decreases. The size of effective radiating aperture thus automatically adjusts or scales with frequency of operation in such a manner that the antenna behaves the

* The logarithmic spiral was first discussed by Descartes (1638) and later (1691–1693) studied by Jacques Bernouilli, who gave it its name. Bernouilli was so delighted by the property of the spiral reproducing itself under various transformations that he requested that the spiral be engraved on his tomb with the inscription "Eadem Mutata Resurgo."

same at all frequencies. Because of the spiraling of the arms, this scaling is accompanied by a rotation of the radiated field about the axis of the antenna.

It is now known that this automatic scaling of the radiating aperture is a condition for operation in a frequency-independent manner. It is interesting to note that Springer observed this phenomenon on the expanding helix, but unfortunately the methods of construction and excitation limited the bandwidth obtainable to something over an octave, so the importance of scaling of effective aperture with frequency was not fully recognized. Chatterjee's measurements also show evidence of scaling with frequency in the near-field amplitude plots from which he calculated radiation patterns; but again, possibly because of the physical configuration and method of feed, the full significance of this scaling does not appear to have been appreciated. In a similar manner, the radiating aperture of the Archimedes spiral antenna tends to scale with frequency; however, because the width and spacing of the spiral arms in the radiating region are not constant in wavelengths, as frequency is varied, the antenna characteristics change (albeit slowly) with frequency.

At this point it is necessary to define the term "frequency independent" when it is used with a practical finite-sized structure. If the antenna is excited by a voltage applied between the two arms at the origin, it has an impedance and radiation pattern that are essentially constant* (that is, independent of frequency) for all frequencies above that for which the outer diameter of the truncated structure is approximately half a wavelength up to the frequency at which the diameter of the feed region (as determined by the transmission line feed) is comparable with a half wavelength. Since these two dimensions can be specified independently, the design bandwidth can be made arbitrarily large; actually it is limited only by practical considerations of construction—that is, how large the outer diameter is made and how finely the geometry at the feed region can be modeled.

The equiangular spiral antenna, which is bidirectional, radiates a very broad, circularly polarized beam on both sides of its surface. This bidirectional characteristic severely restricts its utility in practice, but a modified version, to be described later, provides a highly practical, extremely broadband antenna.

**Log-periodic antennas**

In 1955, working with Rumsey on broadband antenna development, DuHamel[12] proposed that it should be possible to force radiation from otherwise "angle structures" by the use of appropriately located discontinuities. One of the first geometries chosen to investigate the validity of this concept was that shown in Fig. 5. Here two wedge-shaped metallic angle structures have teeth cut into them along circular arcs. The radii of the arcs which define the location of successive teeth are chosen to have a constant ratio $\tau = R_{n+1}/R_n$. This same ratio $\tau$ defines the lengths and the widths of suc-

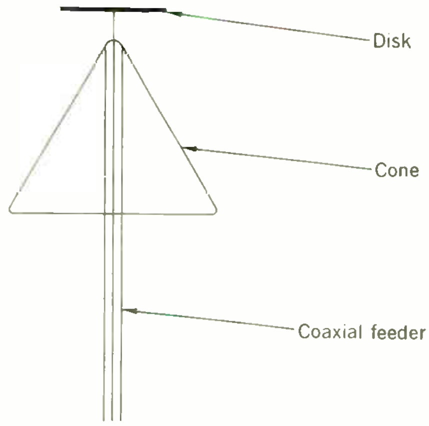



Fig. 3. The discone, a successful early broadband antenna having a useful bandwidth of approximately four to one.

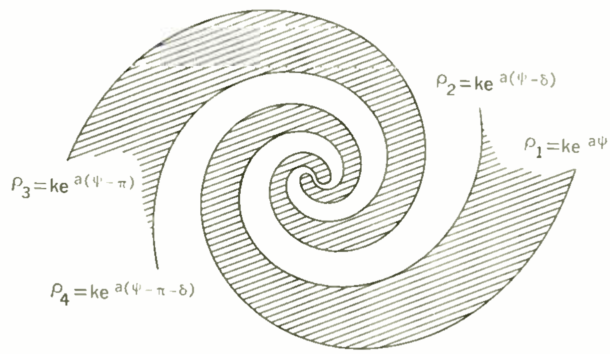
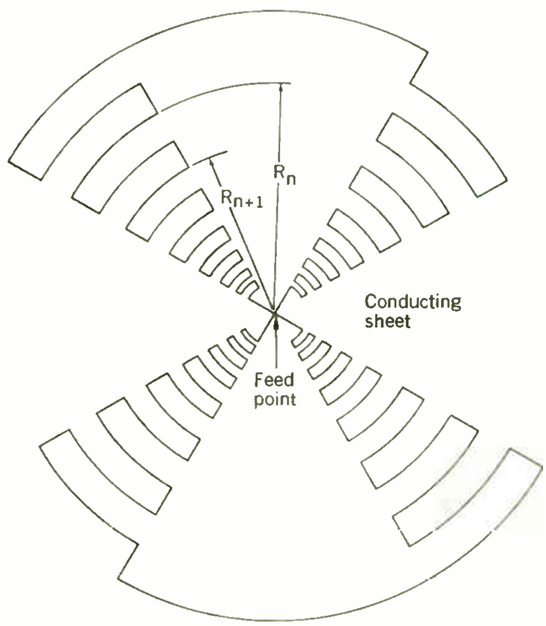



Fig. 4. Sketch showing the geometry of the equiangular (or logarithmic) spiral antenna with equations of the edges.

Fig. 5. A sheet-metal log-periodic antenna.




* The pattern actually rotates with frequency about an axis perpendicular to the plane of the spiral. If the pattern-measuring coordinate system is allowed to rotate at the same rate, the measured pattern remains constant; otherwise there will be a (generally small) periodic variation of magnitude proportional to the rotational asymmetry of the pattern.

cessive teeth. From the principle of modeling it is evident for this structure, extending from zero to infinity and energized at the vertex, that whatever properties it may have at a frequency $f$ will be repeated at all frequencies given by $\tau^n f$, where $n$ is an integer. When plotted on a logarithmic scale, these frequencies are equally spaced with a period equal to the logarithm of $\tau$; hence the name "log-periodic" structure. Log-periodicity guarantees only periodically repeating radiation pattern and impedance. However, for certain types of such structures and for values of $\tau$ not too far from unity, variation of characteristics over a period can be quite small, and an essentially frequency-independent structure results. It is important to note that only a relatively few of the nearly infinite variety of log-periodic structures will make successful broadband antennas in the sense that the impedance and pattern characteristics will remain constant when the structure is truncated to a finite length. It happens that the geometry of Fig. 5 did result in a successful log-periodic antenna.

The antenna of Fig. 5 was designed to have one other rather special property; namely, that the metal cut away from the plane sheet to form the antenna arms has identical shape with the metal that remains. In other words, the complementary slot antenna has the same size and shape as the metallic dipole antenna. Now by an extension of Babinet's principle it is known that complementary-dipole and slot antennas have impedances $Z_d$ and $Z_s$, respectively, related by $Z_d Z_s = (60\pi)^2$. Because the slot antenna and dipole antenna are the *same* (for the geometries chosen) it follows that $Z_d = Z_s = 60\pi \approx 189$ ohms, a result that is independent of frequency. Hence this particular geometry assured constant impedance, although not constant radiation pattern, independently of the other consideration of log-periodic geometry. In view of this use of Babinet's principle in the design of these planar structures, the next step to be taken was a bigger one than might at first appear.

**Unidirectional frequency-independent and log-periodic antennas.** Both the equiangular spiral antenna (Fig. 4) and the log-periodic antenna (Fig. 5) radiate equally on both sides of the plane of the antennas, a result that severely limits their usefulness. A major step forward was made in extending the range of practical application when Isbell[13] bent the two arms of the planar log-periodic structure toward each other (out of the plane) to form the nonplanar V-shaped antenna of Fig. 6. Two rather surprising results were observed. As the angle between the two arms of the antenna was decreased from 180° the radiation pattern changed from bidirectional to undirectional, with the major radiation off the apex of the antenna—that is, in the backward direction. Moreover, although one of the necessary conditions for Babinet's principle (that of a plane surface) was now violated, the impedance continued to remain nearly constant with frequency, but at a different value, which depended upon the angle between the arms. This nonplanar version of the log-periodic structure, radiating a plane-polarized unidirectional beam, greatly increased the utility of the log-periodic structures.

The frequency-independent logarithmic spiral structure also found wider use when Dyson developed a unidirectional version by wrapping the balanced spiral arms on the surface of a cone, as shown in the antenna of Fig.

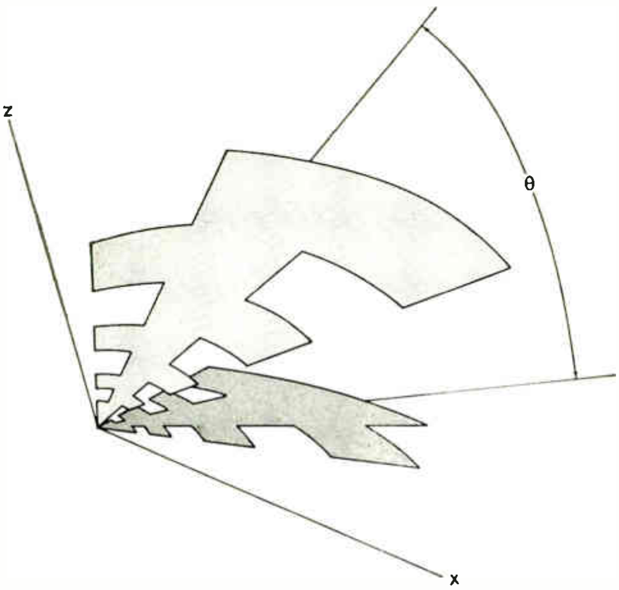
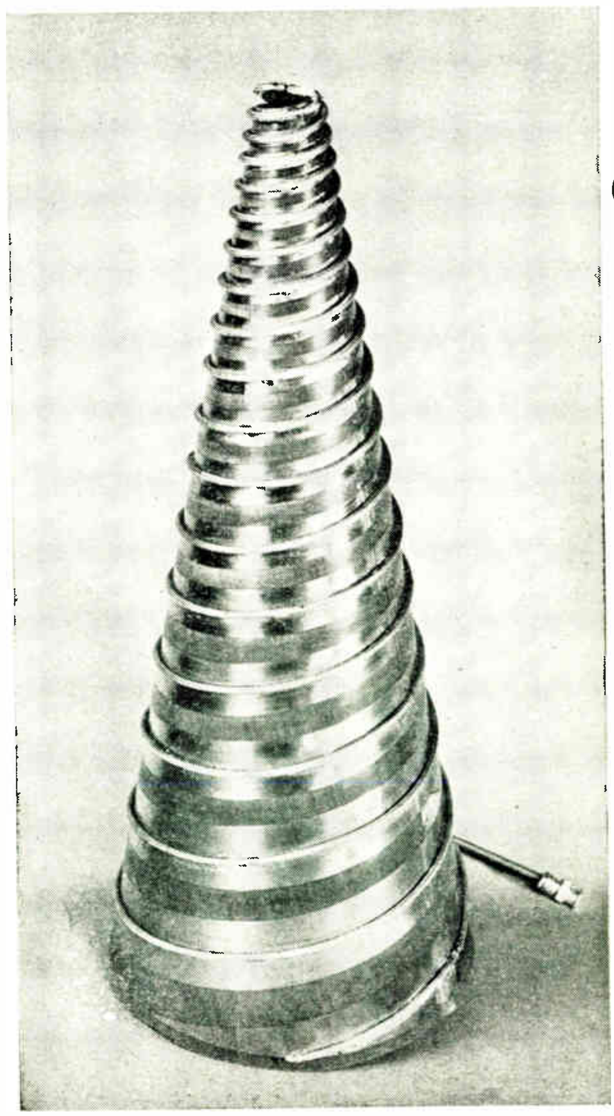



Fig. 6. Nonplanar, unidirectional log-periodic antenna.

Fig. 7. Unidirectional conical spiral antenna.



7. For appropriately chosen rates of spiral this modified version continued to yield essentially frequency-independent performance. For cone angles of less than about 45° the pattern became unidirectional with a broad-lobed beam, again in the backward direction off the apex of the cone.

The conical equiangular spiral antenna is a balanced structure, which may be fed (at the apex) by means of a balanced transmission carried up inside and along the axis of the cone. Alternatively, it may be fed as illustrated in Fig. 7 by a coaxial cable carried along and soldered in contact with one of the arms. Because the amplitude of antenna current on the arms, and also on the outside of the coaxial cable, falls off quite rapidly with distance from the apex, the ends of the arms where the cable enters is essentially a field-free region. This type of feed automatically provides a frequency-independent balun (balanced converter), permitting the balanced antenna to be fed by means of an unbalanced coaxial line. To maintain physical symmetry a dummy cable is usually soldered to the other arm. Conical equiangular or log-spiral antennas have been constructed to operate over bandwidths of higher than 40 to 1. The bandwidth obtained is at the discretion of the designer. The upper usable frequency is determined by the truncated region at the apex, which must remain small in terms of wavelengths, and the lowest usable frequency is set by the base diameter of the cone, which must be at least $^3/_8$ wavelength at the lowest frequency of operation for spirals that are wrapped fairly tightly.

A further modification of the conical equiangular spiral results in a very practical, easily constructed antenna. If the width of the expanding arms is narrowed and they are allowed to degenerate to constant-width structures, the cables alone can form the arms. For fairly tightly spiraled antennas there is little change in the characteristics from those of an antenna with narrow expanding arms.

**Other types of log-periodic antennas.** The practical value of the log-periodic approach was enhanced even further when DuHamel[11] and co-workers demonstrated that successful log-periodic antennas could be made with wire structures as well as sheet structures. This development extended the range of application down from microwaves through the high-frequency band. A typical wire version of a log-periodic antenna is shown in Fig. 8. It was also demonstrated that for higher gain a frequency-independent array of log-periodic antennas could be constructed by arranging the antennas like the spokes of a wheel with the origins of the individual antennas at the hub.

Still another application of the log-periodic principle is the log-periodic dipole array[15] of Fig. 9. As with all log-periodic geometries, all dimensions are increased by a constant ratio in moving outward from the origin. Thus the lengths and spacings of adjacent elements must be related by a constant scale factor $\tau$, as follows:

$$\frac{l_n}{l_{n-1}} = \frac{d_n}{d_{n-1}} = \tau$$

Although at first glance this antenna might appear similar to the early fishbone antenna with $\tau = 1$, there are several essential differences. For successful operation, the log-periodic dipole array must be fed with a transposition of the transmission line between adjacent dipole elements. The antenna is then caused to radiate in the backfire direction (that is, toward the source), a condition which appears to be necessary for successful unidirectional frequency-independent or log-periodic operation. The

Fig. 8. Log-periodic wire antenna for frequencies of 11 to 60 Mc/s. (Photo courtesy Collins Radio Company.)

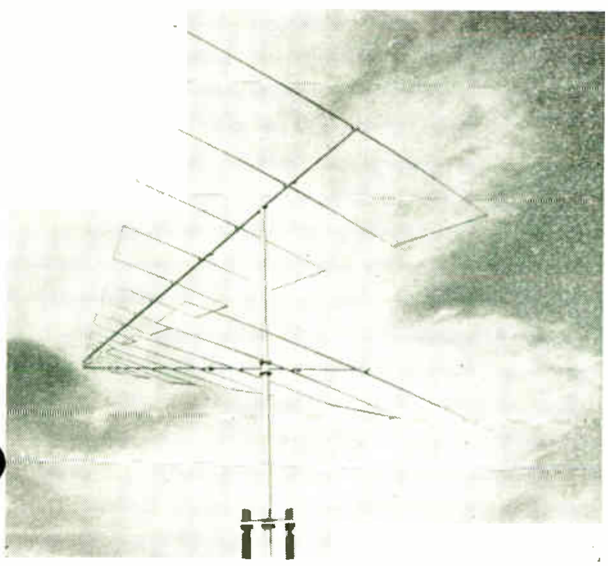
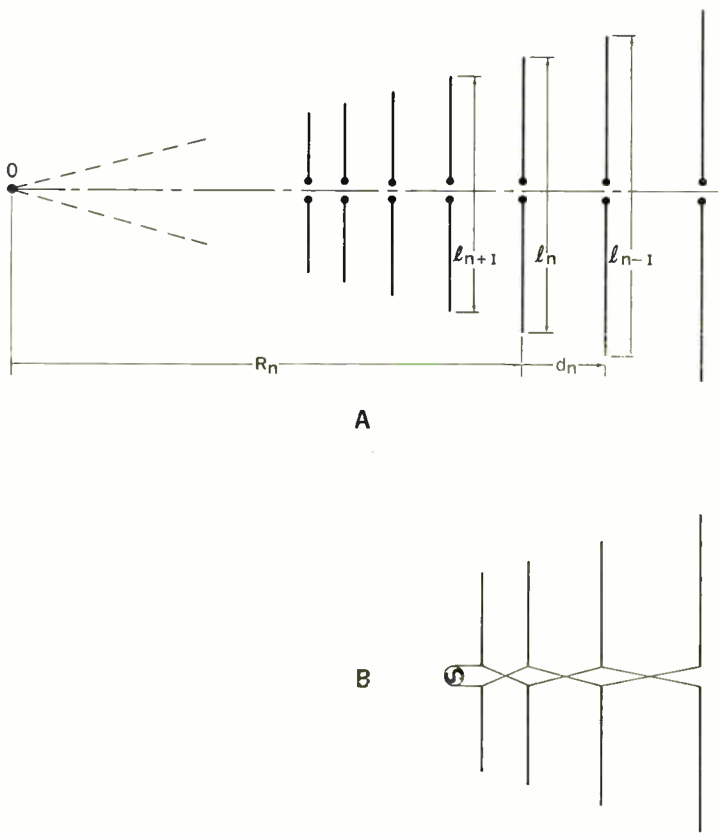


Fig. 9. Log-periodic dipole antenna array. A—Lengths and spacings of elements. B—Method of feeding.



active portion of the array from which most radiation occurs is centered around those elements near resonance (for which $l_n$ is somewhat less than $\lambda/2$). As the frequency is changed the active region moves back or forth along the array. Because practically all of the input power is absorbed in and radiated by the active portion, the larger elements to the right of the active region are not excited. Moreover, because the beam is directed toward the feed point at the left, these larger elements are in an essentially field-free region, and so do not adversely affect the operation. The shorter elements to the left of the active region are in the beam but, because of their short lengths, close spacings, and alternate phasings, have small influence on the pattern.

**Basic principles of operation of log-periodic and frequency-independent antennas.** Of the almost unlimited variety of log-periodic structures that can be devised, only a small fraction will produce successful antennas when truncated. It is interesting to search out the essential requirements for successful design. The operation of the log-periodic dipole array of Fig. 9, being simple and easily understood, will be analyzed in some detail. From the understanding so gained it should be possible to extend the analysis to less familiar geometries, and then to frequency-independent antennas in general.

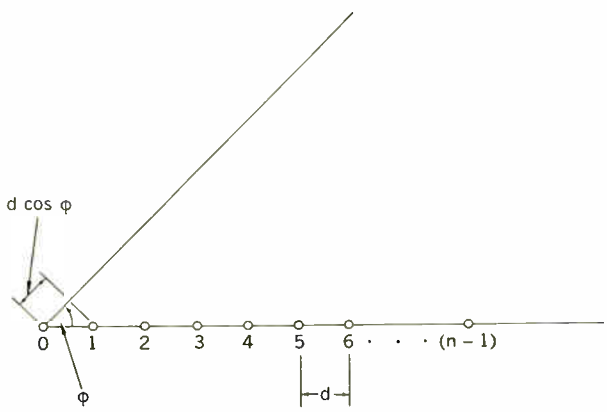



Fig. 10. Array of equispaced isotropic radiators.

At this point it will be advantageous to recall some of the basic notions of antenna array theory. Consider an $n$-element array of equispaced isotropic radiators (Fig. 10) having equal current amplitudes and a spacing $d$ less than one-half wavelength. (An isotropic radiator is one that radiates uniformly in all directions; a simple dipole antenna is an isotropic radiator in the H plane perpendicular to its axis.) At a distant point the electric fields from these radiators will add with a phase angle between them which is dependent upon the relative phasings of the radiator currents and the relative phase delays produced by the difference in path lengths to the distant point. For the array shown the phase difference due to path length difference between adjacent elements is $(2\pi/\lambda)d\cos\phi$ radians. If the elements of the array are fed with a progressive phasing of currents equal to $\alpha$, where $\alpha$ represents the angle by which the current in a given element *leads* the current in the preceding element, then at the distant receiving point the phase difference of the fields produced by adjacent elements will be

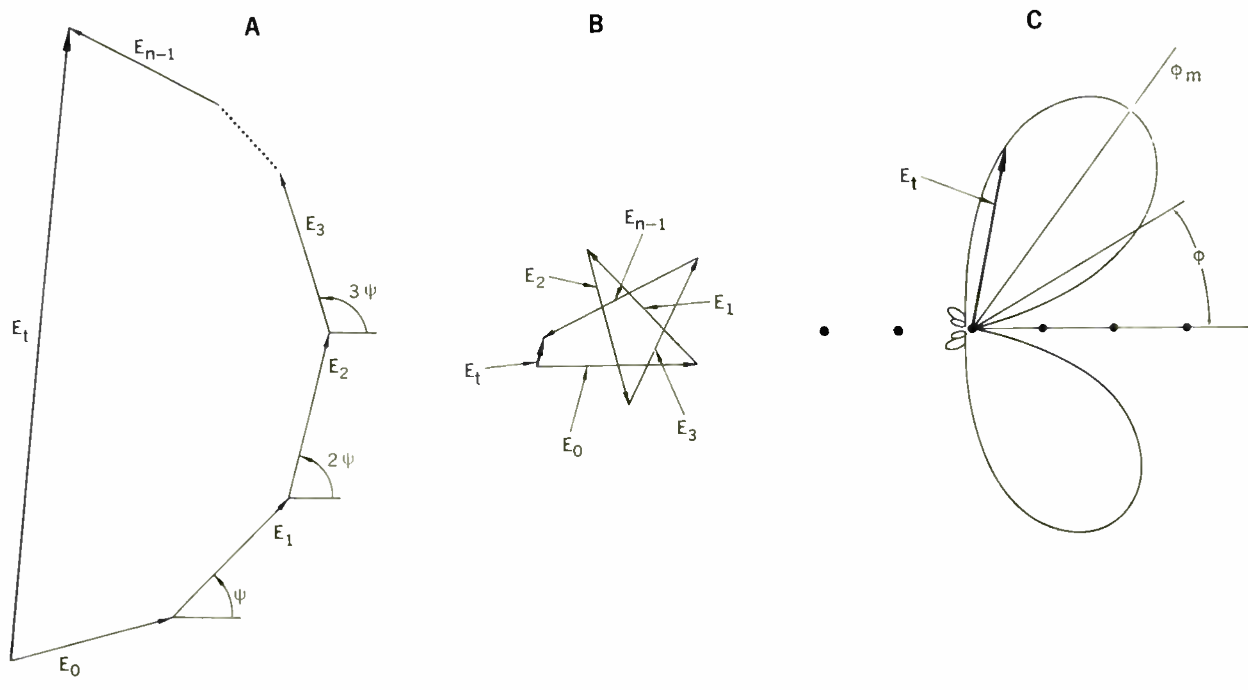
$$\psi = \alpha + \frac{2\pi}{\lambda} d\cos\phi = \alpha + kd\cos\phi \qquad (1)$$

where $k = 2\pi/\lambda$ is the free-space phase-shift constant. The total electric field at any distant point will be given by the phasor sum

$$E_t = E_0 1 + e^{j\psi} + e^{j2\psi} + \ldots + e^{j(n-1)\psi} \qquad (2)$$

where $E_0$ is the field intensity at the reception point produced by current $I_0$. $E_t$ can be obtained graphically

Fig. 11. A and B—Graphical construction for $E_t$. C—Calculated radiation pattern, $E_t$ vs. $\phi$.



from the construction of Fig. 11(A). Using the particular value of $\alpha$, and computing $\psi$ from (1) for various values of $\phi$, the construction of Fig. 11(A) can be used to determine a radiation pattern of the array; see Fig. 11(C). It is evident that the total field intensity will be maximum when $\psi = 0$, so that all fields add in phase.

Therefore, for a maximum, $\psi = \alpha + kd \cos \phi = 0$. The angle $\phi_m$ for maximum radiation is given by

$$\cos \phi_m = -\frac{\alpha}{kd} \quad \text{or} \quad \phi_m = \cos^{-1}\frac{-\alpha}{kd} \tag{3}$$

If the elements are fed in phase, $\alpha = 0$, and $\phi_m = 90°$, so the maximum radiation is broadside. If successive elements are fed with a lagging phase of value, $\alpha = -kd$, than $\phi_m = 0$, so the maximum radiation is endfire in the forward direction. If successive elements are fed with a leading phase of value, $\alpha = +kd$, than $\phi_m$ will equal 180°, and the maximum radiation will be endfire in the backward direction. For values of $\alpha$ between $-kd$ and $+kd$, the angle of maximum radiation is at an angle between 0 and 180° as given by Eq. (3). By symmetry about the axis of the array, there is another maximum at an angle between 0 and −180°, which is also given by (3). When $|\alpha| > kd$, Eq. (3) cannot be satisfied for any real value of $\phi$; that is, there is no value of $\phi$ in the "visible" range between 0 and 180° (hence, also between 180° and 360°) that will produce a maximum—in the sense that all the radiations add in phase. However, if $|\alpha|$ is only slightly greater than $kd$, so that $\psi$ is not much larger than zero, the total field can still be quite strong in the forward direction ($\phi = 0$) for negative $\alpha$, or in the backfire direction ($\phi = 180°$) for positive $\alpha$. This case is illustrated by the sketch of Fig. 11(A). On the other hand, if $|\alpha|$ is considerably greater than $kd$ (that is, the phase shift between elements is large), the phase diagram might be as illustrated in Fig. 11(B), with a resulting small total $E_t$ for all values of $\phi$.

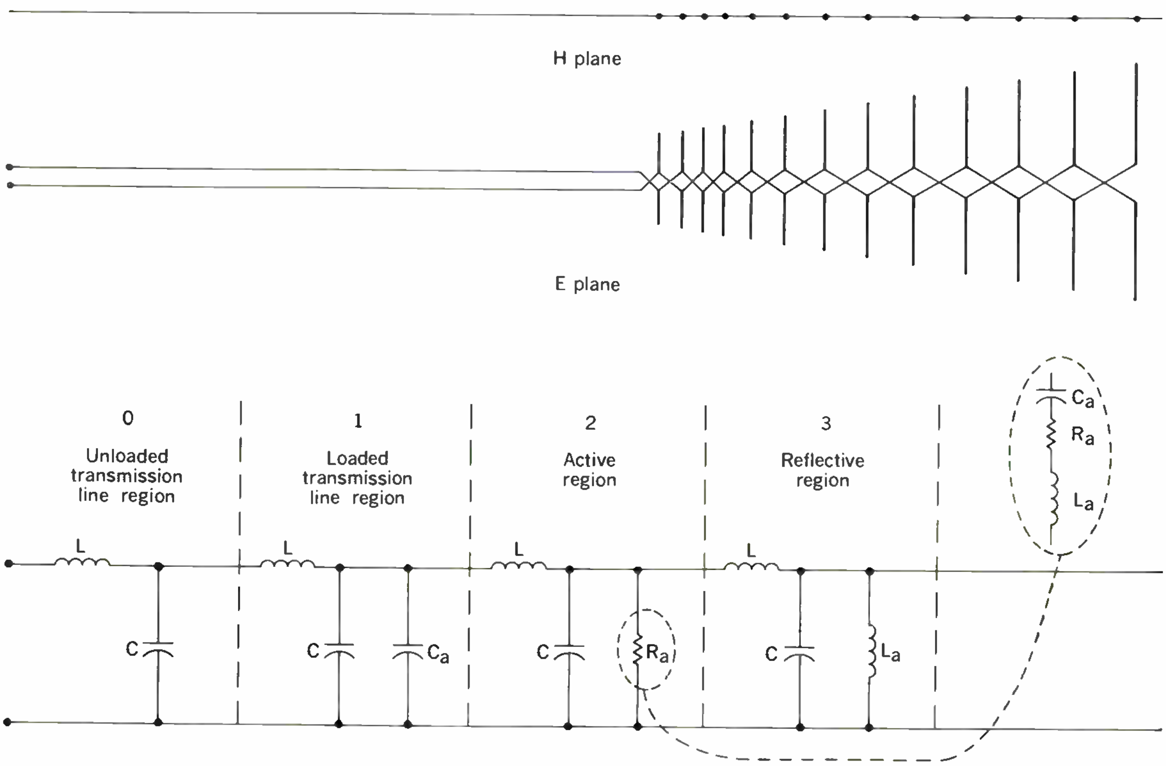
For these cases of large phase shift, as shown in Fig. 11(B), there is no major lobe anywhere, and the array radiates only feebly, scattering its small radiated energy in various directions.

The elementary notions just discussed can be applied with some slight modification to an analysis of the log-periodic dipole array sketched in Fig. 12. For this purpose, it is helpful to consider separately three main regions of the array.

1. *Transmission-line region.* The antenna elements in the transmission-line region are short compared with the resonant length (that is, $l \ll \lambda/2$), so the element presents a relatively high capacitive impedance. The element current is small and leads the base voltage supplied by the transmission line by approximately 90°. The element spacing is small in wavelengths and the phase reversal introduced by transposition of the transmission line means that adjacent elements are nearly 180° out of phase. More precisely, each element current leads the preceding element current approximately by $\alpha = \pi - \beta d$, where $d$ is the element separation and $\beta = 2\pi/\lambda = \omega/v$ is the phase-shift constant along the line. In general $\beta$, $\lambda$, and $v$ will differ from their free-space values owing to the loading effect of the elements on the transmission lines. Because of the phasing and close spacing of the elements, radiation from this region will be very small and in the backfire direction.

2. *Active region.* In the active region the element lengths approach the resonant length ($l$ slightly less than

Fig. 12. Transmission-line representation of log-periodic dipole array.



λ/2), so the element impedance has an appreciable resistive component. The element current is large and more nearly in phase with the base voltage; the current is slightly leading just below resonance and slightly lagging just above resonance. The element spacing is now sufficiently large to allow the phase of current in a given element to lead that in the preceding element by an angle $\alpha = \pi - \beta d$, which may approximate $\pi/2$ radians. This combination of conditions will produce a strong radiation in the backfire direction.

3. *Reflection region.* The element lengths in the reflection region are greater than the resonant length ($l \geq \lambda/2$), so the element impedance becomes inductive and the element current lags the base voltage. The base voltage provided by the transmission line is now quite small, because in a properly designed array nearly all of the energy transmitted down the line has been abstracted and radiated by the active region. The element spacing may now be larger than λ/4. However, as will be shown later, the phase shift per unit length along the line in this region is small, so the resulting phasing between elements (including the phase reversal introduced by the transposition) is such that any small amount of radiation is still in the backfire direction. In addition, it will be demonstrated later that the characteristic impedance of the transmission-line becomes reactive in this region. Thus, any small amount of incident energy transmitted through the active region is not accepted in the reflection region but is reflected back toward the source.

**The array as a loaded transmission line.** Some of the remarkable properties of log-periodic and frequency-independent antennas are attributable to the propagation characteristics of the equivalent loaded transmission line that conveys energy from the source to the radiating portion of the antenna. These effects are particularly easy to see in the case of the log-periodic dipole array, shown in Fig. 12. On the feed line to the antenna, region 0, the series inductance and shunt capacitance per unit length are shown as $L$ and $C$, respectively. In the transmission region of the antenna, region 1, the transmission line is loaded by a capacitance per unit length $C_a$ that represents the loading effect of the short dipoles, which have a capacitance reactance. It is noted that to the first approximation $C_a$ is nearly constant throughout this region because at the beginning of the region the capacitance per element is small, but the elements are closely spaced, whereas near the end of the region the capacitance per element is larger, but so is the spacing. The effect of the augmented shunt capacitance of the line ($C + C_a$) is to increase the phase delay per unit length, and since $\beta = 2\pi/\lambda = \omega/v$, this means a decrease of wavelength λ and a decrease of phase velocity $v$ along the line below the free-space values. This is said to be a "slow wave" region of the transmission line. Note, however, that because of the transposition of the feed line between elements, successive elements are fed with a leading phase shift of $\pi - \alpha$ per section. This rapid phase shift in the reverse direction corresponds to a slow wave in the backward direction along the antenna elements.

In region 2, the element lengths approach the resonant length and the transmission line loading becomes resistive, designated by the shunt resistance $R_a$ in series with the antenna capacitance $C_a$ and antenna inductance $L_a$. The phase shift per unit length, the wavelength, and the phase velocity all approach their free-space values. Because of the transposition between elements, and accounting for the fact that the element current leads the base voltage by lesser amounts in successive elements as the resonant length is approached, it turns out that phasing of currents in the elements corresponds to a backward traveling wave having a velocity $v$ somewhat less than $c$, the velocity of light.

In region 3 the element lengths become longer than the resonant length, the antenna inductive reactance pre-

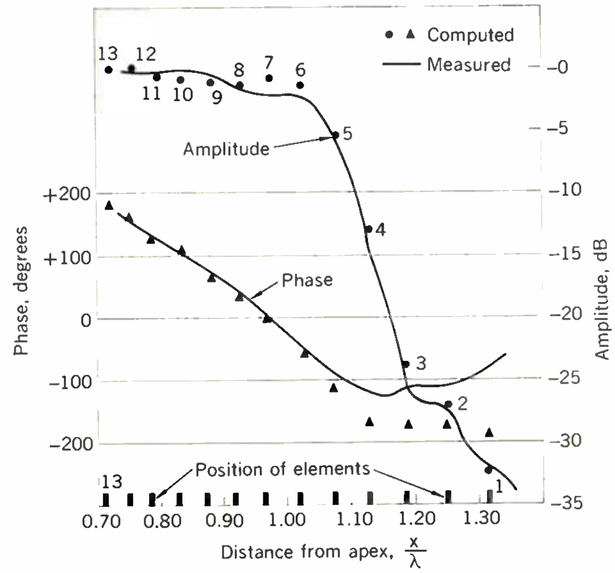
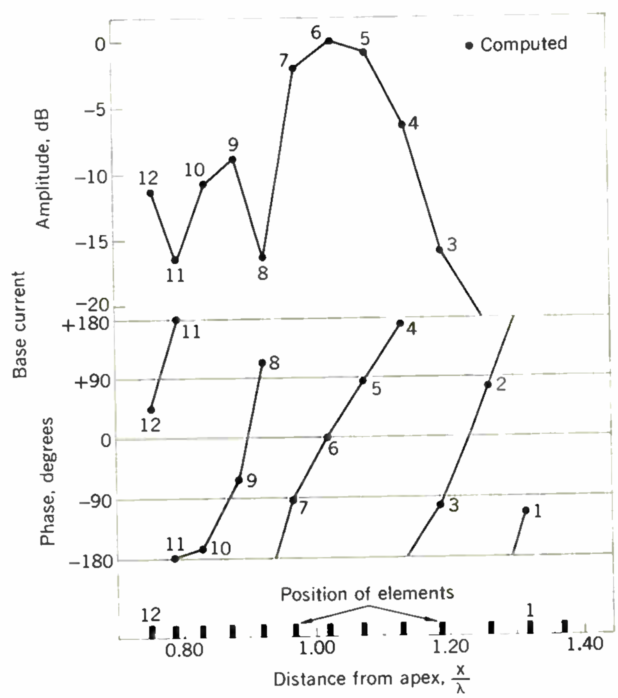



Fig. 13. Base voltages along a typical 13-element log-periodic dipole array at a frequency for which element 4 is half of a wavelength.

Fig. 14. Element currents corresponding to Fig. 13.



dominates, and the loading effect on the line is represented by the shunt inductance $L_a$. If the parallel combination of $L_a$ and $C$ is inductive, we have the equivalent of the attenuation region of a filter. The phase shift per unit length is then zero (for the lossless case) and the phase velocity is infinite; that is, there is no wave motion. The incident energy propagating down the line is no longer accepted but is reflected back toward the source. These results are strictly true only in the case of a lossless filter, but they form the first approximation in the case of a lossy filter.

The general features outlined in the foregoing discussion will be illustrated for a particular log-periodic dipole array, which has been analyzed in considerable detail.[16] Fig. 13 shows the amplitude and phase of the transmission line voltage along a particular 13-element log-periodic dipole array. Distance is shown measured from the apex of the array, and the elements are numbered starting with the largest element as number 1. This set of data is for a frequency $f$ for which element number 4 is $\lambda/2$ long. Several interesting aspects of the data are immediately apparent: In the transmission region (elements 13 to 7), the amplitude of voltage along the line is approximately constant and the phase shift between element positions increases gradually from about 20° to 30°. (Because of the transposition between elements, this means that adjacent elements are fed with a progressive phase lead 160° to 150°.) In the active region (elements 7 to 4) the amplitude drops sharply because of power absorbed by the strongly radiating elements, and the phase shift averages about 90° between adjacent elements. Finally, in the unexcited or reflection region (elements 3 to 1), the amplitude drops to very low values and the phase shift between element positions is nearly zero (corresponding to the zero phase shift or infinite phase velocity in the attenuation region of a low-pass filter).

The resulting element currents for the log-periodic dipole array of Fig. 13 are shown in Fig. 14, both in amplitude and phase. From the current amplitudes (noting that small contributions from elements 12 through 8 tend to cancel one another because of the nearly 180° phase shift between them), it is evident that the only elements that will contribute appreciably to the radiation are elements 7, 6, 5, and 4. For these elements, the phase difference between adjacent members is approximately 90° leading, so a backfire radiation will be expected. The phasor diagrams for $\phi = 0°, 90°$, and 180° are shown in Fig. 15 and the resulting radiation patterns are shown in Fig. 16. (The E-plane pattern is the H-plane pattern modified by the directivity of the individual elements in this plane.)

As operating frequency is decreased or increased the active region moves up or down the array, but radiation pattern and input impedance remain almost constant.

**General properties of log-periodic and frequency-independent antennas.** The manner of operation of the log-periodic dipole array has been described in some detail because of the insight it gives into what are believed to be general requirements for successful frequency-independent operation. These appear to be as follows:

1. An excitation of the antenna or array from the high-frequency or small end of the antenna.

2. A backfire radiation (in the case of unidirectional radiators), so that the antenna fires through the small

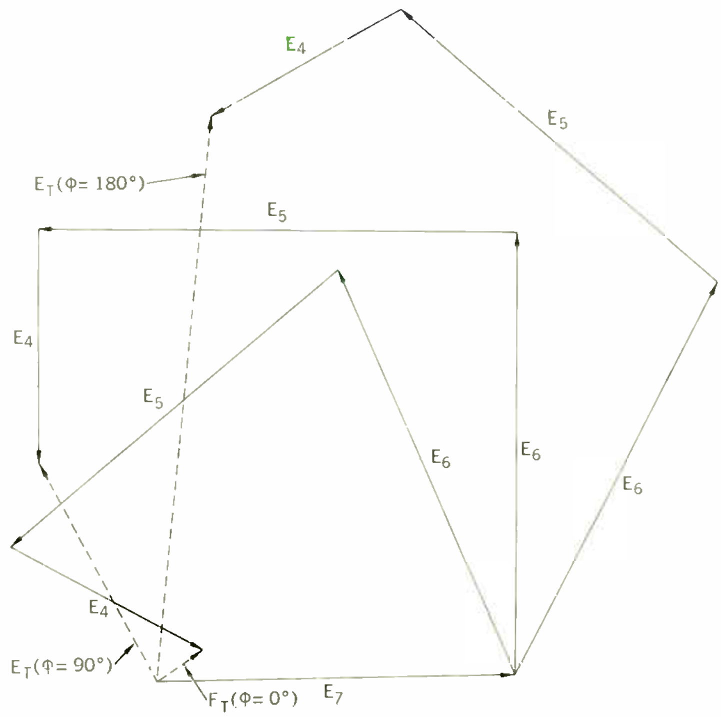
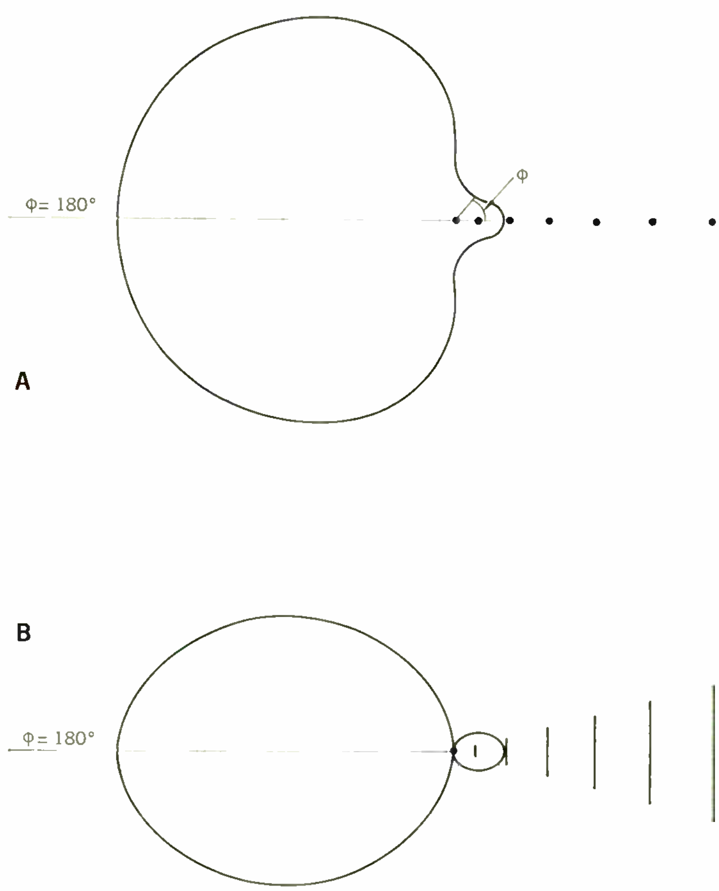



Fig. 15. Phasor sum of radiated fields from currents shown in Fig. 14, for $\phi$ values of 0°, 90°, and 180°.

Fig. 16. Radiation patterns resulting from fields shown in Fig. 15. A—H-plane pattern. B—E-plane pattern.



part of the antenna, with the radiation in the forward direction being zero or at least very small. For bidirectional antennas the backfire requirement is replaced by a requirement for broadside radiation. In any case, the radiation in the forward direction along the surface of the antenna (which theoretically extends to infinity) must be zero or very small.

3. A transmission region formed by the inactive portion of the antenna between the feed point and the active region. This transmission line region should have the proper characteristic impedance and negligible radiation.

4. An active region from which the antenna radiates strongly because of a proper combination of current magnitudes and phasings. The position and phasing of these radiating currents are such as to produce a very small radiation field along the surface of the antenna or array in the forward direction, and a maximum radiation field in the backward direction (broadside for bidirectional antennas). For successful backfire antennas these requirements are frequently met with separations less than a quarter wavelength and phasings near 90° leading, for adjacent elements in the active region. For broadside radiation the phasings must, of course, be zero.

5. An inactive or reflection region beyond the active region. All successful frequency-independent antennas must exhibit a rapid decay of current within and beyond the active region, so that operation will not be affected by truncation of the structure. A major cause of the rapid current decay is, of course, the large radiation of energy from the active region. An additional cause, in at least some types of frequency-independent and log-periodic antennas, is the attenuation resulting from the rejection of incident energy by the reflection region (the filter stop-band effect mentioned previously). The prevalence and importance of this latter filter action are still uncertain.

Finally, two other observations may be made. Although we have tended to think of the structures of Figs. 4, 5, 7, and 8 as single antennas and the structure of Fig. 9 as an antenna array, it appears that most frequency-independent and log-periodic antennas may be thought of as antenna arrays, with the array factor playing an important role in the formation of a proper endfire or broadside pattern. The localization of the individual radiating elements may be easier to see for the case of the log-periodic dipole array of Fig. 9, but the array action can also be observed in the other cases; it is particularly evident in the case of the fairly tightly wrapped conical log-spiral.

The second observation relates to the similarity between antennas derived from the angle concept and log-periodic concept.[17] Both lead to a solution of the unlimited-bandwidth problem and for this reason both have come to be known as frequency independent.

An example of the similarity between these two antenna types can be demonstrated in the case of the log-periodic wire antenna of Fig. 8, which produces a linearly polarized beam off the apex with the electric vector parallel to the transverse elements. If two such antennas are arranged in space quadrature along a common axis, and with a common origin but with one structure scaled a quarter period from the other, the resultant combination produces a circularly polarized beam with a pattern that rotates about the axis with frequency, exactly as in the case of the conical equiangular spiral antenna. Conversely, of course, if the pattern of a conical equiangular spiral is probed with a linear receiving antenna of fixed plane of polarization, the measured pattern will vary log-periodically with frequency, as does the pattern of the antenna of Fig. 8.

In addition, it is pertinent to note that if a narrow-armed conical equiangular spiral (an angle structure) is flattened sideways (along the axis), it becomes a log-periodic zigzag antenna.

**Recent developments**

The log-periodic and angle concepts have been used to generate many highly useful antennas of large bandwidth. Fig. 17 shows a very practical two-element array of log-periodic dipole arrays capable of maintaining a nearly constant radiation pattern and a 50-ohm input impedance over the frequency range from 450 to 2000 Mc/s. The 50-ohm input impedance results from feeding two 100-ohm arrays in parallel. Although the dipole array is a balanced structure, it can be fed as shown with a coaxial cable running up the inside of one of the hollow transmission lines, utilizing the frequency-independent balun effect previously noted in connection with the conical log-spiral antenna.

Fig. 17. Two-element array of log-periodic dipole arrays.

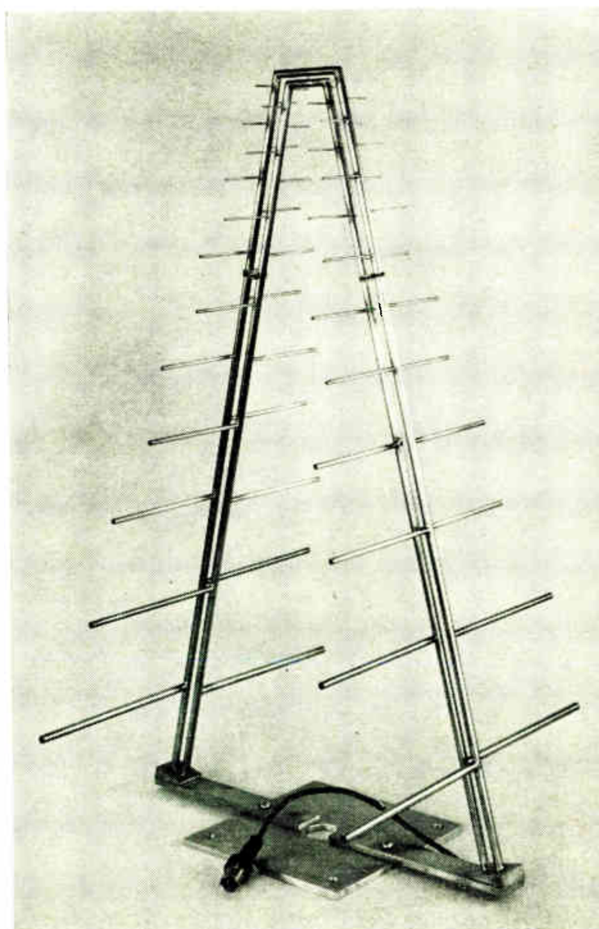


Fig. 18 shows one version of the log-periodic resonant-V developed by Mayes and Carrel.[18,19] This antenna was designed to overcome one of the major shortcomings of the ordinary log-periodic dipole array—namely, the long physical length of array required to cover a very wide band of frequencies. The antenna of Fig. 18 is designed to operate in several modes. In the lowest order $\lambda/2$ mode, the operation is similar to that of the log-periodic dipole array because the forward tilt of the elements has small effect for this mode. However, as the frequency of operation is increased beyond that at which the shortest elements are resonant—that is, when the active region runs off the front end of the array—the largest elements at the rear become active in the $3\lambda/2$ resonance mode. In this mode the forward tilting of the elements ensures a good unidirectional pattern of high directivity. As the frequency is further increased, the active region moves forward through the array in the $3\lambda/2$ mode until once again it runs off the front end, to return to the rear in the $5\lambda/2$ mode. This scheme makes it possible to obtain large bandwidths of the order of 20 to 1 with a relatively compact array. The pattern and impedance characteristics remain good over the entire frequency spectrum except for intervals about the mode-transition frequencies. Based on these principles, arrays have been designed to cover all of the television channels from 2 through 83, corresponding to a frequency range from 54 to 890 Mc/s.

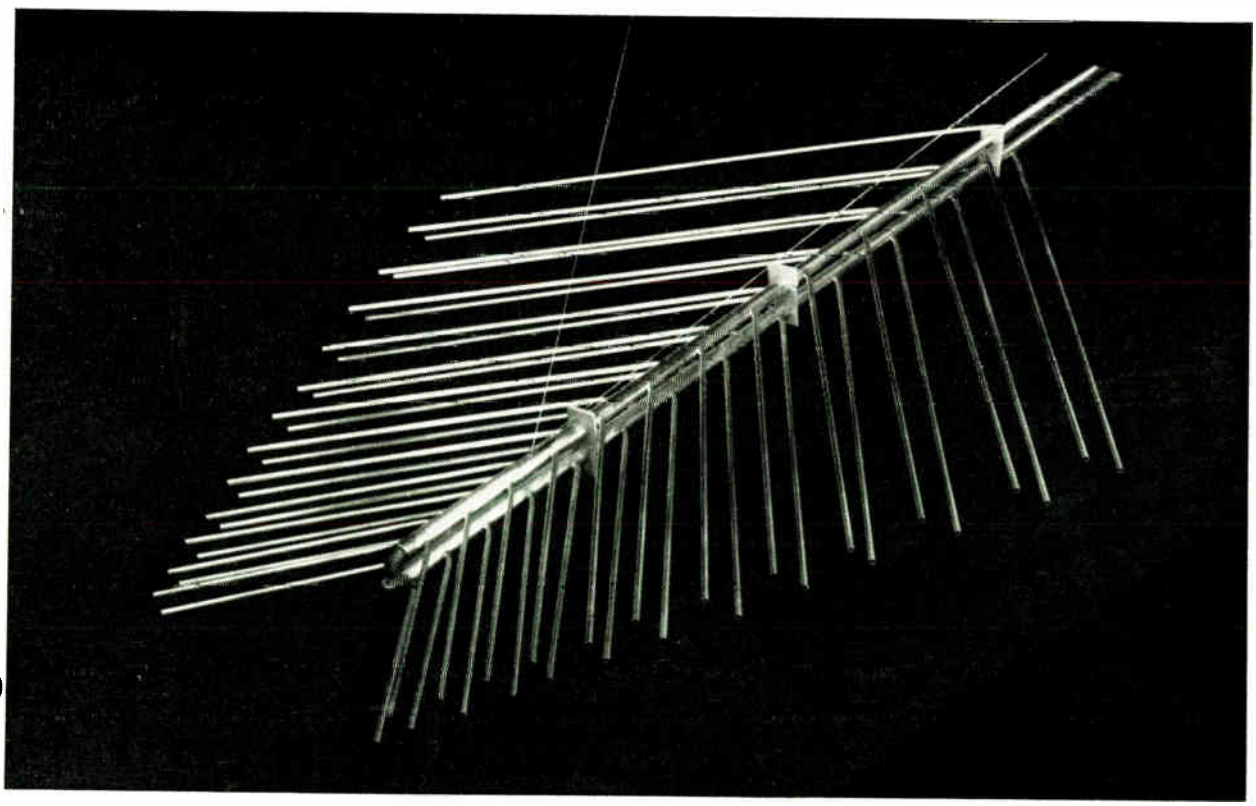
Another interesting development is that of a log-periodic folded-dipole array. At first thought it would appear that such an array could not work because the short elements at the front of the array present a very low impedance, thus short-circuiting the transmission region leading to the active region. This difficulty is circumvented[20] by connecting the folded dipoles in series with the transmission line, rather than in shunt, and recognizing that the active region will occur near first resonance, that is near the element length ($\lambda/4 < l < \lambda/2$) where the capacitive reactance of the short antenna resonates with the inductive reactance of the folded dipole viewed as a short-circuited transmission line. This unusual operating mode for the folded dipole results in a shorter element length for resonance, and consequently a narrow width for the resulting folded-dipole array.

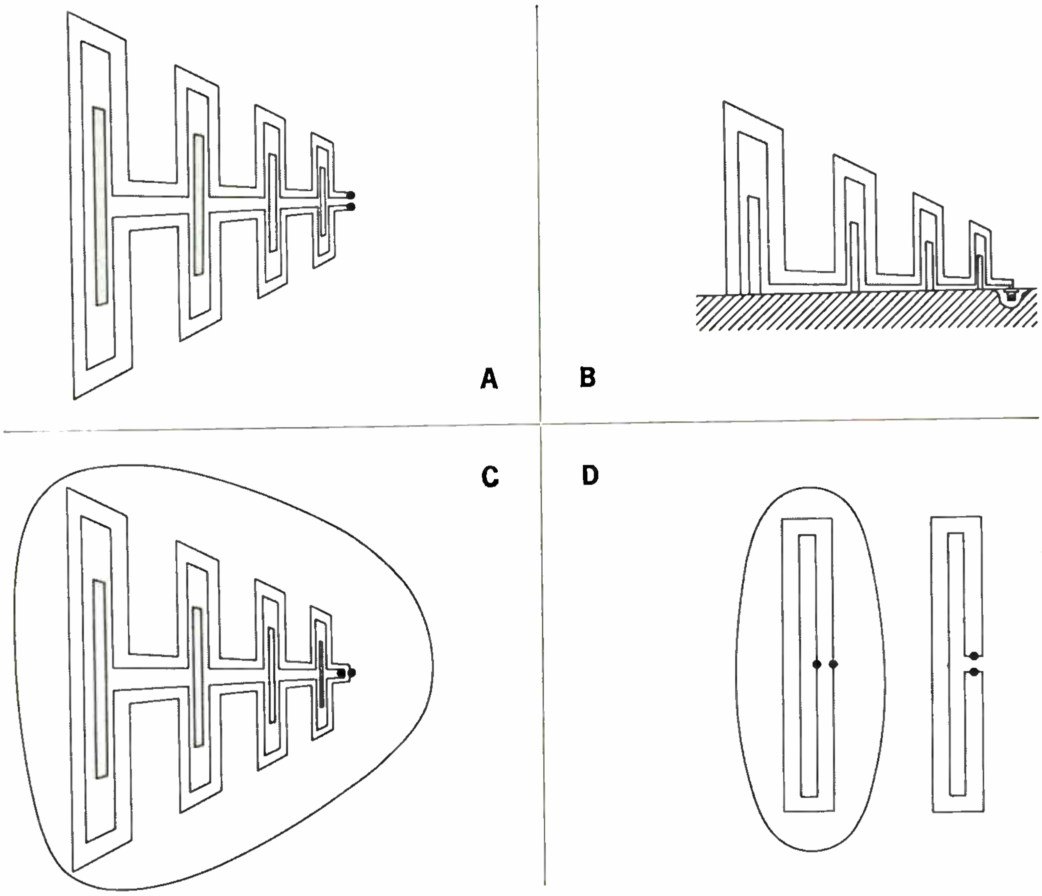
A major problem with log-periodic structures has been the design of an antenna that will operate successfully when fed against a ground plane to produce vertical polarization. One half of the antenna of Fig. 8 can be operated over ground to produce horizontal polarization, as can an inclined horizontal log-periodic dipole array. For vertical polarization, particularly in the high-frequency band (3–30 Mc/s), it is desirable to use the equivalent of a log-periodic monopole array that has a height of only approximately $\lambda/4$ at the lowest operating frequency, rather than $\lambda/2$. Because of the necessity for introducing a transposition between elements (or otherwise producing the required phase difference between elements) it is not possible simply to use one half of a log-periodic dipole array fed against ground.

Several solutions to this problem, having varying degrees of success for different applications, have been developed by a number of workers in the field.[21–23] A quite recent development[24] using folded monopoles with added phasing elements promises to be very useful.

Three versions of this antenna are shown in Fig. 19.

Fig. 18. Log-periodic resonant-V array, for operation in several modes.




A
B
C
D

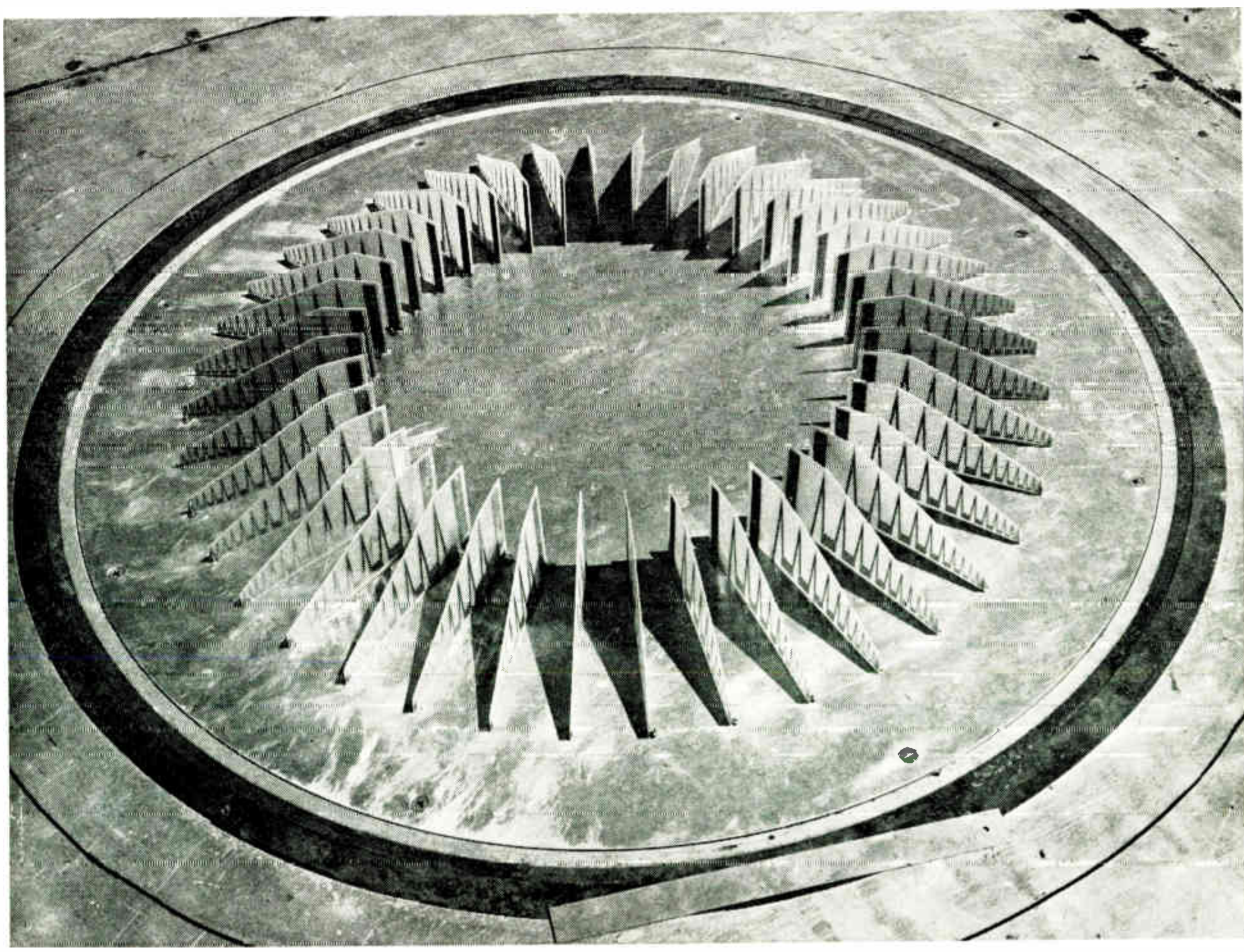


Fig. 19. Log-periodic arrays of folded elements. A—Log-periodic folded-dipole array. B—Log-periodic folded-monopole array. C—Log-periodic folded-slot array. D—Duals: folded slot and folded dipole.

Fig. 20. Model of wide-aperture log-periodic array for high-frequency radio direction finding (3–30 Mc/s).

The log-periodic folded-slot array (A) was conceived first, but by duality, the log-periodic folded-dipole array (B) is obtained automatically. Because this array possesses the proper image symmetry about the horizontal axis (horizontal currents in opposite directions, vertical currents in the same direction), one half of the array can be fed against a ground plane to produce the folded monopole array of (C). The duals, folded slot and folded dipole, are illustrated in (D). The dimensions of the phasing slots in (A), or phasing strips in (B) and (C), are adjusted experimentally to provide the required phasing between successive dipoles or monopoles to produce a good backfire beam.

For greater directivity than can be achieved with a single frequency-independent antenna (or array) it is possible to use the frequency-independent structure as the broadband feed of a large paraboloid. Although the resultant combination is no longer frequency independent, high-gain antennas having a usable bandwidth as high as ten to one have been built by use of this approach.

Some of the high-gain paraboloid tracking antennas for the Atlantic Missile Range have been modified to use two conical log-spiral antennas as a circularly polarized broadband feed in a conical scan system. This application covers a frequency band of 215 to 1000 Mc/s, but the feed elements themselves are capable of operating continuously to 2300 Mc/s.

An alternative approach to the high-gain broadband problem is illustrated in the model of a broadband (3–30 Mc/s) wide-aperture radio-direction-finding array shown in Fig. 20. For frequency-independent arraying, the individual elements should lie along radials and be arranged to fire inward toward the common origin (toward the hub of the wheel). Unfortunately this arrangement requires opposite elements to fire through each other, and severe pattern deterioration results. In the array of Fig. 20 the log-periodic antennas fire outward. A 100° sector of elements is connected together through an appropriate phasing network and rotating switch or goniometer to form a narrow beam, which rotates with the goniometer as the latter connects in elements on one side of the sector and disconnects them on the other side. Again, this arrangement is far from being frequency independent, but the use of broad-band log-periodic structures as array elements is an improvement over the earlier use of frequency-sensitive elements.

This last example indicates that although truly remarkable progress has been made in the past decade in achieving broadband antenna operation there still remain some challenging problems for the future. Among these challenges are the design of broadband antennas having very high gain, and the design of frequency-independent antennas to produce specified radiation patterns.

REFERENCES

1. Kandoian, A. G., "Three New Antenna Types and Their Applications," *Proc. IRE*, vol. 34, Feb. 1946, pp. 70W–77W.

2. Nash, C. C., and Martin Chase, "Antennas for High Speed Aircraft," Rept. TR-5, Univ. of Illinois Antenna Lab., Urbana, USAF Contract W33-038-AC20778, Feb. 1949.

3. Kraus, J. D., "Helical Beam Antenna for Wide-Band Applications," *Proc. IRE*, vol. 36, Oct. 1948, pp. 1236–1242.

4. Springer, P. S., "End-Loaded and Expanding Helices as Broadband Circularly Polarized Radiators," *Proc. Nat'l Electronics Conf.*, vol. 5, 1949, pp. 161–171.

5. Chatterjee, J. S., "Radiation Field of a Conical Helix," *J. Appl. Phys.*, vol. 24, May 1953, p. 550.

6. Chatterjee, J. S., "Radiation Characteristics of a Conical Helix of Low Pitch Angles," *Ibid.*, vol. 26, Mar. 1955, pp. 331–335.

7. Turner, E. M., "Spiral Slot Antenna," Tech. Note WCLR 55-8, Wright Air Development Center, Dayton, Ohio, June 1955.

8. Rumsey, V. H., "Frequency Independent Antennas," *IRE Nat'l Convention Record*, pt. I, 1957, pp. 114–118; also Rept. TR-20, Univ. of Illinois Antenna Lab., Contract AF33(616)-3220, Oct. 1957.

9. Dyson, J. D., "The Equiangular Spiral Antenna," *Proc. Fifth Symp. on USAF Antenna Res. and Dev. Program*, Allerton Park, Monticello, Ill., Oct. 1955 (classified).

10. Dyson, J. D., "The Equiangular Spiral Antenna," Rept. TR-21, Dept. of Elec. Eng., Univ. of Illinois, Contract AF33 (616)-3220, Sept. 15, 1957.

11. Dyson, J. D., "The Equiangular Spiral Antenna," *IRE Trans. on Antennas and Propagation*, vol. AP-7, Apr. 1959, pp. 181–187.

12. DuHamel, R. H., and D. E. Isbell, "Broadband Logarithmically Periodic Antenna Structures," *IRE 1957 Nat'l Convention Record*, pt. I, pp. 119–128.

13. Isbell, D. E., "Non-planar Logarithmically Periodic Antenna Structure," Rept. No. 30, Univ. of Illinois Antenna Lab., Feb. 1958.

14. DuHamel, R. H., and F. R. Ore, "Logarithmically Periodic Antenna Designs," *IRE 1958 Nat'l Convention Record*, pt. I, pp. 139–151.

15. Isbell, D. E., "Log-Periodic Dipole Arrays," *IRE Trans. on Antennas and Propagation*, vol. AP-8, May 1960, pp. 260–267.

16. Carrel, R. L., "Analysis and Design of the Log-Periodic Dipole Antenna," Rept. No. 52, Univ. of Illinois Antenna Lab., Contract AF33(616)-6079.

17. Deschamps, G. A., and R. H. DuHamel, "Frequency Independent Antennas," Chap. 18 in *Antenna Engineering Handbook*, Henry Jasik, ed. New York: McGraw-Hill Book Co., Inc., 1960.

18. Mayes, P. E., and R. L. Carrel, "Log-Periodic Resonant-V Arrays," presented at Wescon, San Francisco, Calif., Aug. 22–25, 1961.

19. Mayes, P. E., "Broadband Backward Wave Antennas," *Microwave J.*, vol. 6, no. 1, Jan. 1963, p. 2.

20. Balmain, K. G., and J. D. Dyson, "The Series-Fed, Log-Periodic Folded Dipole Array," *Digest IEEE PTG-AP Internat'l Symp.*, Boulder, Colo., 1963, pp. 143–148.

21. Ore, F. R., "Investigation of the Log-Periodic Coaxial Fed Monopole Array," Tech. Note No. 1, Univ. of Illinois Radiolocation Research Lab., Contract NOBSR 85243, Oct. 1963.

22. Berry, D. G., and F. R. Ore, "Log-Periodic Monopole Array," *IRE 1961 Internat'l Convention Record*, pt. I, pp. 76–85.

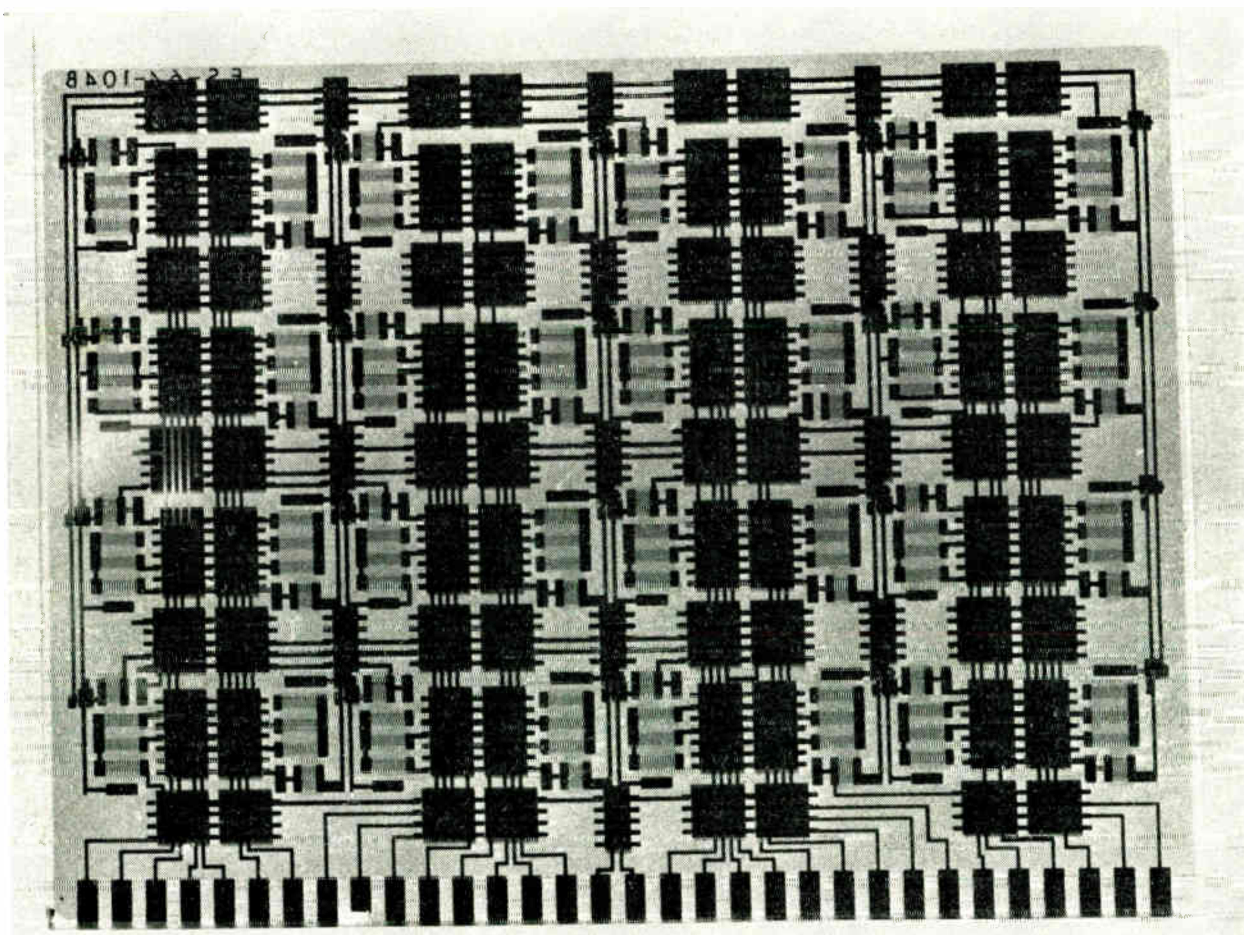
23. Wickersham, A. F., R. F. Franks, and R. L. Bell, "Further Developments in Tapered Ladder Antennas," *Proc. IRE*, vol. 49, Jan. 1961, p. 378.

24. Greiser, John, "Research on Log-Periodic Arrays of Slots," Tech. Rept. No. 3, Univ. of Illinois Antenna Lab., Contract NOBSR 85243, Feb. 1964.

The original work on the logarithmic spiral and logarithmic periodic antennas at the University of Illinois was made possible by the sponsorship of the Aeronautical Systems Division, Wright Air Development Division, U.S. Air Force. Current work is being supported by Wright Air Development Division, Wright-Patterson Air Force Base, The Naval Electronics Laboratory, and the U.S. Army Signal Supply Agency.

# Thin-film circuit technology

*Miniaturization requirements have brought about the integral fabrication of many components. The thin-film approach permits the integration of numerous precision circuit elements and their interconnections. Part I of this three-part series deals mainly with the two deposition techniques of cathode sputtering and vacuum evaporation, and with their use in the fabrication of film resistors, capacitors, and R-C networks. Subsequent articles will discuss thin-film transistors and cryogenic thin films*



# Part I—Thin-film R-C networks

*A. E. Lessor, L. I. Maissel, R. E. Thun*
*International Business Machines Corporation*

The most pronounced trend in present electronic developments is the integral fabrication of many components, together with their functional interconnections. This approach offers many advantages: reliability is increased and cost is saved by the elimination of assembly operations and discrete joints; weight and volume can be shrunk because of reduction in packaging levels and reduction of electronically inactive structural material; and the compact and regular arrangement of the components often results in an increase in functional performance.

In such an integral function or circuit, the various components and their interconnections represent a multilayered pattern, which can be fabricated by three basic categories of processes, or a combination thereof. These categories include local changes of the physical properties of a basic material (solid-state diffusion or surface reactions); the addition of materials of the desired properties and geometric configurations by various deposition processes utilizing mechanical masks; and the subtractive processes of machining or etching selectively a previously deposited coating. In the last-mentioned case, the pattern may be generated by photolithographic processes or, alternatively, by the programmed control of a machine tool, such as an electron beam.

For practical purposes, and for two reasons, the choice of a microscopically planar configuration of these integral networks of components is mandatory. First, circuit theory today extends to elements with freely distributed parameters in only a few selected cases; and, second, optical transfer systems of sufficient resolution for generating the circuit pattern can reasonably be designed only for a transfer from an object plane to an image plane.

Two major microelectronic technologies of a truly functional character have reached the state of development at which they are about to enter mass production and find major systems applications—namely, monolithic semiconductor circuits based on an extension of silicon junction transistor technology, and thin-film circuits. At present, thin-film circuits assume the form of "hybrid" circuits (for example, deposited *R-C* networks with attached discrete chip transistors and diodes), since the development of practical deposited active devices is still in the laboratory stage. Monolithic circuits offer the advantage of a complete integration of all circuit components. Film circuits, on the other hand, permit the use of circuit elements with considerably closer tolerances and the integration of a larger number of passive components and their interconnections. In other words, monolithic circuits today offer functional integration at the circuit level, and thin films at the passive component plus intercircuit interconnection level.

Numerous methods of thin-film deposition are under investigation. This article will describe cathode sputtering and vacuum evaporation, the two methods that are most prominently used because they offer the most convenient control of the dominant process parameters and, consequently, film properties.

A freshly deposited film circuit panel containing 32 transistor logic NOR gates and most of their interconnections. The transistors are not yet attached.

### Cathode sputtering

**Techniques.** Although the use of cathodic sputtering as a method for the deposition of thin films predates vacuum evaporation by many years,[1] the latter has received far more widespread application because evaporation is more convenient for many materials and generally gives higher deposition rates. In recent years, however, it has been found that certain materials are more conveniently deposited by sputtering. In some cases it is impossible to deposit materials by any other means.

One of the most significant spurs to the growth in the popularity of sputtering has been the development of a tantalum-film technology for microelectronic resistors and capacitors.

Sputtering is, in essence, the ejection of atoms from the surface of a material as a result of bombardment by atoms or ions. For film deposition applications, the material to be sputtered (commonly in the form of a flat sheet) is used as the cathode in a low-pressure glow discharge in an inert gas, such as argon. Ions of the latter then bombard the surface of the sheet and eject from it atoms, which diffuse away through the glow discharge and deposit as a thin film on any surface nearby. Figure 1 shows a typical sputtering system.

Some of the major differences between sputtering and evaporation are:

1. Sputtering rates depend only weakly on vapor pressure, so highly refractory metals, such as tantalum and tungsten, are just as easy to sputter as are the less refractory ones.

2. Sputtered films have the same chemical composition as the starting material. To ensure this, it is necessary that the cathode does not get too hot during sputtering and that any chemical reaction with the sputtering gas is negligible.

3. There is, in principle, no limit to the planar area over which films of uniform thickness can be deposited by sputtering.

4. Since the sputtered atoms usually reach the substrate by diffusion, sharp shadowing at mask edges will not occur, and sputtered deposits will tend to show round corners.

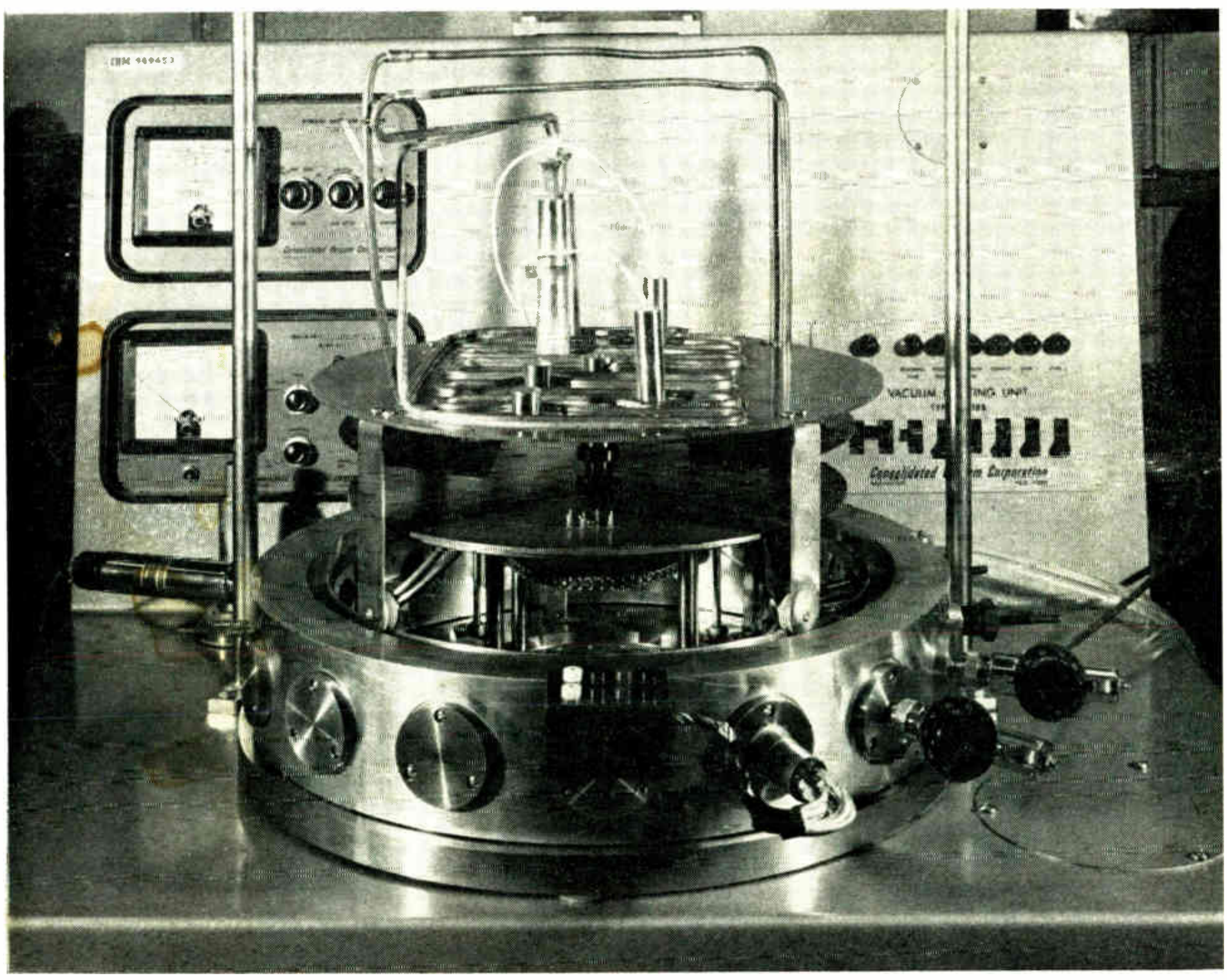
5. The electrons and ions of the glow discharge will bombard the various exposed surfaces in the sputtering chamber. Thus, the problems of heat dissipation and outgassing are usually more severe during sputtering than during evaporation.

6. Because of interference by the glow discharge, monitoring and control of the deposition rate are often very difficult. On the other hand, at constant pressure and power input, deposition rates are usually quite reproducible.

**Reactive sputtering.** When a chemical reaction between the sputtered material and sputtering gas occurs, the process is termed reactive sputtering. Because of the presence of ionized species in the glow discharge, gases are generally more reactive than they would be if they were present at the same partial pressure during film deposition by evaporation. In addition, the total amount of reactive gas that can be used is limited only by the total sputtering pressure. Most commonly, a mixture of inert and reactive gases is used to allow maximum deposition rates, as well as complete reaction.

Because of the very wide range of compounds that it

Fig. 1. Typical cathode-sputtering system with bell jar removed. Water-cooled plate on top is the cathode shield, which suppresses sputtering from rear of cathode. Multisource cathode directly below is rotatable. The substrate to be coated is laid onto the anode plate at the bottom.



I. Stability of tantalum resistors

| Resistor Type | Sheet Resistance, ohms per square | Storage at 150°C, hours | Increase in Resistance, per cent |
|---|---|---|---|
| Gold-loaded grain boundaries | 35 | 2000 | 0.10 |
| | 80 | 4000 | 0.50 |
| Nitrogen-doped with anodized surface | 38 | 1850 | 0.27 |
| | 100 | 1850 | 0.83 |
| Hermetically sealed | 250 | 1000 | 2.00 |

forms, oxygen has received by far the greatest attention as the gaseous component of reactive sputtering. Virtually all the common metallic oxides[2,3] have been laid down in film form by this technique. Of particular interest has been recent work on amorphous glossy films with glasslike characteristics.[4] For further details, the reader is referred to reviews that have been published recently on the subject.[5,6]

**Tantalum-film technology.** Tantalum films possess the unusual advantage of being convertible to both resistors and capacitors. This fact greatly simplifies the deposition technology for *R-C* networks. Since tantalum is difficult to evaporate at reasonable rates, sputtering has become the generally preferred method for the deposition of these films. Despite the fact that tantalum is a highly reactive metal, well known for its properties as a getter, remarkably pure films can be obtained. Methods include chemical neutralization of the reactive impurities in the sputtering gas,[7] careful elimination of contaminants by use of ultraclean systems,[8] and the "cleaning" of films *in situ* by the technique of alternate deposition and ion bombardment.[9]

**Tantalum-film resistors.** The outstanding attraction of tantalum as a resistor material is its refractoriness. Because of this property, defects quenched into the film during its deposition are very stable against the effects of annealing. It is usually necessary to heat tantalum films to temperatures of the order of 600°C before irreversible decreases in the resistance can be initiated. The refractoriness of tantalum also gives it a very low tendency to agglomerate (break up into fine droplets) at high temperatures.[10]

Although bulk tantalum is known to form a self-limiting oxide at temperatures below about 400°C[11] it has been found that tantalum films heated in air increase in resistance far more than can be explained in terms of the measured thickness of the surface oxide.[12] This has been shown[13] to be caused by grain-boundary diffusion of oxygen into the film, which creates an oxide layer in series as well as in parallel with the film. These oxides "platelets" perpendicular to the plane of the film influence both the high-frequency behavior of tantalum films and their temperature coefficients of resistance.

There have been several approaches to solving this problem of long-term instability against oxidation. The simplest but most expensive solution has been to seal all tantalum components hermetically in an inert atmosphere.[14] Much the same effect as an hermetic seal is obtained by forming a thick anodic oxide on the surface of the resistor.[15] A totally different approach has been to prefill the grain boundaries of the film with a suitable metal (such as gold), which virtually excludes grain-boundary oxidation.[16] The electrical properties of the film are unaffected by the grain-boundary filling technique, but a much more stable resistor is obtained.

In addition to diffusing from the surface, oxygen trapped during deposition can also diffuse internally, to grain boundaries so that tantalum films heat-treated *in vacuo* may sometimes show an irreversible increase in resistivity. This can be prevented by partial reactive sputtering in nitrogen.[8] Table I compares the stability of tantalum resistor films of various types.

According to Mathiesson's rule, the product of electrical resistivity and temperature coefficient should be a constant for any metal. Thus, films of very high resistivity should have temperature coefficients close to, but not less than, zero. In practice, most high-resistivity resistor films (including tantalum) have large, negative temperature coefficients. For this reason, tantalum films of high sheet resistivity, such as those obtained by oxygen doping[13] have very limited application.

Resistance uniformity of tantalum films over fairly large areas to better than ±5 per cent[17] is readily obtained. Achievement of absolute resistance values is somewhat more difficult, particularly if some sort of stabilizing heat treatment is applied after deposition. If anodic oxide films are used for resistor protection, the precise thickness of the protective oxide is easily varied from batch to batch. Such anodic oxidation can consequently be used as a method for trimming the resistors at little extra cost. Individual trimming, as opposed to batch trimming, is of course not as convenient. A rather different technique for trimming tantalum resistors consists of the application of a series of high-voltage pulses to individual resistors. Joule heating causes the temperature of the film to rise briefly to 400–500°C during each pulse, resulting in a small, permanent increase in the resistivity (about 0.05 per cent per pulse). Resistance is measured between each pulse, and pulsing is continued until the desired value has been reached.

**Tantalum-film capacitors.** When bulk tantalum is anodized[18] and an electrode is subsequently evaporated onto the surface of the oxide, short-circuiting through the oxide to the underlying tantalum generally occurs. However, a good capacitor can be produced if, instead of an evaporated-metal electrode, a conducting oxide (such as $MnO_2$) is used.[19] The latter serves to re-form any bad spots in the oxide, provided an anodic bias (tantalum electrode positive) is maintained. For even small cathodic bias, the devices are destroyed.

Berry and Sloane[20] have shown that these limitations are absent in tantalum films sputtered or evaporated onto smooth glass substrates. Metal electrodes can be evaporated into direct contact with the surface of anodized tantalum films to give capacitors of good quality. High yields can be achieved and a substantial fraction of the anodic breakdown voltage can be withstood in the reverse direction.

Using breakdown voltage times capacity per unit area as a figure of merit, tantalum film capacitors have a rating of about 5.5 $\mu F \cdot V/cm^2$. The corresponding figure for evaporated silicon monoxide films is about 2. In addition, dielectric thicknesses down to about 150 Å can be used without significantly affecting the figure of merit, whereas most evaporated dielectrics have a minimum usable thickness of 500–1000 Å.

The yield of tantalum film capacitors depends on the metal used for the top electrode and on its area. Certain metals, such as cadmium and gold, lead to good capacitors at high yields for electrode areas up to about 2 $cm^2$, whereas other metals, such as chromium or titanium, result in poorer quality capacitors whose yield begins to fall off for electrode areas in excess of a few tenths of a square centimeter. Silcox and Maissel[21] have demonstrated that these effects are due to the presence of microfissures in the anodic oxide. Metals that tend to replicate the surface of the oxide more closely have a greater probability of filling the microfissures and thus creating short circuits or high-leakage paths. Particularly for very thin oxides, other metal properties such as the work function also play a significant role in determining the electric properties of tantalum film capacitors.[22]

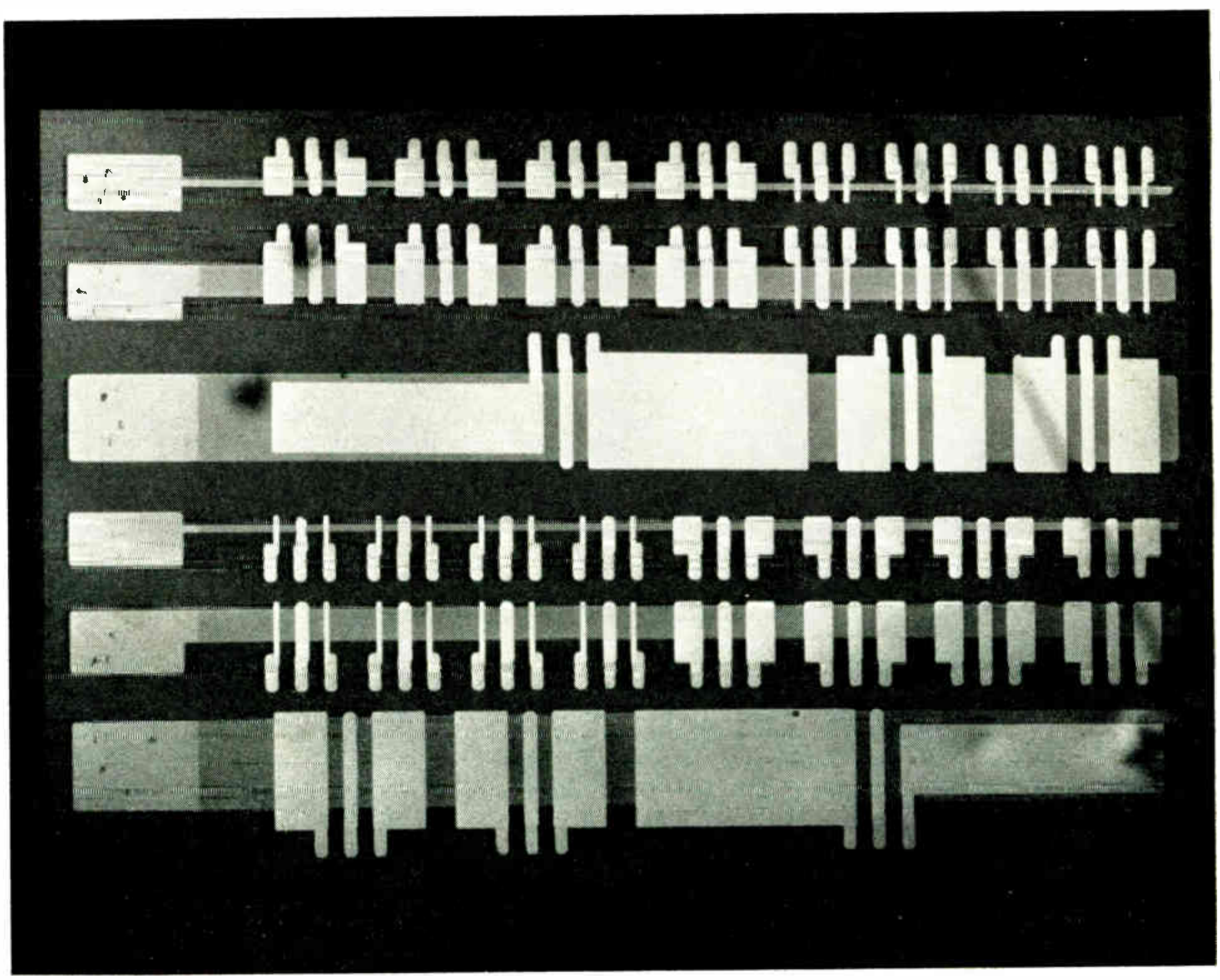
The dissipation factor of tantalum film capacitors at high frequencies is limited by the combined series resistance of the two electrodes. Mechanically stable tantalum films with sheet resistance less than 1 or 2 ohms per square are difficult to produce so that tantalum film capacitors intended for use at high frequencies have to be underlaid with some other, more conductive, metal. Aluminum is commonly used for this purpose. Since aluminum will itself anodize, any pinholes that may have penetrated all the way through the tantalum film are automatically healed. Devices of this type with capacitances in excess of 1000 pF and dissipation factors less than 5 per cent at 10 Mc/s have been reported.[7] In Fig. 2, various capacitor patterns are shown.

**Tantalum R-C circuits.** As already mentioned, sputtering of sharp lines through masks is usually not possible. Most tantalum technologies, therefore, have centered around postdeposition etching of circuit patterns.[23,24] Most commonly, both the resistive and conductive layers are deposited first and are etched subsequently in several steps. In some instances, tantalum technology has been combined with silicon integrated device technology,[25] with promising results.

### Evaporated circuits

As previously mentioned, cathode sputtering frequently is the preferred deposition method for high-

Fig. 2. Test array of tantalum oxide film capacitors. The long, dark strips are anodized tantalum films. The bright areas on top are the counter electrodes, which usually are deposited by means of vacuum evaporation techniques.



### II. Resistivity vs. composition, Cr-SiO cermet resistors

| Atomic per cent Cr | Resistivity, ohms per square |
|---|---|
| 90 | 10–120 |
| 65 | 100–600 |
| 60 | 300–700 |
| 50 | 600–1000 |

melting or refractory materials and many compounds. Evaporation *in vacuo*, on the other hand, is commonly chosen for lower-melting metals, such as gold, copper, chromium, and aluminum, and those simpler inorganic dielectrics that do not decompose appreciably in the vapor phase. Examples of the latter group of compounds are SiO, $MgF_2$, ZnS, and a few other oxides and fluorides.[26-28]

To avoid solid-state reactions and, in particular, electrolysis between the various films, a compatible set of resistor, conductor, and dielectric materials must be carefully chosen. Other considerations of importance are mechanical stresses in the film stacks, suppression of recrystallization leading to rough film surfaces, and many subtle interrelations between deposition parameters and mechanical or electrical film properties, which are outside the scope of this article.

Today a wide variety of materials, processes, and deposition equipment is used throughout industry in the fabrication of evaporated films.[29] To give a descriptive account of the present state of the art, an approach that is capable of economically producing large numbers of film panels under closely controlled conditions will be described.

**Processes and materials.** Three basic materials are used for evaporants: copper, chromium, and silicon monoxide. Silicon monoxide is used as a circuit underlayer to provide a chemically reproducible surface for resistor deposition, as crossover insulation and, where necessary, as a dielectric material for capacitors. Its low dielectric constant ($\epsilon = 6$) results in weakly coupled crossovers, but limits the practically achievable capacitance values. A mixture of chromium and silicon monoxide, flash-evaporated as a premixed powder, provides a resistor material of high stability.[30,31] Stable resistivities ranging from 10 to 1000 ohms per square are available by merely altering the atomic ratio of Cr to SiO. Table II shows the range of resistivities as a function of composition. Copper, with an underlayer of chromium to facilitate adhesion, is used for the conductors and capacitor plates.

Seven evaporations are required to complete a complex *R-C* network. In five depositions, film patterns are formed by evaporating through metal masks. The other two cover the entire substrate with a sheet film and require only a large opening mask to protect the substrate holder from the build-up of a deposit.

**Deposition equipment.** Conventional high-vacuum systems are often employed in the deposition of film circuits. They may contain simple fixtures to hold and register the mask and substrate, or may be equipped with both a mask and source changers to shorten the cycle time and to avoid the exposure of the films to the atmosphere between the depositions of the various layers.

A more elaborate four-chamber in-line deposition system, designed for a high circuit output, was built by IBM for the U.S. Naval Avionics Facility, Indianapolis, Ind.[32] Each of the 2- by 2- by 1.5-foot chambers has its own pumping system, which consists of a mechanical fore pump, a 10-inch oil diffusion pump with a chevron baffle cooled by a three-stage Freon refrigeration unit, and a holding pump. Each chamber also has its own source chamber, which can be valved off from the main chamber for source servicing during a deposition run without breaking the main vacuum.

The four chambers are bolted together to form one continuous vacuum space. A valved entrance magazine on the first chamber allows the loading of 24 substrate holders, each of which holds a 3.75- by 5.187-inch substrate, without breaking the main vacuum. A similar exit magazine on the fourth chamber allows the collection of 24 substrates. Modular transport mechanisms in each chamber transport substrates from the entrance magazine through the four chambers and deposit them in the exit magazine. Each transport mechanism has a mask changer associated with it, which consists of a mask cartridge of six-mask capacity, mask preheaters, and a mechanism for carrying the heated mask from the cartridge to the deposition position and replacing it in the cartridge when contaminated.

Modular panels contain the controls for the transport mechanisms, substrate heaters, mask changers, and the controls for the evaporation sources. Interlocks prevent transport mechanism operation if a mask is engaged with a substrate holder, and indicator lights show the positions of transport mechanisms, masks and substrate holders, and other pertinent process information. Ion gauge rate monitors[33,34] are used to control the evaporation rates of both SiO and Cu, whereas the resistor deposition is controlled by directly monitoring the film resistance on a pilot slide.

**Peripheral processes.** Considerable processing is done outside of the vacuum system. It has been found that savings in masking and evaporation steps can be realized by chemically etching part of the interconnection pattern. If the lower conductor pattern is etched, the line profile can be smoothed out and sharp corners removed by electrolytically micropolishing the etched pattern. This increases the insulation yield and gives a more gentle slope for crossover deposition. The etch pattern can be very complex, having a pattern equivalent of three or more evaporation masks. The combination of several deposition steps into one reduces registration problems and allows rapid response to engineering changes. Photolithographic masks are, of course, less expensive than mechanical masks.

Precision metal masks are either arc-eroded Invar or etched 0.004-inch molybdenum foil brazed to a thick molybdenum backing plate. With either technique, line opening tolerances of $\pm 0.0003$ inch can be held. The positional tolerance of a mask opening is held to within $\pm 0.001$ inch of the reference point. Similar tolerances are held with etch artwork. Each mask or etch artwork is preregistered to a photographic composite of the circuit. In-process registration is maintained by mating V notches in the substrate holder with a fixed pin and a spring-loaded pin of the mask frame (Fig. 3). The deposition

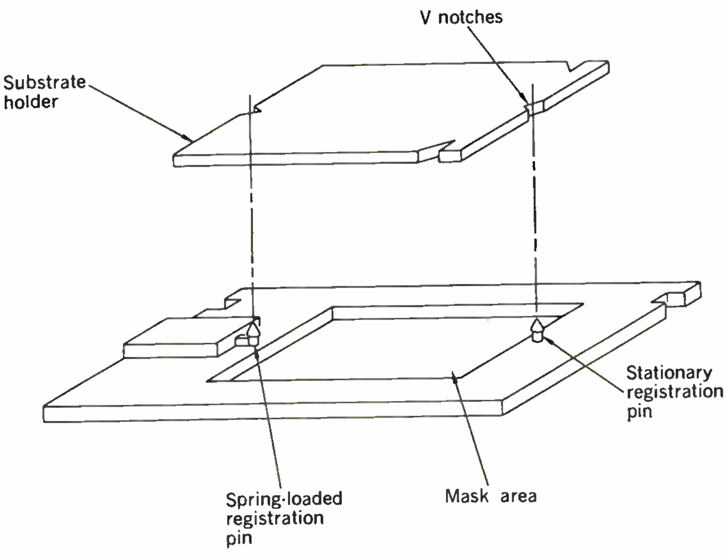
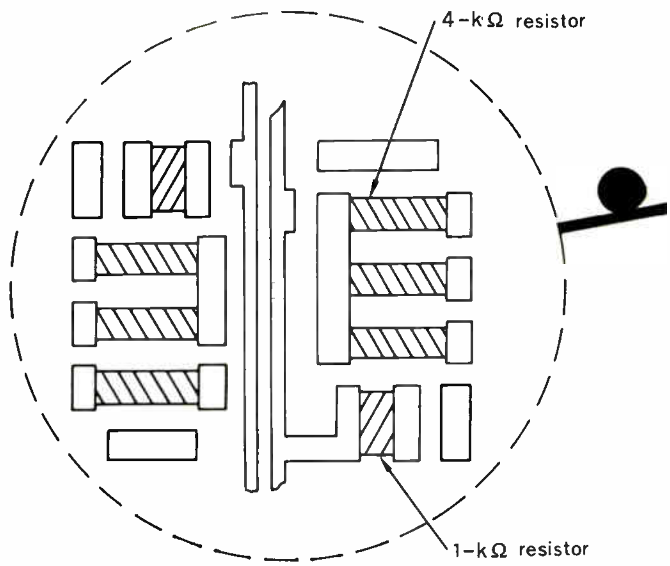


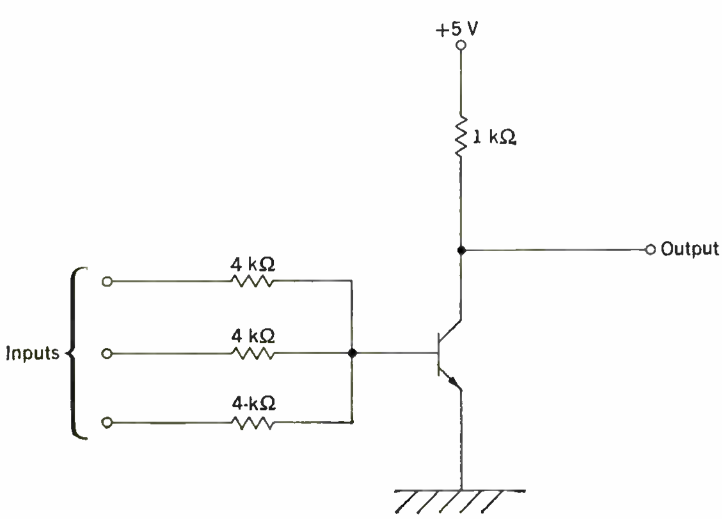

Fig. 3. Mask and substrate holder engagement.




### III. Deposition sequence and process parameters

| Chamber No. | Material | Use | Substrate Temperature, degrees C | Pressure, torrs |
|---|---|---|---|---|
| Pass I: | | | | |
| 1 | SiO | undercoat | 350 | $< 4 \times 10^{-6}$ |
| 2 | Cr-SiO | resistors | 160 | $< 1 \times 10^{-5}$ |
| 3 | SiO | resistor overcoat | 350 | $< 4 \times 10^{-6}$ |
| 4 | Cr-Cu | conductor etch plane | 170 | $< 1 \times 10^{-5}$ |
| Pass II: | | | | |
| 1 | SiO | insulation | 350 | $< 4 \times 10^{-6}$ |
| 2 | SiO | insulation | 350 | $< 4 \times 10^{-6}$ |
| 3 | SiO | insulation | 350 | $< 4 \times 10^{-6}$ |
| 4 | Cr-Cu | crossover jumpers | 110 | $< 1 \times 10^{-5}$ |

Fig. 4. Simple transistor–resistor NOR gate. Because of its simplicity, it is well suited as the basic building block of highly functional film circuit panels.




sequence and critical process parameters are summarized in Table III.

Between the first and second pass, the "personality" interconnection and the land pattern of the panel are etched, and individual components may be measured and tested.

The pressure differences between the connected chambers shown in Table III result from the gettering action of the high capacity, well-outgassed SiO sources. It should be noted that the pumping capacity is such that the system reaches $1 \times 10^{-5}$ torr in 10 minutes from one atmosphere and $5 \times 10^{-6}$ torr in about 25 minutes with all heaters operating. No difficulties are encountered in holding the defined process parameters.

**Circuits and circuit testing.** One problem encountered in fabricating highly functional assemblies is the testing of their electronic performance. By checking only the dynamic characteristic of the entire function, the marginal operation of individual components and elements might be sufficiently masked to introduce uncertainties into the certification procedure. This difficulty has been circumvented for film circuits by the two-pass production process outlined previously. In the first pass, a matrix of individual components (such as resistors with their termination lands) is deposited, after which each component can be tested. The functional network is completed by the second-pass deposition of all functional interconnections only after it has been established that the deposited components meet their specifications.

This testing is performed with multicontact probes, which instantaneously contact all lands of an entire film panel or substrate. Readout may be automatically controlled and evaluated. Testers have been built with a capacity of up to 100 resistors per minute.

While thin-film circuits are adaptable to a wide range

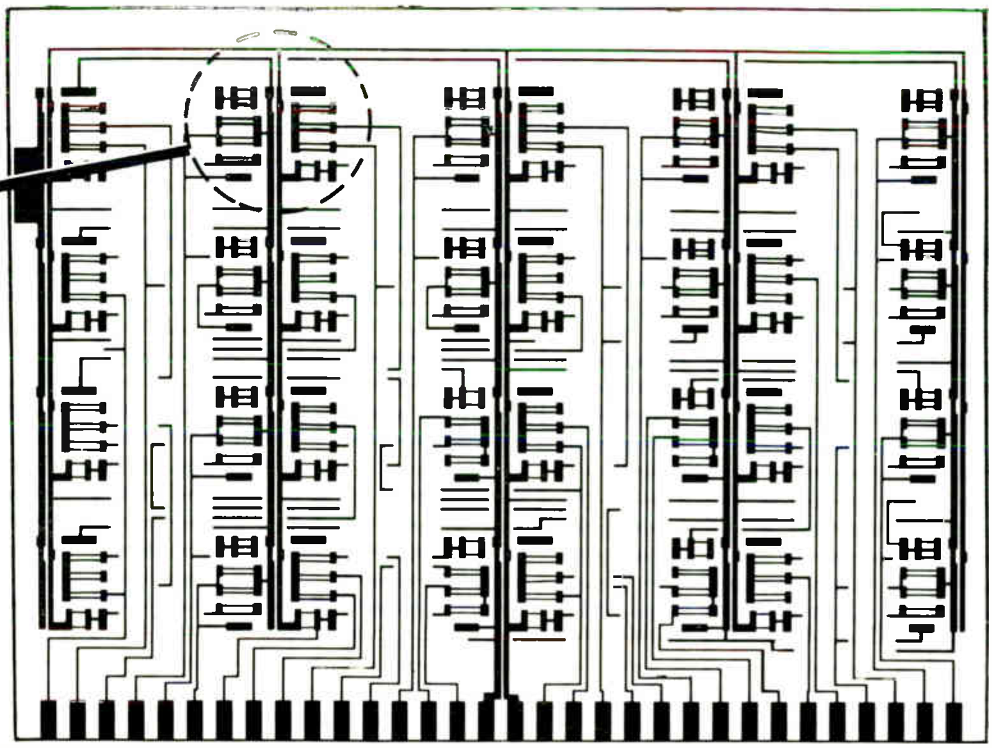
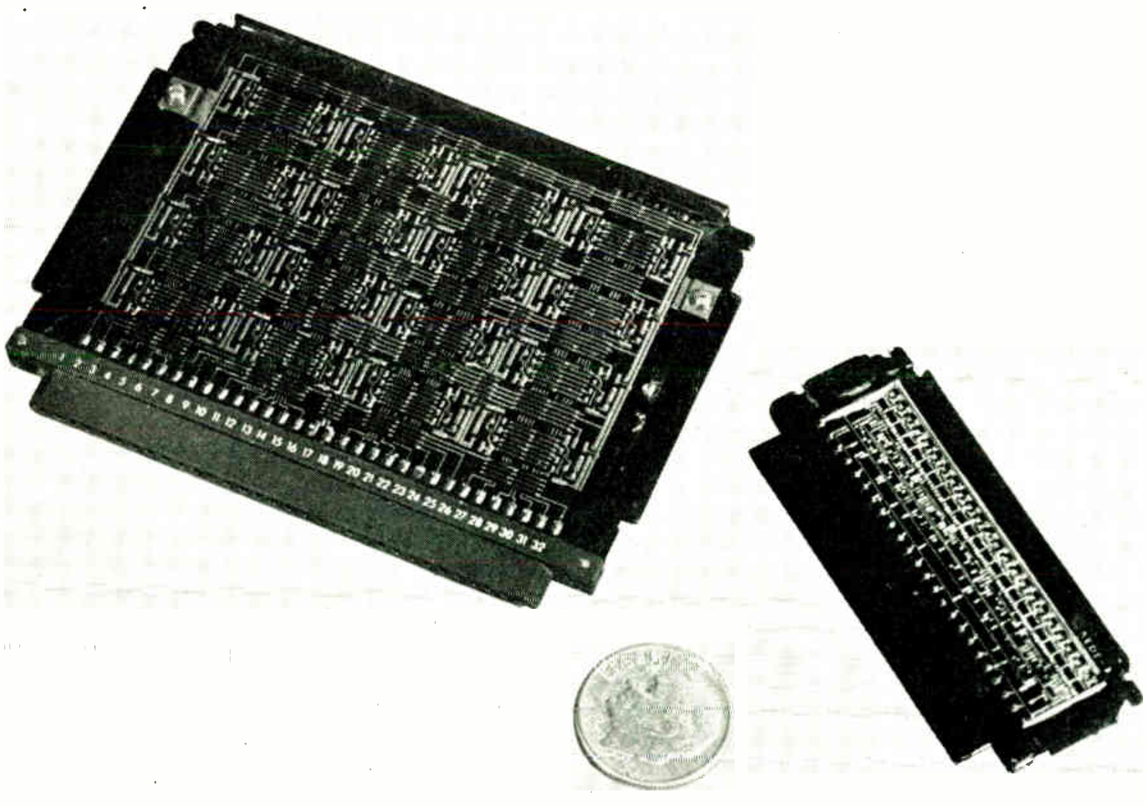


Fig. 5. The basic layout of the panel shown in the title illustration, representing the resistor matrix and the etched interconnection network.

Fig. 6. Left—Packed and coated 32-circuit panel with attached transistors, connectors, and frame. Right—Similar 24-circuit panel of considerably reduced size, now under development.



of component values, functions, and pattern configurations, testing and fabrication is obviously simplified if the desired functions can be formed from a fixed matrix of basic components or circuits by solely changing the interconnection pattern. From this aspect, digital applications are especially attractive in utilizing functional film circuit panels.

Figure 4 shows a simple logic NOR gate which reduces the number of elements to be attached to the passive component network to a single transistor per logic decision. The title illustration shows a freshly deposited and not yet packaged film panel containing 32 of these transistor logic NOR circuits, distributed in a two-dimensional matrix over the glass substrate area. The intercircuit interconnections are arranged in regular vertical and horizontal channels, and the dark patches represent the SiO-insulated crossover lattices, which are identical for all possible connection patterns with regard to the depositions performed through masks. The panel size is 1.8 by 2.5 inches. The pattern layout is shown schematically in Fig. 5.

The bottom edge of the panel displays the row of 32 lands to which a connector can be attached for communication with the outside. The integration of 32 circuits has thus led to a reduction of discrete intercircuit interconnection of six to one, with a corresponding increase of overall packaging density and system reliability.

The packaged and coated film circuit panel with attached transistors, connector, and metal frame are shown on the left in Fig. 6. The transistors are glass-passivated chips of 25-mil edge dimensions, which are soldered with their contact balls directly to the pretinned film land areas. As can be seen, their size is insignificant and therefore the package volume does not increase.

The integration level of 32 circuits represents a reasonable compromise of fabrication yield, reduction of external interconnections, and cost of the packaged unit. This panel type already has found application in an airborne computer developed for the U.S. Naval Avionics Facility in Indianapolis.

A current panel development is shown on the right in Fig. 6. This 24-circuit panel has a single row of circuits along the upper substrate edge. A significant size reduction of the package has been achieved through the corresponding simplification of the interconnection pattern and, more important, through a considerable increase in film pattern density. This panel approaches a degree of microminiaturization beyond which further savings in weight and volume will depend both upon new technological developments for the higher packaging levels and upon peripheral electronic circuits, such as power supplies.

*(The second article in this series will appear in May.)*

REFERENCES


1. Grove, W. R., *Phil. Trans. Roy. Soc. London*, Series B, vol. 142, 1852, p. 87.

2. Hiesinger, L., and H. Koenig, "Optical and Structural Properties of Sputtered Films with Special Consideration of the Reaction with the Residual Gas," *Festsch. 100 Jähr. W. C. Heraeus GmbH Hanau*, 1951, p. 376.

3. Holland, L., and G. Siddall, "The Properties of Some Reactively Sputtered Metal Oxide Films," *Vacuum*, vol. 3, Oct. 1953, pp. 375–391.

4. Sinclair, W. R., and F. G. Peters, "Preparation of Thin Mullite Films," *J. Am. Ceram. Soc.*, vol. 46, 1963, p. 20.

5. Holland, L., *Vacuum Deposition of Thin Films.* New York: John Wiley & Sons, Inc., 1958.

6. Schwartz, N., "Reactive Sputtering," *Proc. 1963 Am. Vacuum Symp.*

7. Maissel, L. I., R. J. Hecht, and N. W. Silcox, "Techniques for Production of Low Cost Tantalum Circuits," *Proc. Electron. Components Conf.*, 1963, p. 190.

8. Gerstenberg, D., and E. H. Mayer, "Properties of Tantalum Sputtered Films," *Ibid.*, 1962, p. 57.

9. Frerichs, R., "Superconductive Films Made by Protected Sputtering of Tantalum or Niobium," *J. Appl. Phys.*, vol. 33, 1962, pp. 1898–1899.

10. Belser, R. B., and W. H. Hicklin, "Temperature Coefficient of Resistance of Metallic Films in the Temperature Range 25° to 600°C," *Ibid.*, vol. 30, 1959, pp. 313–322.

11. Cathcart, J. V., R. Bakish, and D. R. Norton, "Oxidation Properties of Tantalum Between 400° and 530°C," *J. Electrochem. Soc.*, vol. 107, 1960, pp.668–670.

12. Basseches, H., "The Oxidation of Sputtered Tantalum Films and Its Relationship to the Stability of the Electrical Resistance of These Films," *IRE Trans. on Component Parts*, vol. CP-8, June 1961, pp. 51–56.

13. Maissel, L. I., "Electrical Properties of Sputtered Tantalum Films," *Proc. 1962 Am. Vacuum Symp.*, p. 169.

14. Sikina, T. V., "High-Density Tantalum Film Microcircuits," *Proc. Electron. Components Conf.*, 1962, p. 24.

15. McLean, D. A., "Microminiaturization with Refractory Metals," *1959 IRE Wescon Conv. Record*, pt. 6, pp. 87–91.

16. Schaible, P. M., and L. I. Maissel, "Grain Boundary Diffusion in Sputtered Tantalum Films," *Trans. 9th Symp. Am. Vacuum Soc.* New York: Pergamon Press, Inc., 1962, p. 190.

17. Maissel, L. I., and J. H. Vaughn, "Techniques for Sputtering Single and Multilayer Films of Uniform Resistivity," *Vacuum*, vol. 13, 1963, p. 421.

18. Young, L., *Anodic Oxide Films.* New York: Academic Press, Inc., 1961.

19. McLean, D. A., and F. S. Power, "Tantalum Solid Electrolytic Capacitors," *Proc. IRE*, vol. 44, July 1956, pp. 872–878.

20. Berry, R. W., and D. J. Sloan, "Tantalum Printed Capacitors," *Ibid.*, vol. 47, June 1959, pp. 1070–1075.

21. Silcox, N. W., and L. I. Maissel, "Some Factors Controlling Gross Leakage Currents in Sputtered Tantalum-Film Capacitors," *J. Electrochem. Soc.*, vol. 109, 1962, pp. 1151–1154.

22. Standley, C. L., and L. I. Maissel, "Conduction Through Thin Tantalum Oxide Films," *J. Appl. Phys.*, vol. 35, 1964 (to be published).

23. Spector, C. J., and W. H. Jackson, "Tantalum Film Circuits," *Proc. Electron. Components Conf.*, 1961, pp. 31-1–31-5, 32-1–32-5.

24. Maissel, L. I., J. Simmons, and M. Casey, "Integrated Thin-Film Circuits," *IRE Trans. on Component Parts*, vol. CP-8, June 1961, pp. 70–79.

25. Sikina, T. V., M. Casey, J. Saunders, and F. Murray, "Integrated Tantalum-Silicon Microcircuits," *Proc. Electron. Components Conf.*, 1963, p. 182.

26. Halaby, S. A., L. V. Gregor, and S. M. Rubens, "The Materials of Thin-Film Devices," *Electro-Technol.*, vol. 72, Sept. 1963, pp. 95–122d.

27. Maddocks, F. S., and R. E. Thun, "Properties of Evaporated Film Capacitors," *J. Electrochem. Soc.*, vol. 109, 1962, p. 99.

28. Siddall, G., "Vacuum Deposition of Dielectric Films for Capacitors," *Vacuum*, vol. 9, 1959–1960, pp. 274–287.

29. Kernjian, E., ed., *Principles of Microelectronic Engineering.* New York: McGraw-Hill Book Co., Inc., 1963.

30. Beckerman, M., and R. E. Thun, "The Electrical and Structural Properties of Dielectric Metal Mixtures," *Trans. 8th Symp. Am. Vacuum Soc.* New York: Pergamon Press, Inc., 1961, p. 905.

31. Beckerman, M., and R. L. Bullard, "Integrally Fabricated Resistors and Their Performance," *Proc. Electron. Components Conf.*, 1963, p. 182.

32. Himes, W., B. F. Stout, and R. E. Thun, "Production Equipment for Thin-Film Circuitry Panels," *Trans. 9th Symp. Am. Vacuum Soc.* New York: The Macmillan Co., 1962.

33. Giedd, G. R., and M. H. Perkins, "Evaporation Rate Monitor," *Rev. Sci. Instr.*, vol. 31, 1960, p. 773.

34. Perkins, M. H., "An Evaporation Rate Control System," *Trans. 8th Symp. Am. Vacuum Soc.* New York: Pergamon Press, Inc., 1962.

# Laser irradiation of biological systems

*Investigations with the laser are in progress that may more clearly illuminate living processes. Applications to problems in biology and medicine are under exploration*

*S. Fine* *Northeastern University*
*E. Klein* *Roswell Park Memorial Institute*
*R. E. Scott* *Northeastern University*

Components of the electromagnetic spectrum are of prime importance to life on our planet. Considerable information has been obtained in the past on the interaction of electromagnetic radiation with biological systems.[1-11] These studies were in general carried out at low peak-power levels and low peak-power densities, as compared to those now attainable with lasers.

In the short time since Maiman's studies resulted in his achievement of the first operating laser,[12-14] laser operation has been obtained in a number of doped crystals,[15] glasses,[16] plastic, liquids, gases,[17] and semiconductors,[18-20] at various wavelengths. Further studies have resulted in increasing energy output with total energies now exceeding 1000 joules per pulse (obtained from crystalline lasers); continuous operation of crystalline, gaseous, and semiconductor units; high peak-power pulses, often called "giant" or Q-switched pulses,[21] now reported at peak powers exceeding 500 MW per pulse[22]; pulse repetition frequencies exceeding one pulse per second with an energy output exceeding 50 joules per pulse; frequency doubling and frequency tripling using crystals such as quartz,[23] and Raman laser action.[24]

Biological studies on the interaction of laser radiation and biological systems are being carried out by a number of groups within this country and abroad. The rapid advances in this field have required scientists in various disciplines to work in close collaboration with each other. These have included physical scientists and engineers concerned with basic research and development of laser systems for purposes other than biological applications, as well as physicists, engineers, and other scientists concerned with specialized laser systems for biological and medical work, biophysicists with a knowledge of dosimetry, and biological scientists who have applied these units to the investigation of biological problems.

Studies on biological systems have been carried out at various wavelengths, at energy (and energy density) levels from the millijoule region to more than 1000 joules per pulse and at power (and power density) levels from the milliwatt to the multimegawatt region. These studies have been carried out at the molecular level, on cellular components and isolated cells, on microorganism, on tissue culture, on isolated physiologic systems, on individual organs, and on intact animals. Some preliminary studies in man have also been attempted. The studies have depended and will continue to depend upon the basic and applied research by the physical scientists and engineers in this field.

The biological studies include the following areas:

1. Study of the interactions of relatively coherent, monochromatic radiation, as made available by lasers, with biological systems.
2. Application of lasers to assist in understanding biological systems.

3. Adaptation of laser devices for medical diagnosis and therapy.

4. Assessment of the hazards involved at the power and power density levels attainable by lasers in the respective regions of the electromagnetic spectrum, on both a short-term and long-term basis, and development of safeguards.

5. Correlation of effects on biological systems with those on physical systems in order to assist in an understanding of underlying factors.

Laser devices have been available only for a short period of time, and are still, in most cases, prototype in nature. Consequently, it has not been possible to obtain data on long-term *in vivo* biological effects. Even short-term *in vivo* and *in vitro* studies should be viewed as preliminary.

### Free radicals in biological systems

Free radicals have been reported to occur normally in biological materials, including rabbit muscle, brain, lung, liver, mouse liver, and various melanin preparations.[25–27] It has been suggested that some free radicals may be implicated in carcinogenesis, the aging process, and the action of radiation on biological systems.

Measurement of electron spin resonance of biological materials irradiated at 6943 Å has been carried out[28, 29] to determine free radical production. Tissues examined included mouse skin and underlying liver. Enzyme preparations studied included preparations of homolysin and collagenase. Following irradiation, a group of samples was frozen in liquid nitrogen, and maintained in dry ice until and during spectroscopic examination. Nonirradiated controls were similarly treated. Other samples were deliberately warmed to room temperature. An order of $10^{15}$ spins per gram was obtained in irradiated skin of black mice and in homolysin preparations, whereas the respective control samples gave no evidence of resonance at the same sensitivity. Both nonirradiated and irradiated skin of white mice and collagenase preparations did not show a signal. Resonance was observed in both nonirradiated and irradiated liver samples. Estimates of the number of spins per gram were made by a calibration using pitch samples. All lines were observed at a frequency of 34.75 Gc/s, in a magnetic field of 12 400 oersteds, giving an approximate $g$ value of 2.002, which is typical of those associated with free radicals. These experiments therefore indicated with high probability that free radicals are generated in the living organism and in isolated biological preparations by exposure to laser radiation. Further studies will be required to determine the nature of these free radicals, and whether they differ significantly from those produced by heating.

### Macromolecular preparations

Studies were carried out to determine effects of laser radiation on biological properties of macromolecular preparations and included irradiation of blood group substances, enzymes, and plasma proteins.

The blood groups of man may be identified by chemical complexes, called blood group substances A and B, which are associated with red blood cells. Red cells without these complexes are called group O. Individuals with group O blood have proteins (anti-A and anti-B antibodies) in their plasma which will cause agglutination (clumping) of A or B cells, respectively. Individuals with group B cells carry anti-A in their plasma, and those with group A cells have anti-B in their plasma. Addition of A cells to anti-A antibodies, (or B cells to anti-B antibodies) in a test tube results in hemagglutination (clumping). Addition of blood group substances prior to adding the red cells inhibits this reaction. It might be expected that irradiation of the blood group substances would tend to decrease their activity, that is, reduce their effectiveness in inhibiting hemagglutination.

Laser irradiation of blood group substances A and B was studied[29] and solutions of the blood group substances were irradiated at 6943 Å and 10600 Å with single and multiple pulses at energy levels ranging from 70 to over 500 joules per pulse. An attempt was made to measure the degree of absorption by the solution at these wavelengths. Following irradiation, the inhibition of hemagglutination by anti-A and anti-B antibodies was increased considerably rather than decreased, in comparison with the inhibitory activities of nonirradiated control samples. This increase in inhibitory action could have been brought about by the liberation of smaller (active) chains of sugars, resulting from degradations of the blood group polysaccharides. Alternately, the increased inhibitory activity could have been brought about by nondegradative changes in molecular structure as the result of laser irradiation. Physical and chemical techniques to assist in the characterization of end products of the interaction are being carried out. Studies on the blood group substances indicate that laser irradiation may enhance the biological reactivity of a molecule rather than decrease its reactivity.

Enzymes are a class of macromolecules (proteins) which accelerate specific chemical transformations. An enzyme may catalyze more than one transformation. For example, fibrinolytic enzymes can catalyze the breakdown of fibrin (the fibrous protein in blood clotting), of other proteins such as casein, of synthetic esters such as tosyl argenyl methyl ester, and can convert an inactive precursor of fibrinolysin to the active enzyme. Irradiation of fibrinolytic and proteolytic enzymes has been carried out[29, 30] to determine whether laser irradiation can differentially affect one or more of these activities in comparison with its effect on the remaining activities. Irradiation has been carried out at 6943 Å and 10 600 Å on enzymes in solid form and in aqueous solutions. Studies are in progress to determine whether production of free radicals in these preparations[28] can be correlated with differential changes in the enzymatic activities.[30]

Human gamma globulin preparations were irradiated at 6943 Å and tested for their ability to interact with antiserum and rheumatoid factor.[30] Irradiated gamma globulin had more precipitative activity with antiserum and less with rheumatoid factor than the nonirradiated control specimens. The effects of laser irradiation on gamma globulin therefore differ from those of thermal denaturation.

### Microscope-focused laser radiation

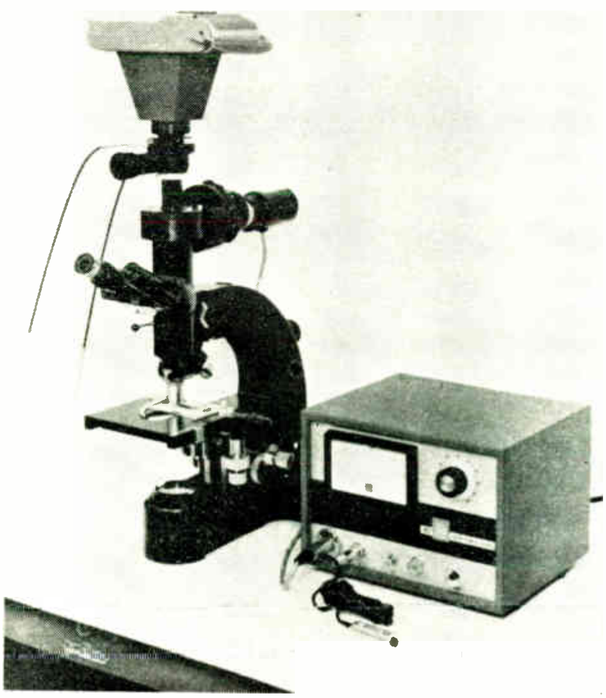
Laser units have been coupled with microscopes to permit irradiation of biological specimens at high levels of magnification. (Fig. 1 shows a laser–microscope unit developed by the Technical Research Group.) The laser beam can be focused to a spot size of less than 5 microns in diameter. A continuous gas laser operating at 6328 Å was focused through a trinocular microscope, and a spot size of 2.5 microns was obtained.[31] The laser image or spot size is a function of the laser, the microscope, and the specimen. If the laser is operating in a single mode, the distribution of the energy will be an Airy disk, the diameter of which is determined by the numerical aperture of the objective. If the material is homogeneously highly absorbent at the laser wavelength, then the area of absorptivity will be that of the Airy disk. If the laser is not operating in a single mode, then the energy is effectively spread over a greater area than the Airy disk. Heterogeneity throughout the biological specimen will result in absorption of a fraction of the energy and this, possibly, over a smaller area than is represented by the disk. As the radiation passes beyond the focal plane, the energy density decreases.

Studies on single cells at 6943 Å using a laser coupled through a microscope have been carried out at the (French) National Transfusion Center.[32] The results of the interaction were observed through a closed-circuit television system. Breakdown of red cells occurred immediately to within a few minutes after irradiation. At the spot where the beam struck, some coagulation of the red cells appeared to occur. White blood cells and tissue cultures have also been irradiated.[33] Janus green was used as a vital stain for mitochondria, in the attempt at selective irradiation of these cellular components.

Studies have been carried out on plant cells.[34] The laser microscope facilitated destruction of a small section of the cell wall to allow microneedles and micropipettes to enter the cell readily and to destroy or alter selectively sections of the cell. Sections of the cell wall and of cytoplasmic and nuclear constituents of the alga spirogyra were destroyed or altered with the use of a microscope laser unit at energy levels in the millijoule region. Methylene blue, a dye with absorption in the red region, was used to provide vital staining to enhance differential absorption of the radiation by cellular constituents. The authors concluded that portions of the cell could be subjected to low levels of energy, with little evident damage to areas other than those that had been irradiated.

Single cells, *in vivo* skin flaps, and isolated biological preparations have been irradiated.[35] Single cells have included mammalian red and white blood cells, reptilian (nucleated) red blood cells, and bacteria. The studies on skin flaps have been concerned with the microcirculation and blood flow patterns, coagulation, and hemorrhage. These studies were carried out with and without methylene blue. Methylene blue was added prior to and following irradiation to determine its possible effect on energy transmission. Spectral analysis of methylene blue has been carried out after irradiation at higher energy levels to determine the effects on the methylene blue itself. Threshold levels were investigated for destruction of the whole cell and for injury of the cell wall without evident damage to the nucleus with and without the addition of methylene blue. In general, the energy required for a specific effect was less when methylene blue had been absorbed by the cell. Initial attempts at inducing coagulation without hemorrhage have been carried out to assist in assessing the laser microscope as a clinical tool for microsurgery, as well as to obtain information concerning interactions of laser radiation with isolated biological systems and organs.

Fig. 1. View of a laser-microscope unit. The laser beam is focused through the microscope on the target.



### Cell culture

Tissue culture consists of the continuous cultivation of cells (or tissue fragments) in natural and artificial media. Conditions necessary to permit maintenance of cellular integrity, growth, and reproduction, including the composition of the growth media, oxygen tension, temperature, and irradiation, have been under intensive investigation for several decades. The effects of laser irradiation of tissue cultures have been studied.[36] Studies were carried out at energy density levels ranging from 1 joule per square centimeter to over 700 joules per square centimeter with a pulsed ruby laser. In other experiments, Q-spoiled systems were employed. Cells derived from the nervous system (dorsal root ganglion), heart, retina, and amnion of the chick embryo, as well as the retina of man and rabbits, were irradiated and observed directly and by time-lapse photography. Results included destruction of satellite cells and pigmented retinal cells. Prolongations of the early phase of cellular division (prophase) of amnion cells were reported. Inhibition of growth and reduction in $O_2$ consumption was observed in irradiated Hela cells. The dye Janus green considerably increased the sensitivity of amnion cells to the lethal effects of the radiation. The rate of contraction of the chick embryo heart was altered by laser radiation, while the same radiation appeared to have no effect on the cardiac cells in tissue culture. An estimate of the depth of penetration of the beam was obtained on the basis of histological evidence. Correlation between the degree of pigmentation and the degree of damage was studied.

Studies of the effects of irradiation of microorganism at 6943 Å were carried out[30] and lethal effects on several species of cocci, coliform, and chromogenic organisms observed. Surviving organisms of the pseudomonas group showed reduction in growth and temporary decrease in pigment formation on subculture.

### Spectroscopic ultramicroanalysis

A laser spectroscope has been developed[37] and used for analysis of metallurgical faults in specimens. The principle of operation is this: The pulse from a Q-switched ruby laser is focused through a microscope on a selected target. The preselected sample is vaporized by the laser beam. Carbon electrodes are then used to raise the temperature of the gases to spectral emissive levels. The resultant spectra from the vaporized material is then recorded photographically by a spectrograph. The use of the laser microprobe in this case is twofold—to select a predetermined small area or volume and to vaporize the material so that it can be raised to a suitable spectral emission level by the cross-firing carbon electrodes.

This type of unit has been used to obtain spectra from various normal tissue sections which had been removed from the animal prior to exposure.[38] The sample vaporized is about 50 microns wide and its weight about $10^{-7}$ grams. In preliminary *in vivo* spectroscopy,[39] the skin of normal, intact black mice and white mice and tumors (including melanomas) were examined in the living animals. These observations were compared with data obtained by *in vitro* spectroscopy.

A number of problems arise with this technique including spectroscopic contamination of the specimens by extraneous material. The technique, however, may provide spectroscopic *in vivo* analysis.

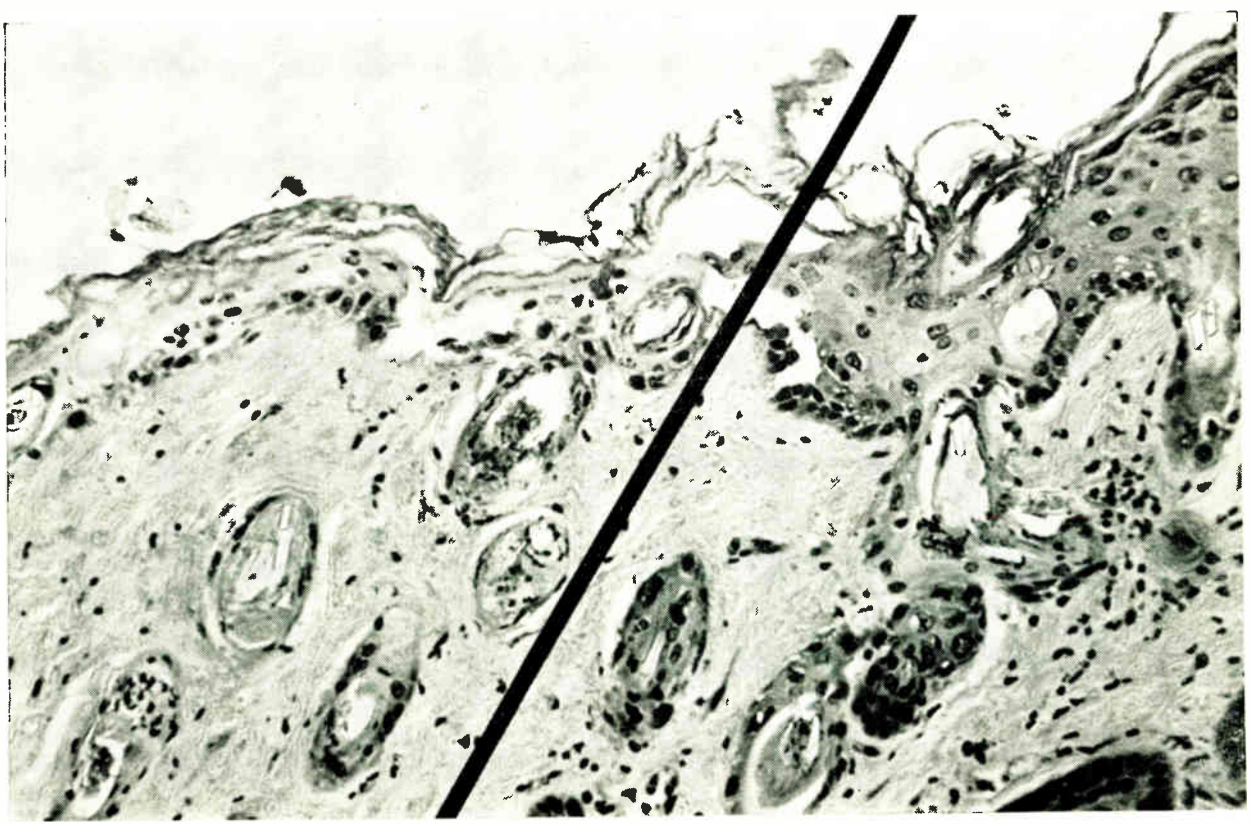
### Intact animals

Initial and subsequent studies[29, 30, 35, 40–44] were carried out at 6943 Å and 10 600 Å, at energy levels ranging from one joule to over 500 joules per pulse, at peak power levels exceeding 100 MW, and at pulse repetition frequencies exceeding one pulse per second. Effects of both unfocused radiation and radiation focused by lens systems were investigated on hamsters, mice, rats, and monkeys. Studies were also carried out on excised human tissues. The purpose of these experiments was:

1. To determine the short-term and long-term effects of the interaction of laser radiation with skin and underlying normal and tumor tissue.
2. To determine thresholds for gross and histological changes.
3. To determine the effect on biological pigments.
4. To evaluate the hazards associated with laser radiation.
5. To orient future studies of the biological interaction of laser radiation with biological systems *in vivo* and *in vitro*.

The initial studies were concerned with laser irradiation of the skin and mucous membranes as well as effects on deeper structures and on tumors in normal and tumor-

Fig. 2. Microscopic appearance of border of skin lesion. Line indicates sharpness of demarcation. Relatively normal skin (histologically) is to the right of the line.



bearing rodents including hamsters and mice. Emphasis in these initial studies was placed on the skin since it is a primary site of interaction in the intact organism. Observations on the skin of the rodents, furthermore, could be carried out more readily than on less superficially located organs.

**Skin and deeper structures.** In their early studies[41–43] on single pulses at energy levels of 30 to 50 joules and pulse durations of the order of 1 ms, Fine and Klein observed that the interaction of the radiation with the skin was associated with a plume of back-scattered material and radiation, seen at the point of impact and believed to be partially a consequence of vaporization of the superficial layers of the skin. The skin lesions subsequently went through the stages of blister formation with surrounding inflammation, exudations, crusting, and recovery. The majority of the animals survived irradiation at these energy and power levels. In the animals that died or were sacrificed, however, lesions were found deep beneath the skin lesion. The deeper lesions consisted of hemorrhage and edema in the abdominal wall and the viscera (i.e., intestine, stomach). Between the injured areas of the abdominal wall and the lesions in the underlying viscera, however, intervening layers of peritoneum appeared normal. Lack of free blood in the peritoneal cavity further indicated that the integrity of the peritoneal layers had been retained, despite extensive hemorrhages and severe injuries to adjacent structures. These observations were further substantiated by the histological findings of alternating injured and uninjured layers of tissues in the path of the radiation.

Microscopic findings included sharp delineation between nonviable and viable cells. In the skin, the sharpness of the boundary was indicated by the close proximity of the depigmented hair follicles to pigment-containing follicles, and of pigment-free granules to pigment-containing granules within the same follicle. The microscopic changes of laser-induced skin lesions thus differ from those of a thermal burn. Blisters following laser radiation, furthermore, were located within the epidermis, while following a thermal burn they are subepidermal (Fig. 2).

These findings suggested a degree of selectivity of the interaction of laser radiation with different biological materials or structures that could not be attributed to thermal factors alone. They suggested that the factors of importance might be related to the short pulse duration (shortness of the burst) and high peak power and peak-power density. The energy could consequently be considered as delivered in the form of an impulse.[42]

It was evident even from these early studies[42] that the relative mildness of the superficial skin lesions did not correlate with the marked severity of the associated deeper lesions. These observations were subsequently confirmed and extended by studies at higher energy and peak-power levels and at relatively high pulse-repetition frequencies.[29, 30, 35, 44]

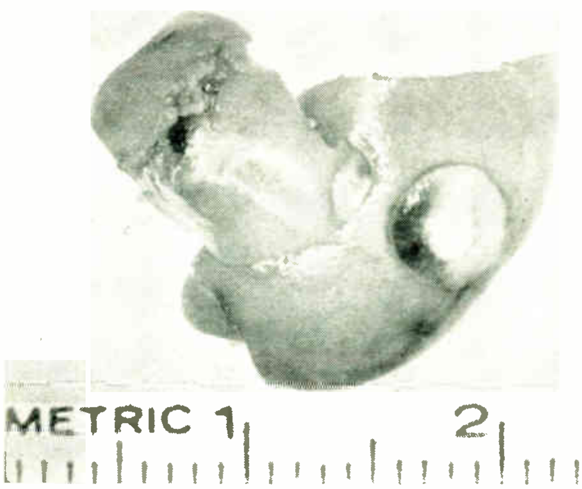
At higher total energy,[30, 35] the severity of the superficial as well as of the deeper lesions was increased. The relation between severity of injury and energy delivered, however, did not appear to be linear. The disparity between the severity of the superficial skin lesions and the severity of the deeper lesions in the muscles and viscera remained (Fig. 3). Thus relatively minor skin injuries were associated with extensive lesions of the internal structures. Even at high energy (over 500 joules), and in other studies[44] at power levels exceeding 100 MW, interposition of uninjured with injured layers of tissues were observed.

Microscopically, the lesions in the liver and other viscera also showed sharp, almost linear demarcation from the adjacent normal areas. The cells of the liver capsule (of Glisson) appeared to have remained intact in the area overlying the lesion. The architecture of lobulation and the outlines of cells and of some intracellular structures had been retained. Since the blood supply to the adjacent viable tissues appeared adequate, the usual causes of avascular necrosis could be excluded. The straight lines along which demarcation occurred at high as well as at low levels of peak power and energy suggested that the area of interaction was delineated by the incident beam, and that the interaction was due to effects other than thermal. This is also indicated by the presence of normal, intact tissues intervening between the superficial and the deeper lesions. The alternation of damaged and normal tissues excluded purely thermal causes. If the lesions observed had been produced by heat in the usual way, the resulting injuries would have been continuous and layers of normal (uninjured) tissues would not alternate with injured tissues along the path of the radiation.

Sharp delineation was apparently limited to the sites of the primary interactions. Subsequent secondary reactions (i.e., inflammatory exudates, hemorrhage, or reparative activity) were not sharply demarcated. The secondary reactions, furthermore, did not occur along straight lines, but followed the distribution of the blood supply.

Another aspect of selectivity of the radiation is suggested by the interaction with melanin (cutaneous pigment). In the skin lesions produced by laser irradiation, melanin is not apparent, although the structure of the hair follicle and its ability to produce hair has been retained. This selective effect on melanin was further indicated by the failure of melanogenesis during the recovery

Fig. 3. Round circumscribed lesion in liver after irradiation to skin of mouse. Irradiation through intact skin.



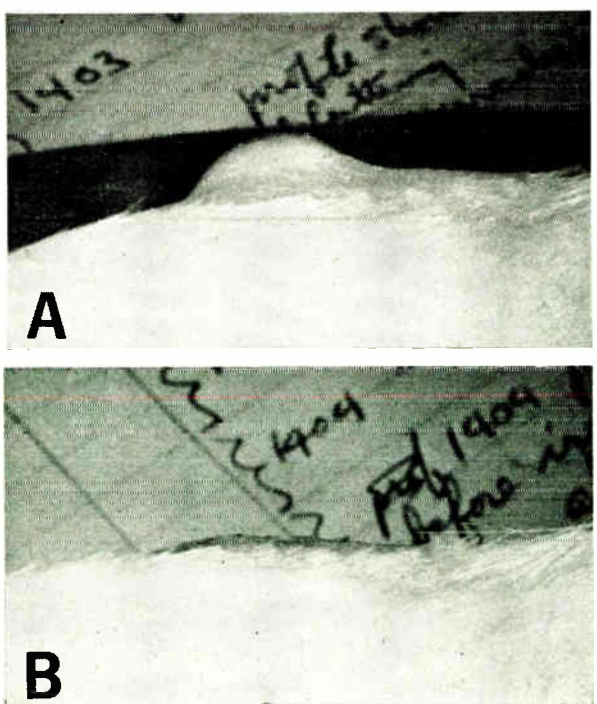
period. Thus in some experiments[44] the hair over sharply circumscribed circular areas, as it grew back at the sites of laser irradiation, was not pigmented.

The superficial lesions were smaller but more severe with focused than with nonfocused radiation. No definite conclusions, however, could be reached in regard to differences between the overall effects of focused vs. nonfocused irradiation. Several aspects, such as spread of the focused beam as it passes beyond the focal plane and factors other than thermal, require consideration in analyzing the overall effects of focusing.

Since there are differences in the interaction of laser radiation with different types of metal, differences between the interactions with tissues and those with metals were not unexpected.[42] Some of the factors of importance are: absorptivity and reflectivity of the tissue; heterogeneity of the tissue; the high water content of tissue with its high heat capacity; and the thermal conductivity of tissue. As the energy is probably dissipated in a relatively short period, the circulating blood does not appear to participate significantly in its dissipation, although the distribution of the blood supply may be of consequence in this respect.

Studies have been made of injuries to the skin and underlying structures at high peak powers (exceeding 100 MW) and high peak-power densities (exceeding $10^8$ watts per square millimeter) at energy levels of approximately 3 joules delivered in less than 10 ns.[44] Skin, underlying connective tissues, muscles, and viscera (liver, intestine) were injured. Uninjured layers of peritoneum and connective tissue intervened between the lesions, as had been observed at 30 to 100 joules delivered at pulses of about 1 ms duration.[42,43] At lower peak powers (approximately 10 kW) and energy levels of 7 joules (corresponding to pulse duration of 1 ms) considerably less damage or no damage to the skin was observed. These studies indicate that the peak power (and peak-power density) is of importance in biological interactions as well as in interactions with physical systems.

Fig. 4. A—Profile of tumor preirradiation. B—Profile of tumor immediately postirradiation at level exceeding 500 joules.



Studies at pulse repetition frequencies exceeding one pulse per second have been carried out.[29,30,35] The effects do not appear to be quantitatively similar when the same energy (of the order of 300 to 500 joules) is delivered at pulse-repetition frequencies of one pulse per second (several joules per pulse) as compared to delivery of this energy as a single pulse of 1 ms duration. One of the reasons for this may be that interaction of one pulse with tissue tends to alter the tissue in such a manner that the absorption of radiation during the folowing pulse differs from that of the previous one.[29]

Studies of the effects of single pulses of laser radiation at 6943 Å and energy levels up to 5 joules (pulse duration approximately 1 ms) on skin have been reported.[45,46] These studies were carried out on normal skin and superficial hemangiomas of man and on the skin of rabbits. Intact animals studied[46,47] included Mexican beetles and other insects. In humans and rabbits, exudation of fluid in the irradiated area and some alteration in blood flow patterns was reported. There was also some loss of pigment.

Investigation has been made of the effects of single pulses at 6943 Å and at energy levels ranging from approximately 30 millijoules to 100 joules on hemostasis and skin flaps in normal mice.[30,35] Irradiation of skin flaps, in which adequate circulation was maintained, resulted in the production of intravascular thrombosis. The degree of injury to the blood vessel wall varied from nondetectable to frank interstitial hemorrhage. Factors affecting these results included direction of the beam, energy level, spot size, and size of the artery or vein. With the laser unit focused through a microscope, thrombosis (clotting within the blood vessel), followed by partial to complete blood-vessel obstruction was observed in the mice.

Hemorrhage produced in the tail was studied at energy levels in the 50 joule range.[30] Defocused radiation produced hemostasis (arrest of hemorrhage) in heparinized animals, while unfocused radiation at these energy levels did not stop the bleeding.

**Experimental tumors.** Effects of interaction of radiation with tumors as well as with normal tissue may be direct or indirect, primary or secondary. Some direct effects may be a result of ionization and excitation within molecules of the cells of the tumor itself. Indirect effects may be a result of transformations brought about by active radicals produced in the surrounding media. Secondary effects on tumors can result from interference with the blood supply, alteration of defense mechanisms of the animal, effects on supporting connective tissue, substances released by degenerating cells, and radiosensitivity of the tumor.

Initial studies of the effects of laser irradiation on experimental tumors in the cheek pouch of Syrian hamsters[41,42] and in mice,[53] have been reported. Further studies[48] employed fractionation of dose (over several weeks), high pulse-repetition frequency and continuous irradiation as well as single pulses. These studies were carried out at 6328 Å, 6943 Å, and 10600 Å. Energy levels ranging from 5 to 500 joules per pulse, peak power levels exceeding 100 MW, and pulse-repetition frequen-

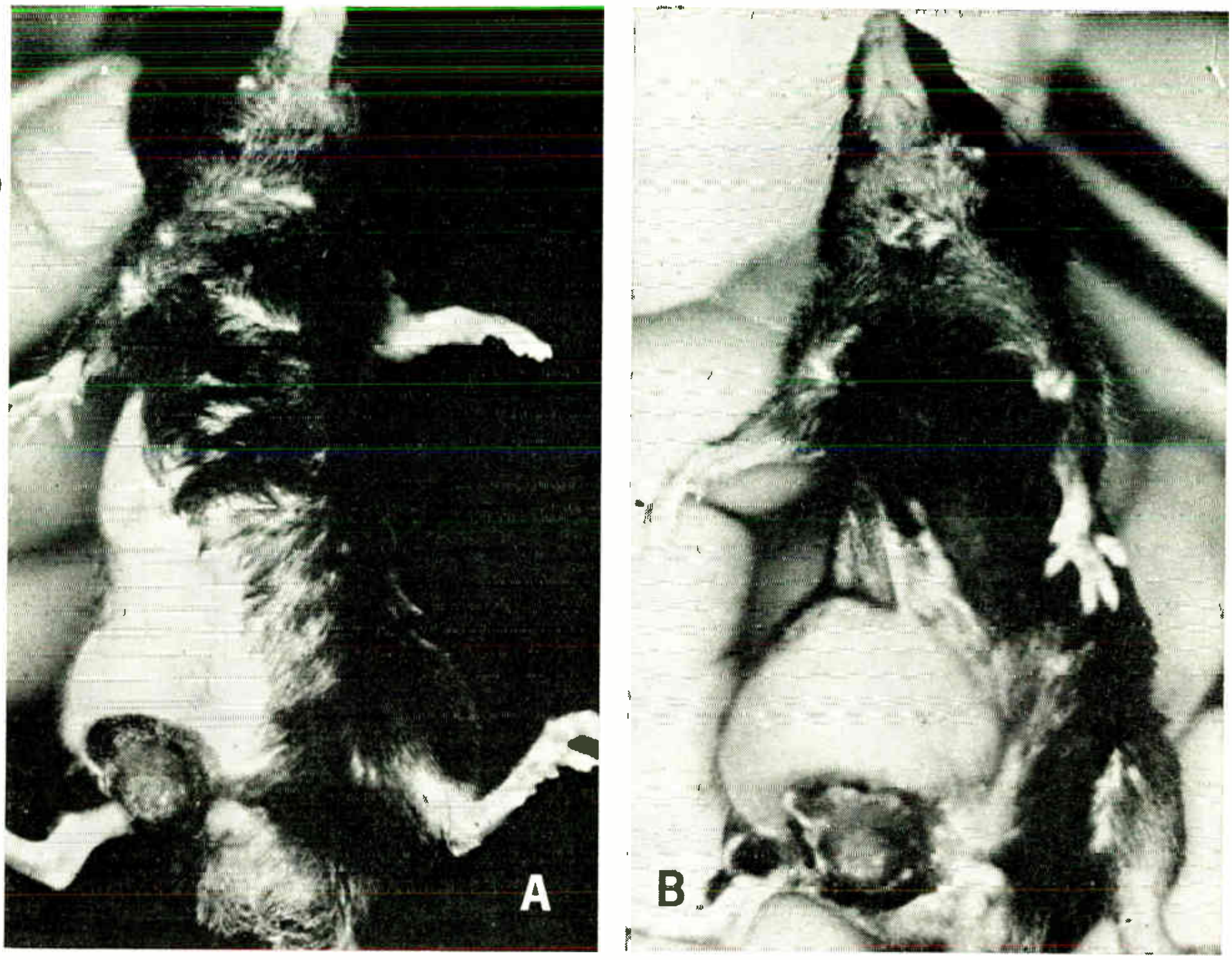


Fig. 5. A—Tumor (Lewis bladder) three days postirradiation at pulse-repetition frequency of one per second. Total energy greater than 300 joules. Note decreased growth in region of irradiation; B—Tumor 16 days postirradiation, showing considerable tumor growth, particularly in nonirradiated area. Animal died three weeks postirradiation.

cies of one per second were employed. The tumors studied included melanomas (Harding-Passey, Cloudman), osteogenic sarcomas (Ridgway) and bladder carcinomas (Lewis) in mice, and HAS no. 1 hamster sarcomas transplanted to the cheek pouch of Syrian hamsters. Preliminary observations indicated that the effects of laser radiation on tumor-bearing animals varied from insignificant to apparently complete regression on the one hand and exacerbations on the other. The characteristics of the laser radiation, focusing, spot size, and the diameter of the cross section of the beam were some of the physical factors which determined the effects on tumors. Biological factors were the type of tumor, time between implantation and irradiation, size and degree of dissemination, general state of the animal, pigmentation, blood supply, and location of the tumor. Small tumors, which were superficially located and had not invaded the surrounding tissues, could be destroyed entirely (Fig. 4). Biopsies were obtained from the sites of irradiated Harding-Passey melanomas and Ridgway sarcoma several weeks after the nonirradiated controls had died. There was no evidence of tumor cells, although histiocytes containing pigment were found in the specimens obtained from albino mice. Complete regressions, however, can occur spontaneously in transplanted tumors, and more extensive studies are required to establish statistical significance.

With larger tumors, if the irradiation was limited to a part of the tumor, the tumor was not destroyed entirely and the remaining viable cells continued to grow and eventually led to death (Fig. 5). Statistical evaluation is not possible yet to determine whether partial irradiation of experimental tumors can prolong life expectancy, or whether irradiation of large advanced tumors leads to earlier death.

The degree of damage to the tumors increased with the energy. The extent of the effects did not bear a linear relationship to the total energy, whether delivered as a single pulse or as a series of pulses, delivered rapidly or fractionated over a period of weeks.

While the total energy was of significance, the peak power and the power density appeared to be of considerable importance in the effects on tumors.[44] Single pulses at 100 MW and 3 joules produced considerably more damage in melanomas (as well as in normal tissues) than did single pulses at 10 kW and 7 joules.

At energy levels above about 100 joules or at power levels above 100 MW, however, the primary damage caused by the direct interaction of the unfocused radiation with the tumor tissue is sufficient to produce secondary reactions which result in delayed extension of the injury to beyond the irradiated part. Results indicated that a larger section of the tissue may be affected by unfocused (or defocused) than by focused radiation. The total final damage produced by the unfocused (or defocused) beam may be more extensive than that produced by the focused beam. If the threshold for irreversible injury is exceeded, the focused radiation may result in charring and total destruction of the irradiated part alone, thus expending the energy on a process which is well beyond the level required for irreversibility.

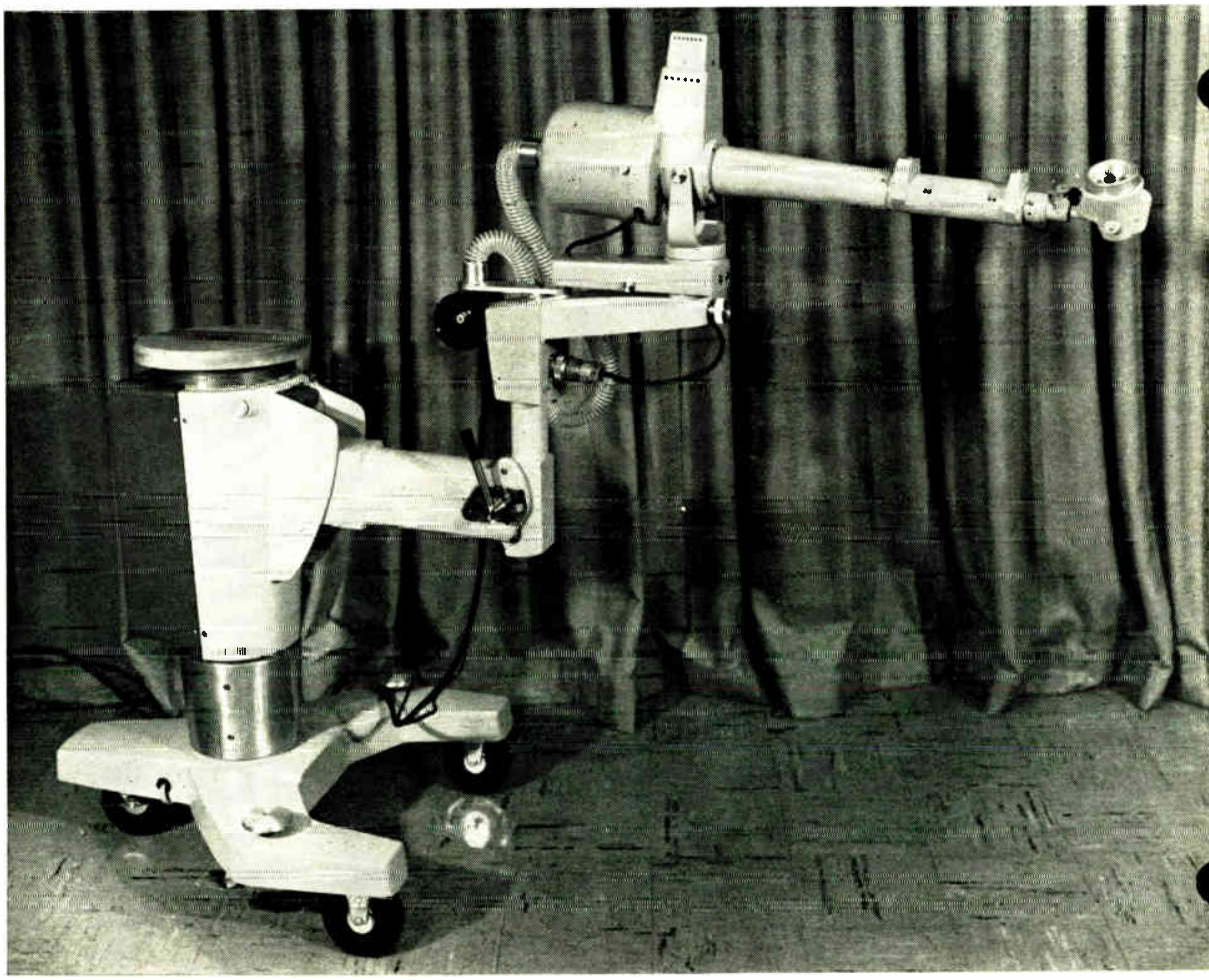
Biological factors affecting the interaction include pigmentation, blood supply, and physiological status of the animal. The severity of the lesion is increased in the presence of pigments that absorb the radiation. In order to assess this factor, highly pigmented tumors (melanomas) containing large amounts of dark brown pigment (melanin) were studied. It was noted that the lesions in the skin overlying the tumors were considerably less extensive than the damage to the tumors.

Blood supply affects the extent of the radiation-induced lesion. In tumors the blood supply is frequently characterized by a shell of dilated tortuous blood vessels which surround the tumor. The interaction causes damage to the blood vessels, which can in turn lead to an inadequate blood supply and ensuing death of those tumor cells that were not lethally injured by direct interaction with the laser radiation.

The general status of the tumor-bearing animal requires consideration of the decreasing life expectancy by laser irradiation. In the late stages of the tumor, the animal is considerably debilitated. Additional trauma and induction of (surgical) shock at the time of irradiation may result in shortening the life of the tumor-bearing animal. The sequelae of injury, such as necrosis, inflammation and infection, may further accelerate death.

A third possible reason for decreasing life expectancy is that the impact of the radiation may result in dissemi-

Fig. 6. Laser light coagulator of the American Optical Company. Patient is positioned under extreme right end of instrument, where a mirror is used to direct laser beam downward into eye. Physician views from above. Ruby laser is contained in cylindrical housing at opposite end of instrument. Hose attached to this housing provides air cooling. Aiming light is in vertical column and light from this source is projected along same path that laser beam follows through the instrument and into the eye. The power supply is a separate unit, not shown.



nation of tumor cells, which were not destroyed. Earlier and wider dissemination of the tumor may thus occur than in the natural course of the disease. This may result from direct extension into adjacent normal tissues, leading to their invasion by tumor cells, or by the entry of tumor cells (induced by the impact of the radiation on the tissue) into the blood or lymph vessels, whence the tumor cells are carried to distant sites in the body to set up metastatic growths.

It has been noted that growth of tumor cells continued at the periphery of the irradiated areas and proceeded in a direction away from the irradiated (and necrotized) area.[43, 48] Thus the adjacent normal tissues were invaded, while the residual debris of the tumor did not support its recurrence at the irradiated site. In tumors that have undergone spontaneous necrosis, it is also found that the necrotized areas are not invaded, although the tumor continues to grow centrifugally.

The microscopic changes were in agreement with the gross observations. Two types of necrosis were observed. One type, caused by primary interaction of the radiation with tumor, proceeded along sharply demarcated straight lines, with an abrupt transition from viable (nonirradiated) to nonviable (irradiated) cells. This type of necrosis resembled that found in irradiated nonneoplastic tissues. The second type of necrosis was secondary to the primary interactions, focal in distribution with gradual transition from nonviable cells through less severely damaged cells to viable cells, and without sharp linear demarcation. This type resembled the spontaneous necrosis of tumors, which occurs when the rapid proliferation of tumor cells has exceeded the capacity of the blood supply to support the respective areas of the tumor. Experimenters have irradiated Pitt-42 tumors; melanomas transplanted to the cheek pouch of the hamster; and fibrosarcomas, originally induced by methylcholanthrene and transplanted to the flank of the hamster.[49] Studies were carried out at 0.5 joules to 360 joules per pulse at 6943 Å. From 60 to 80 per cent of the methylcholanthrene-induced fibrosarcomas were destroyed. The cheek pouches containing the melanomas and Pitt-41 tumors were reported free of viable tumor on gross and microscopic studies when examined from the 14 to the 39 day. It appears, they concluded, that radiation at 6943 Å has a selective effect upon tumors, since the treated tumor was destroyed, while adjacent normal tissue was relatively unaffected. Studies have been reported on the irradiation of excised human basal cell carcinomas,[47, 50] with and without prior staining.

#### Hazards of laser radiation on eyes

Damage to the eyes caused by exposure to electromagnetic radiation is well documented.[51, 8] Retinal burns have been receiving increasing attention in regard to atomic explosions,[52] in the exploration of extraterrestrial space,[53] and the investigations involving laser design, development, and utilization.

Damage to the eyes, either from exposure to the laser beam or to backscattered radiation and volatilized material[43, 51, 54–56] resulting from the interaction of the beam with the target is a distinct hazard. The chorioretinal layer is particularly vulnerable because of its pigmentation, the focusing of the beam on it by the focusing mechanism of the eye, and the transparency of the ocular constituents anterior to the retina.

Initial calculations regarding the hazard to the eyes has been discussed.[54] Production of experimental ocular lesions with laser units have been carried out.[55, 57–62] Koester obtained retinal lesions at 6943 Å at an energy output of 0.04 joules. Chorioretinal lesions were obtained by Zaret using a 0.1-joule ruby laser, directed at both nonpigmented (albinotic) and irregularly pigmented retinas of rabbits. He obtained an approximate threshold dose for retinal damage of 2.5 to 25 joules per square centimeter for a lightly pigmented retina. Experimental iris lesions were also produced by Zaret.[57] A preliminary study of hazard-evaluation exposure of eyes of anesthetized mice[29] to an unfocused ruby beam with an output of about 100 joules resulted in complete destruction of the eye, with hemorrhage evident in the remaining socket.

Although 95 per cent of the light at 6943 Å is transmitted by the transparent ocular media (cornea, anterior chamber, lens, vitreous body), and interacts with the retina and choroid,[60, 61] the long-term effects of absorption of the remaining 5 per cent of the energy by the ocular media has not been assessed. Damage may be produced by the back-scattered radiation, which may be at other than the primary wavelength, and by the back-scattered volatilized material produced at the site of interaction of the radiation with the retina and choroid. At longer wavelengths, attenuation by the ocular media increases[60, 61] and interaction with the ocular media due to incident infrared, and consequently invisible, radiation may be of considerable significance.

Hazards due to light output from the flash tubes used to pump the crystal must also be considered. Although the energy density and power density of the flash system is much less than that of the crystal, the light output often exceeds 2000 joules. During experimental testing of flash systems, damage to the conjunctiva and superficial layers of the eye has occurred when flash lamps were tested without shielding and suitable protective glasses were not used.

To limit the hazards to the eyes, several techniques have been considered and reported. Studies have been made of the protection offered by glasses,[51] and investigations made on BG-18 color filter glass, which possesses high absorption extending from the ruby wavelength into the infrared, including the Nd-doped laser wavelength with adequate transmittance in the remaining portion of the visible spectrum. To prevent cracking of the glass when it is exposed to direct impact at 200 joules, a high-pass dichroic coating of the front surface, with high reflectivity at ruby and neodymium laser wavelengths, was added. The American Optical Company, Bausch and Lomb Company, and The Technical Research Group each have developed glasses to provide mechanical strength and wavelength attenuation. A closed-circuit television system to limit exposure to direct or reflected radiation has been used.[56] Studies of effects of pulsed unfocused and focused irradiation of the forehead in mice have been carried out.[63] Unfocused irradiation was followed by death within less than 30 seconds in 10 out of 23 animals irradiated. Of 41 animals irradiated, 31 died within 24 hours of irradiation. Although there was damage to the overlying skin and to the underlying brain, there was no gross penetration through the intervening skull. They concluded that effects were due to transmission of a shock wave, rather than due to direct thermal effects, and that effects could be more severe at a deeper level than those produced in the tissue at the site of irradiation. General as well as opthalmic hazards of laser radiation[64] have been comprehensively discussed.

### Clinical studies

There are a number of indications for the use of light coagulation in the treatment of retinal lesions.[65] These indications, which may also apply to laser coagulation, are holes in the macula (area of central vision) without gross detachment, peripheral retinal tears without detachment, peripheral retinal changes in eyes predisposed to retinal detachments, retinal detachments with little or no elevation, blood vessel abnormalities, tumors of the eye including melanoma and retinoblastoma, and an abnormal position of the iris which interferes with vision after removal of a cataract. Inasmuch as only a proportion of patients to whom these conditions apply will benefit from light coagulation, careful evaluation of the indications is necessary.

Laser units are being specifically modified or designed for ophthalmic light coagulation. Possible advantages of these units over the nonlaser type of coagulator include: potential selection of the most effective wavelength for treatment; ability to focus the beam to a small spot size, thus achieving high energy density over the desired area with a reduced energy density through the remainder of the eye tissue; use of wavelengths other than that of the laser for constant visualization of the target; and control of energy, power, and power density. Units for clinical application are being developed by Schepens, Havener, and Abraham in collaboration with Franks of Maser Optics Inc., Campbell in association with Koester of the American Optical Co., Zaret in conjunction with the Technical Research Group, and Flocks and Zweng collaborating with Kapany of Optics Technology. The Maser Optics unit uses an indirect stereoscopic ophthalmoscope for observation and illumination. Intensity is controlled by means of filters, in order to maintain consistency of the desired energy. Units developed by the American Optical Co. (Fig. 6) and by the Technical Research Group employ similar general principles. The unit developed by Optics Technology is a hand-held direct ophthalmoscope laser unit. The various units operate at 6943 Å. These groups have carried out extensive investigations on animals.

Therapeutic retinal coagulation in a small number of human subjects has been reported.[58] Fiber optics systems in conjunction with laser photocoagulators to reach otherwise inaccessible areas have been considered.[62]

The skin of volunteers has been exposed to laser radiation at energy levels up to five joules.[46] The irradiated areas were examined under the skin microscope, excised, and subjected to routine histological and histochemical studies.[31] Under investigation are the use of laser light in a fiber bundle within a cardiac catheter to provide illumination for motion pictures inside the beating heart. Clinical uses in tumor therapy are under consideration.[64]

### Summary and conclusions

Information on interactions of laser radiation with biological systems and on effects resulting from these interactions is limited and largely preliminary at this time. Laser devices have been available for a relatively short time and the complex nature of the area under study requires the close collaboration of physical and biological scientists. Present investigations include studies on macromolecular preparations; single cells; normal animals with emphasis on eyes and skin; tumors; the production of free radicals; and application of spectroscopy to biological systems. Definite hazards to the eyes exist. Long-term hazards to other structures cannot be ruled out at this time.

Exploration of laser devices for application to biology, including microscopy and microsurgery have been discussed,[66] and studies on biological effects of lasers and their implications reviewed.[64] Hazards are discussed and the possible use of fiber optics systems suggested for the transmissions of laser radiation for the diagnosis and treatment of deep-seated pathological tissues, as well as for cauterization and specialized surgical procedures.

In laser devices the major potential clinical application appears to be in the retinal coagulator (photocautery unit), when used in conjunction with good clinical evaluation. Irradiation of the skin or of tumors in man should be guided by further basic investigations, before clinical studies are attempted, since complications including tumor spread, bleeding, irreparable damage to normal structures, and long-term detrimental effects may occur. Studies of physiological processes related to the visual process are being considered. Determination of the threshold for eye damage on both short- and long-term basis must be carried out prior to studies on man.

Although investigations up to now have been carried out primarily with crystal (or glass) systems and with gaseous lasers, semiconductor lasers may prove of interest for basic investigation and clinical applications. Laser operation with semiconductor units has been attained at room temperature. Electrical pumping offers good control. Should suitable energy and power levels be obtained at room or body temperatures, the relatively small size of these units may permit them to be readily manipulated for application to the eyes and skin, and may allow intracavity and intravascular insertion for purposes of photography, diagnosis, and therapy.

Other areas due for attention include disinfection and sterilization, particularly at ultraviolet wavelengths, under usual conditions or for special purposes, such as the space program; synthesis of macromolecules of biological interest from smaller components; and possible existence of communication systems basic to communication between organisms of the same species using modulated coherent electromagnetic radiation.

#### REFERENCES


1. Blair, *Biological Effects of External Radiation*, New York: McGraw-Hill Book Co., Inc., 1954.
2. Claus, W. D., *Radiation Biology and Medicine*, Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1958.
3. Law, L. W., "Radiation Carcinogenesis," *Advances in Bio. and Med. Phys.*, vol. 7, 1960, pp. 295–342.
4. Harris, R. J. C., *The Initial Effects of Ionizing Radiation in Cells*. New York: Academic Press, Inc., 1961.
5. Hollaender, A., *Radiation Biology*. New York: McGraw-Hill Book Co., Inc., 1955.
6. Blum, H. F., *Photodynamic Action and Diseases Caused by Light*. New York: Reinhold Publishing Corp., 1941.
7. Lorincz, A. L., "Physiological and Pathological Changes in Skin from Sunburn and Suntan," *J. Am. Med. Assoc.*, vol. 173, July 1960, pp. 1227–1231.
8. Friedenwald, J. S., *Ophthalmic Pathology*. Philadelphia, Pa.: W. B. Saunders Co., 1952.
9. Fitzpatrick, T. B., M. A. Pathak, I. A. Magnus, and W. L.

Curwen, "Abnormal Reactions of Man to Light," *Annual Rev. Med.*, vol. 14, 1963.

10. Slager, U. T., *Space Medicine*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962.

11. Knauf, G., "The Bio-Effects of Radar Energy," *Aerospace Med.*, vol. 31, March 1960, pp. 225–238.

12. Maiman, T. H., "Stimulated Optical Radiation in Ruby," *Nature*, vol. 187, 1960, p. 493.

13. Maiman, T. H., "Optical and Microwave Optical Experiments in Ruby," *Phys. Rev. Letters*, vol. 4, June 1, 1960, pp. 564–566.

14. Maiman, T. H., "Stimulated Optical Emission in Fluorescent Solids: I. Theoretical Considerations," *Phys. Rev.*, vol. 123, 1961, p. 1145.

15. Yariv, A., and J. P. Gordon, "The Laser," *Proc. IEEE*, vol. 51, Jan. 1963, pp. 4–29.

16. Snitzer, E., "Optical Maser Action of $Nd^{3+}$ in a Barium Crown Glass," *Phys. Rev. Letters*, vol. 7, Dec. 1961, pp. 444–446.

17. Javan, A., W. R. Bennett, and D. R. Harriott, "Population Inversion and Continuous Optical Maser Oscillation in Gas Discharge Containing a He-Ne Mixture," *Phys. Rev. Letters*, vol. 6, Feb. 1961, p. 106.

18. Hall, R. N., G. E. Fenner, J. D. Kingsley, T. J. Soltys, and R. O. Carlson, "Coherent Light Emission from GaAs Junctions," *Phys. Rev. Letters*, vol. 6, Feb. 1961, p. 106.

19. Nathan, M. T., W. P. Dumke, G. Burns, F. H. Dill, and G. Lasher, "Stimulated Emission of Radiation from GaAs p-n Junctions," *Applied Phys. Letters*, vol. 1, Nov. 1962, p. 62.

20. Quist, T. M., R. H. Rediker, R. J. Keyes, W. E. Krag, B. Lax, A. L. McWorter, and H. J. Zieger, "Semiconductor Maser of GaAs," *Ibid.*, Dec. 1962, pp. 91–92.

21. McClung, F. J., and R. W. Hellwarth, "Giant Optical Pulsations from Ruby," *J. App. Phys.*, vol. 33, no. 3, Mar. 1962, pp. 828–829.

22. Gillmer, A. H., and H. Schultz, "500 Megawatt Pulsed Laser," presented at the 8th Annual Technical Symposium of Society of Photographic Instrumentation Engineers, Los Angeles, Calif., Aug. 5–9, 1963.

23. Franken, P., A. E. Hill, C. W. Peters, and G. Weinreich, "Generation of Optical Harmonics," *Phys. Rev. Letters*, vol. 7, 1961, pp. 118–119.

24. Porto, S. P. S., and D. L. Wood, "Ruby Optical Maser as Raman Source," *J. Opt. Soc. Am.*, vol. 52, 1962, p. 251.

25. Commoner, B., J. Townsend, and G. Pake, "Free Radicals in Biological Materials," *Nature*, vol. 174, Oct. 9, 1954, pp. 689–691.

26. Sogo, P. B., and B. M. Tolbert, "Nuclear and Electron Paramagnetic Resonance and its Application to Biology," in *Advances in Bio. and Med. Phys.*, vol. 5, 1957, pp. 1–35.

27. Mason, H. S., D. J. E. Ingram, and B. Allen, "The Free Radical Property of Melanins," *Arch. Biochem. Biophys.*, vol. 86, 1960, p. 225.

28. Derr, V., E. Klein, and S. Fine, "Electron Spin Resonance Tests of Laser Irradiated Biological Systems," *Appl. Optics* (submitted for publication).

29. Fine, S., E. Klein, J. Ambrus, E. Cohen, V. Derr, and C. Ambrus, "Interaction of Relatively Coherent Laser Radiation and Biological Systems," *Fed. Proc.*, vol. 23, 1964.

30. Klein, E., S. Fine, E. Cohen, J. Ambrus, E. Neter, R. Lyman, and R. E. Scott, "Effects of Laser Radiation on Biological Systems," presented at American College of Physicians Meeting, Atlantic City, N. J., Apr. 10, 1964.

31. Malt, R. A., C. H. Townes, "Optical Masers in Biology and Medicine," *New Eng. J. Med.* vol. 269, p. 1417, 1963.

32. "Laser Aids Study of Effects of Radiation on Living Cells," *Electron. Design News*, 1962.

33. Bessis, M., F. Gires, G. Mayer, and G. Nomarski, "Irradiation des Organites Cellulaires a Paide d'un Laser a rubis," *Compt. Rend. Acad. Sci.*, vol. 255, 1962, pp. 1010–1012.

34. Saks, N. M., and C. A. Roth, "A Ruby Laser as Microsurgical Instrument," *Science*, vol. 141, 1963, p. 46.

35. Klein E., S. Fine, V. Laor, J. Donaghue, and L. Englander, "Laser Irradiation of the Skin," *Proc. Soc. Invest. Dermatol.* (accepted for publication), 1964.

36. Rounds, D. E., and E. C. Chamberlain, "Preliminary Observations of the Effects of Laser Irradiation on Cells Grown in Tissue Cultures," presented at the Second Boston Laser Conference, Mass., Aug. 1, 1963.

37. Brech, F., and L. Cross, "Optical Micromission Stimulated by Ruby Maser," *Appl. Spectroscopy*, vol. 16, no. 2, 1962, p. 59.

38. Rosan, R. C., M. K. Healy, W. T. McNary, Jr., "Spectroscopic Ultramicroanalysis with Lasers," *Science*, vol. 142, 1963, p. 236.

39. Fine, S., E. Klein, F. Brech, W. T. McNary, Jr. (unpublished data).

40. Fine, S., E. Klein, R. E. Scott, R. Seed, and A. Roy, "Some Effects of Laser Radiation on the Skin of the Syrian Hamster," *Life Sci.*, no. 1, Jan. 1963.

41. Fine, S., E. Klein, S. Farber, R. E. Scott, A. Roy, and R. E. Seed, "In Vivo Effects of Laser Radiation on the Skin of the Syrian Hamster," *J. Invest. Dermatol.*, vol. 40, 1963, p. 123.

42. Fine, S., E. Klein, R. E. Scott, and R. Seed, "Laser Radiation in the Syrian Hamster," *Skin*, vol. 2, 1963, p. 43.

43. Fine, S., E. Klein, R. E. Scott, C. Aaronson, and J. Donoghue, "Biological Effects of Laser Radiation," presented at the Second Boston Laser Conference, Aug. 1, 1963.

44. Fine, S., T. H. Maiman, E. Klein, and R. E. Scott, "Biological Effects of High Peak Power Radiation," *Life Sci.*, Mar. 1964.

45. Goldman, L., D. J. Blaney, D. J. Kindel, E. K. Franke, "Effect of the Laser Beam on the Skin," *J. Invest. Dermatol.*, vol. 40, 1963, p. 121.

46. Goldman, L., D. J. Blaney, D. J. Kindel, D. Richfield, and E. K. Franke, "Pathology of the Effect of the Laser Beam on the Skin," *Nature*, vol. 197, 1963, p. 912.

47. Goldman, L., "Comparison of Biological Effects of Low and High Energy Lasers," presented at the Second Boston Laser Conference, 1963.

48. Klein, E., S. Fine, R. E. Scott, and S. Farber, "Observations on Laser Irradiation of Experimental Tumors," *Proc. Am. Assoc. Cancer Res.*, 1964 (accepted for publication).

49. McGuff, P. E., E. Bushnell, H. S. Soroff, and R. A. Deterling, "Studies of the Surgical Applications of Laser (Light Amplification by Stimulated Emission of Radiation)," *Surg. Forum, Proc.*, vol. 14, 1963, p. 143.

50. Goldman, L., "Protection of Personnel Operating Lasers," *Am. J. Med. Electron.*, vol. 2, 1963, p. 335.

51. Straub, H. W. "Protection of the Human Eye from Laser Radiation," Report TR-1153, Harry Diamond Laboratories, Army Material Commands, Washington 25, D. C., July 10, 1963, p. 15.

52. Geeraets, W. J., W. T. Ham, Jr., and D. Guerry, III, "Retinal Burns," presented at the Annual Meeting A.M.A., Atlantic City, N.J., June 16–21, 1963.

53. Severin, S. L., N. L. Newton, and J. F. Culver, "An Experimental Approach to Flash Blindness," *Aerospace Med.*, vol. 33, 1962, p. 1199.

54. Solon, L. R., R. Aronson, and G. Gould, "Physiological Implications of Laser Beams," *Science*, vol. 134, 1961, p. 1506.

55. Zaret, M. M., G. M. Breinin, H. Schmidt, H. Ripps, I. M. Siegel, and L. R. Solon, "Ocular Lesions Produced by an Optical Maser (Laser)," *Science*, vol. 134, 1961, p. 1525.

56. Fine, S., E. Klein, C. Aaronson, G. Hardway, W. King, and R. E. Scott, "Closed Circuit Television in Laser Investigation," (accepted for publication).

57. Zaret, M. M., H. Ripps, I. M. Siegel, and G. M. Breinin, "Laser Photocoagulation of the Eye," *A.M.A. Arch. Ophthalmol.*, vol. 69, 1963, p. 97.

58. Koester, C. J., E. Snitzer, C. J. Campbell, and M. C. Rittler, "Experimental Laser Retina Coagulator," *J. Opt. Soc. Am.*, vol. 52, 1962, p. 607.

59. Campbell, C. J., M. C. Rittler, and C. J. Koester, "The Optical Maser as a Retinal Coagulator: An Evaluation," *Trans. Am. Acad. Ophthalmol. Otolaryngol*, vol. 67, 1963, p. 58.

60. Ham, W. T., "Ocular Radiation Effects," presented at the Second Boston Laser Conference, Aug. 1, 1963.

61. Geeraets, W. J., R. E. Williams, G. Chan, W. T. Ham, Jr., D. Guerry III, and E. H. Schmidt, "The Loss of Light Energy in Retina and Choroid," *Arch. Ophthalmol.*, vol. 64, 1960, p. 606.

62. Schepens, and Freeman, H. "Studies on Retinal Coagulation," presented at the Conference on Biological Effects of Laser Radiation, Medical Physics Society at M.I.T., Boston, Mass., 1963.

63. Fine, S., and E. Klein, "Effects of Pulsed Laser Irradiation of the Forehead in Mice," *Life Sci.* (accepted for publication).

64. Litwin, M. and D. Glew, "Biological Effects of Laser Radiation," *J. Am. Med. Assoc.*, (accepted for publication).

65. Meyer-Schwickerath, G., *Light Coagulation*. St. Louis: C. V. Mosby Co., 1960.

66. Townes, C. H., "Optical Masers and their Possible Application to Biology," *Biophys. J.*, vol. 2, pt. 2, March 1962.

This article is based on research supported by Research Contracts DA-49-193-MD-2436 and DA-49-193-MD-2437 from the Surgical Research Branch, U.S. Army Medical Research and Development Command, Office of the Surgeon General, U.S. Department of the Army.

# Regional integration of electric power systems

*The development of EHV and supersized generators have made economically feasible a 7-gigawatt power pool combining the power systems of 10 states and 22 utilities*

*Earl Ewald, D. W. Angland* *Northern States Power Company*

Integration of power systems means the development and use of interconnections between power systems so that their combined loads and resources can be treated as a unit system to achieve the advantage of large-scale operation.

The history of the power industry has been one of steadily expanding integration. At the beginning, small individual community systems were interconnected for combined supply from the early central stations. As the art of generating and transmitting electric power improved, high-voltage transmission networks were gradually developed and further consolidations evolved to form today's modern power systems with large load areas supplied from relatively few favorably located power plants.

These individual power systems, in turn, have been interconnected with their surrounding neighbors to carry out integrated planning and operations on a multisystem basis through power pooling contracts or other forms of interconnection agreements. Recent technological advances in the fields of extra-high-voltage (EHV) transmission and supersized generators offer economic incentive for intersystem coordination on an even broader basis, and the industry is now on the threshold of another dramatic step in its integration advances.

This article reviews the economic fundamentals of large-scale power system planning and reports on specific regional integration plans that have been developed for the north central states.

### Generator cost characteristics

Opportunities for more effective use of large base-load generating units are an important incentive for expanding integration today. Figs. 1 and 2 show how generator investment and operating costs decrease with unit size. These graphs represent average values of estimated costs that have been developed by a number of utilities, and the costs apply to coal-fired plants in the northern section of the country. Costs for any particular generating unit may, of course, depart substantially from these typical values, depending upon the number of units in a single plant and other local conditions.

The solid-line curve of Fig. 1 represents capital costs, or the investment that goes into building the generating units. The operating and maintenance costs of Fig. 2 represent annual expenses for operating labor, materials, overhead, etc., but do not include fuel. The operating and maintenance expenses for a generator that operates at a reasonably high use factor tends to be a fixed amount each year.

For purposes of comparison, the annual operating and maintenance expenses of Fig. 2 have been converted to their equivalent capital or investment values and added to the actual estimated investments to produce the adjusted curve of Fig. 1. A 12.5 per cent carrying charge rate was assumed for this purpose; that is, a \$1.00 investment is assumed equivalent to an operating expense of \$0.125 per year and, conversely, an annual operating expense of \$1.00 is equivalent to \$1.00/0.125 or an \$8.00 investment.

The fuel efficiency of generating units improves slightly with size and offers some additional advantages in reduced operating costs. These advantages are not illustrated here.

The shape of the adjusted capital cost curve of Fig. 1 has special significance to integration. Note that the cost for a 100-MW unit is about \$190/kW, whereas the 200-MW unit cost is only \$160/kW. Hence, replacement of two 100-MW units with a single 200-MW unit would reduce costs \$30/kW for a total savings equivalent to a \$6 million investment. Note also that the curve flattens out and the percentage cost advantage diminishes with size. For example, the adjusted capital cost of a 600-MW unit is only \$4/kW less than the cost for a 500 MW unit.

In opposition to the economies of size just described, the use of larger generators increases the amount of spare or reserve capacity required to continue full service when one or more generators are out for repair. The economic selection of generator unit size involves effecting an economic balance between these two conflicting factors. This will now be demonstrated by solutions to some simple size selection problems for generators.

**Illustrative examples**

For purposes of illustration, consider that a new 1000-MW load is to develop in an isolated region, and be supplied by base-load generators with adequate spare capacity for high-quality utility-type service. The problem to be solved is to select the economical size and number of identical generators to be installed in this area.

The first step in the solution is to determine the

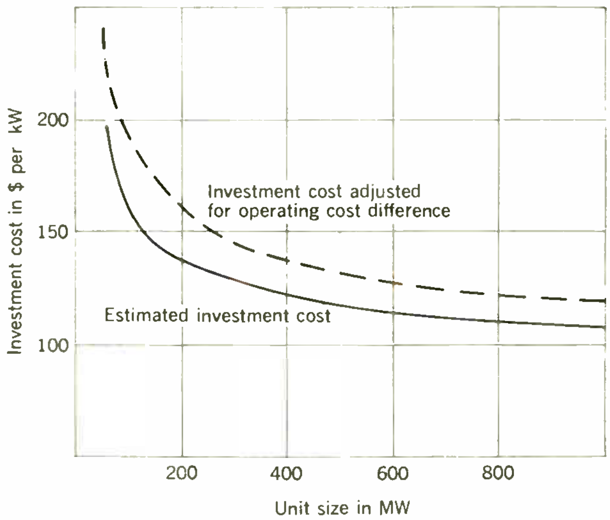
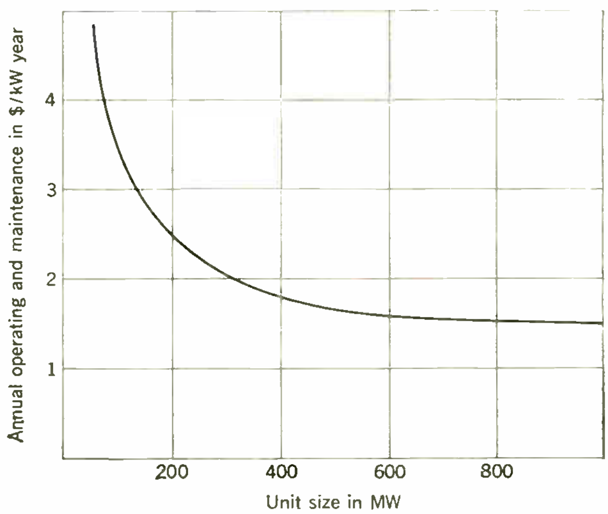



Fig. 1. Generator investment for unit sizes up to 1000 MW. Curves show how costs decrease with unit size.

Fig. 2. Generator operation and maintenance cost for unit sizes up to 1000 MW; costs do not include fuel.



**I. Number of reserve units in a hypothetical system of identical generators**

| Total Number of Identical Units | Number of Spare Units for Reserve |
|---|---|
| 3 to 8 | 2 |
| 8 to 19 | 3 |
| 19 to 32 | 4 |
| 32 to 49 | 5 |
| 49 and 50 | 6 |

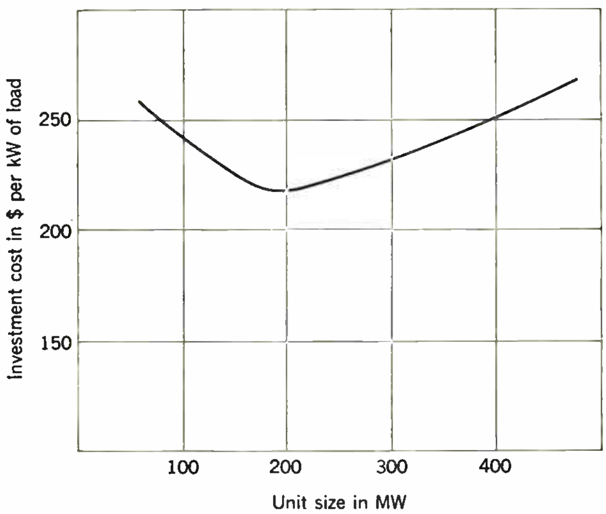
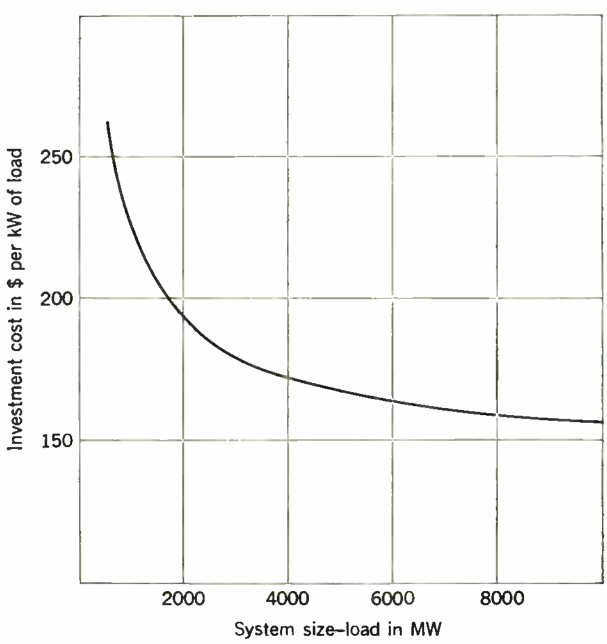



Fig. 3. Optimization curve of system generation cost for a hypothetical 1000-MW power system.

Fig. 4. Curve depicting the optimum generation cost for hypothetical power system of various sizes.




number of spare units that will be required for reserves. This can be handled by probability mathematics as explained in Appendix I Table I lists the number of spares calculated as necessary for reserves in systems of up to 50 identical units.

Table I and the cost data of Fig. 1 can now be used for determining the most economical unit size for the 1000-MW load by a cut-and-try process. As a first trial, assume a 13-generator system, which (from Table I) will require three spare units for reserve. Hence, the combined capacity of ten units must equal the load, and 100-MW units are required. The adjusted capital cost for 100-MW units is $190/kW, and the total cost for 13 units is $247 million or $247/kW of load supplied. Repetition of this process for other judiciously selected numbers of generators provides data for the optimization curve of Fig. 3. This curve shows that the minimum system generation cost would be $220/kW of load and would be attained with a unit size of approximately 200 MW.

Some insight into the influence of system size on economic unit size can be gained by calculating solutions to these simplified generator-size selection problems for a number of different system loads. For example, the optimization curve for a 2000-MW load shows an economic unit size of about 300 MW and a minimum system generation cost of $190/kW of load. Continuation of these calculations for higher loads reveals that system generation costs become relatively insensitive to unit sizes over a wide range. For assumed values of loads in these hypothetical power systems beyond 4000 MW, the decrease in power system generation costs with system size is almost wholly attributable to the percentage reductions achieved in reserve generating capacity.

The relationship between optimum generation costs and system size, as just calculated, is shown in Fig. 4. This graph can be used to extend the simplified generation planning problem into the area of integration. As previously cited, the equivalent capital cost for generation in a 2000-MW system is $190/kW as compared to the cost of $220/kW for the 1000-MW system. Hence, interconnection of two 1000-MW systems could reduce generation costs by $30/kW of load. To carry the process further, five 1000-MW systems might be interconnected into a single 5000-MW pool to attain generation costs of $170/kW, which is a $50/kW cost advantage as compared to isolated operation. These generation economies would, of course, have to be weighed against the cost of the integrating transmission system, and the problem of overall optimization would be critically dependent upon the geographical dispersion of the individual systems.

In addition to the transmission cost of integration, the simplified problems considered here neglect other complicating factors of importance in practical generation planning problems. It is emphasized that the solutions to these hypothetical problems only demonstrate the general nature of cost–size relationships and offer no basis for judgment as to appropriate generator sizes in any particular system.

Reference to an actual power system generation plan may be of value at this point. The Upper Mississippi Valley Power Pool, with which the authors are associated, consists of five investor-owned utilities and four rural electric cooperatives operating in Minnesota and adjacent areas of surrounding states. Load patterns across the pool vary from highly concentrated loads in metropolitan areas to extremely low load densities in sparsely settled rural areas. The current pool plan for adding 3000 MW of new base-load generating capacity during 1967 through 1975 proposes a mix of unit sizes varying from 200 to 600 MW with a weighted average size of about 400 MW. Concentration of generator additions in a few centrally located plants with minimum unit sizes of 500 MW was considered but found less economical. The lesser transmission expense for supply from strategically located 200-MW units in the lighter load areas more than offsets the cost advantages of the larger, centrally located units.

**General nature of integration benefits**

Areas of mutual benefits that are usually exploited in interconnection and pooling arrangements among electric power systems are the following: (1) shared generation reserves, (2) larger generating units, (3) staggered generator installations, (4) load diversity, (5) higher transmission voltages, and (6) the benefit of economy energy exchange.

As previously demonstrated, the first two benefits are interrelated in the general case. However, the nature of this relationship changes with size of the integrated operation, and influences pooling arrangements. In general, the smaller systems or pools with economical unit sizes along the steep portion of the generator cost–size curve of Fig. 1 will have greater percentage cost incentive for expanding integration. Also, at this pool-size level, a relatively high degree of integrated planning will be appropriate. In larger pools or systems where economic unit sizes are along the flat portion of the cost–size curve, the generation economies of further integration are limited to reserve reductions and can be accomplished by relatively simple arrangements for power exchanges during emergencies.

The advantage from staggering generator additions among pools or systems is again dependent upon relative size. In small pools, economical generator sizes may be equal to several years' load growth for an individual system and temporary exchanges of excess capacity are very important aspects of the pooling arrangements. In the case of transmission interconnections between larger pools or systems, the advantage of staggering generating units diminishes as the annual load growth of the individual pools or systems approaches their economic generator size.

Diversity in the times of day at which peak power demands occur in a group of systems may offer an opportunity for reducing generation capacity requirements through integration. Load diversity that can be dependably predicted results from intrinsic differences in daily and in seasonal load patterns, or both. For example, urban systems with large proportions of industrial load will generally have substantially different load patterns from those of systems that supply predominantly rural loads. The climatic differences in the North and South cause significant seasonal load diversity in certain regions. Extensive studies of load diversity made for a large region in the north-central states produced disappointing results with respect to the diversity benefits for integration on a regional basis. The existing subregional pools and interconnected groups effectively exploit daily load-pattern diversity, and the only significant regional diversity benefit found was a minor amount of seasonal diversity.

The per-kilowatt cost of power transmission decreases substantially as voltage levels increase if the full capacity of the higher voltage line can be fully utilized. An interconnected group of utilities may effect substantial economies through development of a common grid to transmit power from generating plants to load centers. In addition, the combined needs of transmission capacity for bulk power distribution and interconnection capacity for general integration benefits may permit economical use of higher levels of transmission voltage than could otherwise be used.

Costs of power generation in a pool may vary widely between plants and between individual generating units in a particular plant. The economic dispatch of total pool power production among these variable cost sources is the principal operating economy associated with pooling. The energy exchanged between systems in connection with this economic power production scheduling is called "economy energy." Although the saving from economy energy transactions may be large in comparison with the effort required to attain it, a number of interconnection studies indicate that economy energy saving is of minor importance in comparison with other integration benefits.

**MAPP regional grid plans**

The Mid-Continent Area Power Planners (MAPP) is a voluntary organization of 22 independent power suppliers established to promote integrated regional planning among its members. The member systems of MAPP, listed in Appendix II, serve, at the present time, a combined peak power load of 7100 MW, and operate through ten states and the province of Manitoba in Canada.

The MAPP plans for a 1980 regional 345-kV grid are shown in Fig. 5. The solid heavy lines represent proposed 345-kV constructions within the MAPP service area. Several future 230-kV constructions are included to indicate better the extent of the area covered. The 345-kV plans were developed in collaboration with nonmember utilities to the east and south of the MAPP area. The planned 345-kV constructions in adjoining areas that were of special importance to the MAPP plans are shown by the dashed lines in Fig. 5. The proposed MAPP regionally integrated grid is joined and somewhat overlapped on the west by the already extensive 230-kV transmission grid of the Federal Missouri River hydro system.

The MAPP systems that will be directly affected by the proposed grid now foresee a need for constructing about 14 000 MW of new base-load generating capacity by 1980. It is not difficult to demonstrate that economies in integrated generation can be expected to more than offset additional transmission costs for the grid at 1980 load levels. The real planning problem was to determine the sequence and timing of construction for the various segments of the grid.

The EHV lines in the higher load density eastern portion of the region have been found to be economical for construction first. The Twin Cities–Milwaukee–Chicago 345-kV line is now under construction and will be in service in 1966. This line will provide interconnec-

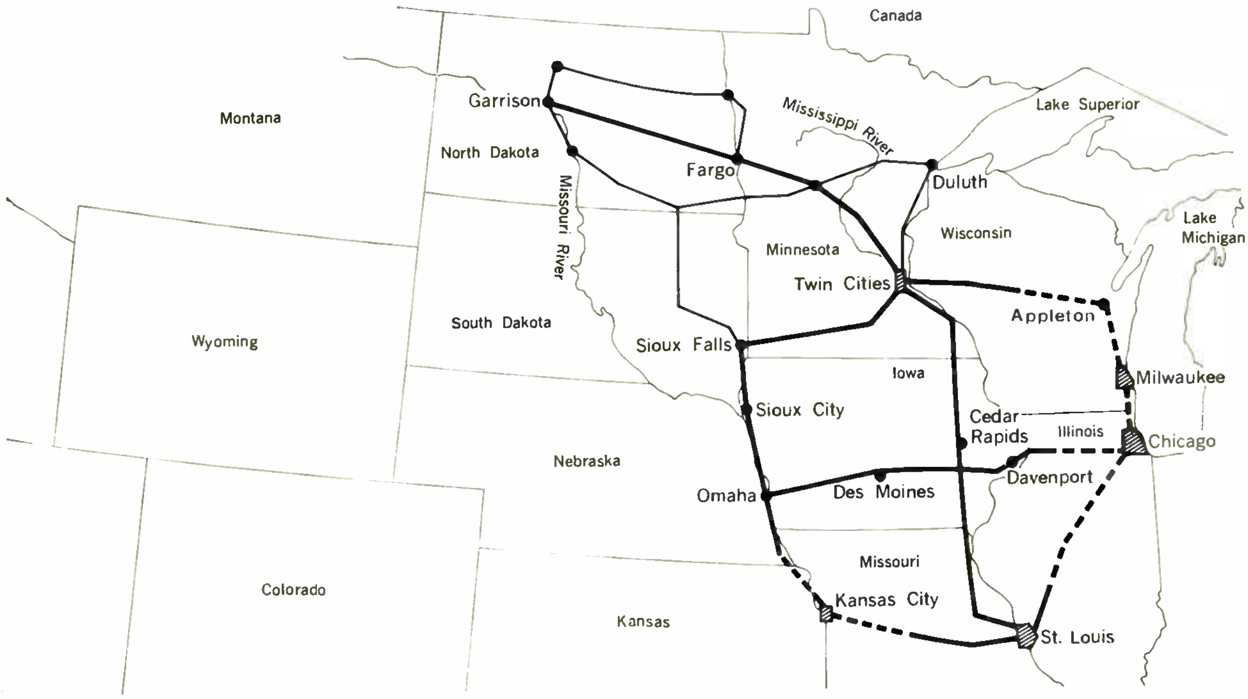



Fig. 5. MAPP plans for proposed 1980 EHV grid covering the north central states.

tion capacity for coordination between the Upper Mississippi Valley Power Pool, the Eastern Wisconsin Utility Group, and the Commonwealth Edison Company. The Commonwealth Edison Company and the eastern Wisconsin utilities are not members of the MAPP organization.

The Twin Cities–Iowa–St. Louis 345-kV line is to be completed in 1967. It will facilitate coordination between the Upper Mississippi Valley, Iowa, and Illinois–Missouri power pools. The line is to be constructed by seven separate members of MAPP. Each of the participants in this EHV project will own, or otherwise provide for the cost of, separate portions of the transmission line and terminal facilities.

The Chicago–St. Louis and St. Louis–Kansas City 345-kV lines are expected to be completed by 1970. The remainder of the lines shown on the map are tentatively planned for construction after 1970, with the exception of some of the 230-kV line sections in northern Minnesota and North Dakota.

A number of alternate transmission patterns with different voltage levels were considered in the process of developing the proposed 345-kV grid plan. One alternate considered was to continue the development of interconnections using existing transmission voltage levels which are in the range of 115 kV to 230 kV. The principal shortcoming of this plan was the excessive cost for providing interconnection capacity between the widely separated major load centers, such as Milwaukee, Minneapolis, Omaha, Des Moines, St. Louis, etc. Also investigated were 500-kV voltage levels; they were found to be less economical than 345-kV because the large interconnection capacities inherent in this voltage could not be utilized effectively in the region for many years. A particular disadvantage of the 500-kV plan was the prohibitive costs of terminal equipment for tap substations in the lighter load density areas intermediate to major load centers.

### 1967 generator coordination

Three to four years of lead time are needed for design and construction of a large base-load generator. Coordinated planning for 1967 generator capacity in the MAPP systems has just been completed and the results offer rather substantial evidence of the degree of large-scale coordination that has been accomplished between MAPP member systems.

Prior to the decision to construct the Twin Cities–St. Louis line in 1967, each of four member systems planned to install a new generating unit in that year. The combination of lower generating reserves, north–south load diversity, and temporary excess capacity in the Illinois–Missouri Power Pool were found to permit postponement of all four generators if the 345-kV line were to be made available. Table II lists the temporary firm power exchanges that have been scheduled for 1967 to accomplish this coordination. The Illinois Power and Central Illinois public service companies are not MAPP members, but they participated, nevertheless, in the 1967 generation coordination through their association with the Union Electric Company in the Illinois–Missouri Power Pool.

### Hydro–thermal integration

Manitoba Hydro joined MAPP in January of this year and potential benefits for extending the grid into Canada are now being explored. The Manitoba generation is predominantly hydro in contrast to the predominantly

## II. 1967 coordination power exchanges

| | Sales, MW | | Purchases, MW |
|---|---|---|---|
| Central Illinois Public Service Co. Springfield, Ill. | 55 | Cooperative Power Assn. Minneapolis, Minn. | 45 |
| Dairyland Power Co-op, La Crosse, Wis. | 28 | Iowa Electric Light & Power Cedar Rapids, Iowa | 55 |
| Illinois Power Co. Decatur, Ill. | 70 | Iowa–Illinois Gas & Electric Davenport, Iowa | 50 |
| Interstate Power Co. Dubuque, Iowa | 75 | Iowa Southern Utilities, Centerville, Iowa | 95 |
| Lake Superior District Ashland, Wis. | 9 | Minnkota Power Co-op, Grand Forks, N. Dak. | 14 |
| Minnesota Power & Light Co., Duluth, Minn. | 67 | Northern States Power Co., Minneapolis, Minn. | 256 |
| Otter Tail Power Co. Fergus Falls, Minn. | 37 | Omaha Public Power District Omaha, Neb. | 35 |
| Rural Power Co-operative Assn. Elk River, Minn. | 34 | | |
| Union Electric Co. St. Louis, Mo. | 175 | | |
| Total | 550 | Total | 550 |

thermal generation in the other MAPP systems. This offers a special incentive for integration because of certain complementary features in these two types of generation. For example, fuel economies can be attained by exchanging thermal energy generated during light load periods for hydro energy generated at peak loads when thermal costs are higher. This can be accomplished by manipulating storage in the hydro reservoirs. The thermal systems can also absorb excess energy or dump power during periods of exceptionally good hydro conditions, and furnish stand-by off-peak energy to firm hydro supplies during drought periods. In addition to these usual hydro–thermal integration benefits, there are large undeveloped hydro resources in Manitoba that may be economically attractive for development on a coordinated basis.

### APPENDIX I.

The calculation of generation reserves shown in Table I was based on 2 per cent forced outage rates for the individual units and a system load loss probability of not more than one day in ten years; that is, the chance that an individual unit will be on forced outage any particular day is 2 out of 100, and the number of spare generators required is such that overlapping outages of units will not be expected to reduce capacity below load levels on the average of more than one day out of ten years.

System load pattern will influence generating reserve requirements. For purposes of simplification, the system load will be assumed continuous at its peak value for 255 days in a year (weekends and holidays excluded) and at negligibly low levels the remainder of the time. Actual utility system loads usually have seasonal valleys. However, scheduled maintenance of generating equipment has the effect of using up the extra spare capacity in these valley periods, so the assumption here of year-around peak loads is a reasonable approximation of practical conditions.

The probability calculations used for the derivation of Table I will now be demonstrated by the solution for reserve requirements in the simple three-generator system. Since the probability that an individual generator will be on forced outage is 0.02, the probability that it will be available for service is 0.98. The generator outages are independent events; hence the daily probability that a particular unit will be on forced outage while the other two remain in service is $0.02 \times 0.98 \times 0.98$. Since there are three units, the probability that any one unit will be out is three times this figure. Similarly, the probability that any two units will be out is three times $0.02 \times 0.02 \times 0.98$, and the probability that all three will be on forced outage is $0.02 \times 0.02 \times 0.02$. Thus the daily probabilities for the various system capacity losses are:

| | |
|---|---|
| One unit | $3 \times 0.02 \times 0.98 \times 0.98 = 0.057\,624$ |
| Two units | $3 \times 0.02 \times 0.02 \times 0.98 = 0.001\,176$ |
| Three units | $1 \times 0.02 \times 0.02 \times 0.02 = 0.000\,008$ |

The daily probability that at least one unit will be on forced outage is the sum of all three of the probabilities listed. Similarly, the daily probability of at least two units being on forced outage is the sum of the probabilities for outage of exactly two units and exactly three units. The average ten-year expectancies for system-capacity loss equal to or exceeding one, two, or three units are the corresponding daily probabilities times the days of exposure or $10 \times 255 = 2550$ days:

| *Capacity Loss* | *Days Per Ten Years* |
|---|---|
| One or more units | 0.058 808 = 2550 = 150.0 |
| Two or more units | 0.001 184 = 2550 = 3.0 |
| Three units | 0.000 008 × 2550 = 0.02 |

It is apparent that a three-unit system would require two spare units for reserve. The resulting service reliability to a system load equal to the capacity of one unit would be measured by an average expectancy of load loss of 0.02 day in 10 years or 1 day out of 50 years.

The number of ways in which various capacity outages could exist increases rapidly with the number of system units and, even in this simple case of identical units, manual computations become impractical for systems with several units. On practical power system reserve problems of even modest size the effect of varying unit sizes and outage rates, load patterns, and many other factors of significance greatly complicate the arithmetic and can only be handled with high-capacity automatic computers.

It is noted that the relationship between spare units and total number of units contains discontinuities. For example, 3 spare units are required in an 18-unit system, and a whole additional spare unit is required for a slight change in unit size in going to a 19-unit system. Data points around these discontinuities were omitted from the graph of Fig. 3 in order to avoid cumbersome multiple inflection points.

### APPENDIX II.

The MAPP members are:

| *Name* | *Location* |
|---|---|
| Black Hills Power & Light Co. | Rapid City, S. Dak. |
| Co-operative Power Assn. | Minneapolis, Minn. |
| Dairyland Power Co-operative | La Crosse, Wis. |
| Eastern Iowa Light & Power Co-operative | Wilton Junction, Iowa |
| Interstate Power Co. | Dubuque, Iowa |
| Iowa Electric Light & Power Co. | Cedar Rapids, Iowa |
| Iowa–Illinois Gas & Electric Co. | Davenport, Iowa |
| Iowa Power & Light Co. | Des Moines, Iowa |
| Iowa Public Service Co. | Sioux City, Iowa |
| Iowa Southern Utilites Co. | Centerville, Iowa |
| Lake Superior District Power Co. | Ashland, Wis. |
| Manitoba Hydro-Electric Board | Winnipeg, Man. |
| Minnesota Power & Light Co. | Duluth, Minn. |
| Minnkota Power Co-operative | Grand Forks, N. Dak. |
| Montana–Dakota Utilities Co. | Bismarck, N. Dak. |
| Northern Minn. Power Assn. | Grand Rapids, Minn. |
| Northern States Power Co. | Minneapolis, Minn. |
| Northwestern Public Service Co. | Huron, S. Dak. |
| Omaha Public Power District | Omaha, Neb. |
| Otter Tail Power Co. | Fergus Falls, Minn. |
| Rural Co-operative Power Assn. | Elk River, Minn. |
| Union Electric Co. | St. Louis, Mo. |

# Matrix functions and applications

## Part II—Functions of a matrix

*In Part II of this series, differential equations with constant matrix coefficients are introduced. These equations may be solved explicitly by finding the eigenvalues and by expressing the functions of a diagonal matrix in terms of its constituent idempotents*

*J. S. Frame* *Michigan State University*

### 2.1 Matrix differential equations

The matrix equations

$$\frac{dX}{dt} = AX \qquad X(0) = C \tag{2.1.1}$$

$$\frac{d^2X}{dt^2} + BX = 0 \qquad X(0) = C \qquad \left.\frac{dX}{dt}\right|_0 = V \tag{2.1.2}$$

in which $A$ and $B$ are $n \times n$ constant matrices and $X$, $C$, $V$ are $n \times 1$ column vectors, provide a simplified notation for the two systems of $n$ linear differential equations in the $n$ unknowns $x_1, \ldots, x_n$:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{ij}x_j \qquad i = 1, 2, \ldots, n \tag{2.1.3}$$

$$\frac{d^2x_i}{dt^2} + \sum_{j=1}^{n} b_{ij}x_j = 0 \quad i = 1, 2, \ldots, n \tag{2.1.4}$$

Solutions in the one-dimensional case would be, respectively,

$$X = e^{At}C \tag{2.1.5}$$

$$X = (\cos B^{1/2}t)C + \left(\frac{\sin B^{1/2}t}{B^{1/2}}\right)V \tag{2.1.6}$$

It is natural to try to define such functions as $e^{At}$, $B^{1/2}$, $\cos B^{1/2}t$, when $A$ and $B$ are square matrices, so that (2.1.5) and (2.1.6) will give the solutions in the general $n$-dimensional case. We shall seek a closed expression for an analytic function $f(A)$ of a square matrix $A$ in terms of corresponding functions of scalars.

### 2.2 Linear transformations and matrix eigenvectors and eigenvalues

An $n \times n$ matrix $A$ maps each $n \times 1$ column vector $X$ into an image vector $AX$. A mapping in which scalar multiples and sums of vectors are mapped into corresponding scalar multiples and sums of their image vectors is called a linear transformation. A change from old coordinates $X$ to new coordinates $Y$ for the same vector may be specified by any nonsingular $n \times n$ matrix $S$ by setting

$$X = SY \qquad Y = S^{-1}X \tag{2.2.1}$$

A corresponding change must then be made in the matrix $A$ that describes the linear transformation $X \to AX$. Since the transformation is linear, it maps $S^{-1}X$ into $S^{-1}AX$. Expressed in the new coordinate system, this is written

$$Y \to S^{-1}ASY \quad \text{whenever } X \to AX \text{ and } X = SY \tag{2.2.2}$$

The new matrix for the linear transformation is $S^{-1}AS$. Among the simplest to describe is a linear transformation that in some suitable coordinate system is represented by a diagonal matrix $\Lambda = S^{-1}AS$ with diagonal entries denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$. It simply multiplies the new $j$th coordinate vector $U_{.j}$ (represented in the old coordinates as the $j$th column $S_{.j}$ of $S$) by the scalar $\lambda_j$, for all $j$, and maps the linear combinations of these so-called eigenvectors into corresponding linear combinations of their images.

Corresponding facts hold for the linear transformation of the row vectors $Z^T$, which are defined by the mapping $Z^T \rightarrow Z^TA$.

**Definition 2.2.1** A nonzero column vector $X$ (or row vector $Z^T$), that is mapped by an $n \times n$ matrix $A$ into a scalar multiple $\lambda X$ (or $\lambda Z^T$) of itself, is called a column (or row) *eigenvector* of $A$. The corresponding scalar $\lambda$ is called an *eigenvalue* of $A$. Conditions that the nonzero column vector $X$ be an eigenvector of $A$ may be written as follows:

$$AX = X\lambda \quad \text{or} \quad (\lambda U - A)X = 0 \tag{2.2.3}$$

Corresponding conditions for a row eigenvector $Z^T \neq 0$ are

$$Z^TA = \lambda Z^T \quad \text{or} \quad Z^T(\lambda U - A) = 0 \tag{2.2.4}$$

Each of the two systems of $n$ equations in $n$ unknowns has the solution vector 0. The vanishing of the determinant of the coefficient matrix $\lambda U - A$, called the *characteristic matrix* of $A$, is a necessary and sufficient condition for the existence of nonzero solutions that determine a column eigenvector $X$ by (2.2.3) or a row eigenvector $Z^T$ by (2.2.4). This determinant $D(\lambda)$ is called the *characteristic polynomial* of $A$. It is the following monic polynomial of degree $n$:

$$D(\lambda) = |\lambda U - A| \tag{2.2.5}$$

$$= \lambda^n + d_1\lambda^{n-1} + d_2\lambda^{n-2} + \ldots + d_{n-1}\lambda + d_n$$

$$= \prod_{j=1}^{s} (\lambda - \lambda_j)^{n_j} \tag{2.2.6}$$

The *characteristic equation* $D(\lambda) = 0$ has characteristic roots $\lambda_j$, with characteristic multiplicities $n_j$, that are the $n$ eigenvalues of $A$. The set $\sigma(A)$ of these $n$ eigenvalues, including repetitions, is called the *spectrum* of $A$. Their product is the determinant of $A$ and their sum is its *trace*

$$|A| = \prod \lambda_j^{n_j} = (-1)^n d_n \tag{2.2.7}$$

$$\text{tr } A = \sum_{i=1}^{n} a_{ii} = \sum_{j=1}^{s} n_j\lambda_j = -d_1 \tag{2.2.8}$$

To each of the $s$ distinct eigenvalues $\lambda_j$ of $A$ there corresponds at least one eigenvector $X \neq 0$ that satisfies (2.2.3), or $Z^T \neq 0$ that satisfies (2.2.4). When all $n$ eigenvalues of $A$ are distinct, the corresponding $n$ eigenvectors $X$ constitute the columns $S_{.j}$ of a matrix $S$, and the eigenvalues are diagonal entries of a matrix $\Lambda$ such that $AS = S\Lambda$. If $S$ is nonsingular this implies $\Lambda = S^{-1}AS$.

*Example 1* Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$$

*Solution:*

$$D(\lambda) = |\lambda U - A| = \begin{vmatrix} \lambda - 1 & -2 \\ -4 & \lambda - 3 \end{vmatrix}$$

$$= (\lambda - 1)(\lambda - 3) - 8 = \lambda^2 - 4\lambda - 5$$

$$= (\lambda + 1)(\lambda - 5) \tag{2.2.9}$$

Thus the eigenvalues of $A$ are $\lambda_1 = -1, \lambda_2 = 5$. Their sum is tr $A = 1 + 3 = 4$ and their product is $|A| = 3 - 8 = -5$. Solution vectors $X$ for $(\lambda_j U - A)X = 0$ are the eigenvectors $S_{.j}$.

$$\begin{bmatrix} (-1-1) & -2 \\ -4 & (-1-3) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$X = S_{.1} = \begin{bmatrix} a \\ -a \end{bmatrix} \text{ for any } a \neq 0 \quad (2.2.10)$$

$$\begin{bmatrix} (5-1) & -2 \\ -4 & (5-3) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$X = S_{.2} = \begin{bmatrix} b \\ 2b \end{bmatrix} \text{ for any } b \neq 0$$

Taking $a = b = 1$, we form the matrix $S = (S_{.1}, S_{.2})$ and the diagonal matrix $\Lambda$ with diagonal entries $\lambda_1, \lambda_2$.

$$S = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix} \quad (2.2.11)$$

$$AS = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 5 \\ 1 & 10 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix} = S\Lambda \quad (2.2.12)$$

**Definition 2.2.2** For any nonsingular matrix $S$, the matrix $S^{-1}AS$ is said to be *similar* to $A$.

**Definition 2.2.3** A matrix is called *diagonable* if it is similar to a diagonal matrix.

Clearly, similarity is a reflexive, symmetric, and transitive relation that divides all $n \times n$ matrices into mutually exclusive classes of similar matrices, each representing the same linear transformation in an appropriate coordinate system. For each class of similar matrices there are important common properties. For each class there are certain "spectral" matrices that are considered to be the simplest for describing the corresponding linear transformation. For a diagonable matrix $A$ the spectral matrix is a diagonal matrix with the same eigenvalues.

For example, the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}$$

in (2.2.12) are similar. $A$ is diagonable and $\Lambda$ is its spectral matrix.

**Theorem 2.2.1** Two similar square matrices have the same characteristic polynomial.

*Proof:* Since the determinant of the product of two square matrices equals the product of their determinants, we have

$$\begin{aligned} |\lambda U - S^{-1}AS| &= |S^{-1}|\,|\lambda S - AS| \\ &= |\lambda S - AS|\,|S^{-1}| \\ &= |\lambda U - A| = D(\lambda) \qquad (2.2.13) \end{aligned}$$

It follows that any two similar matrices $A$ and $S^{-1}AS$ have the same eigenvalues $\lambda_j$ with the same characteristic multiplicities $n_j$ and hence the same trace and determinant. They also have the same rank. For example, the following $5 \times 5$ matrix $A$ is similar to the upper triangular matrix $S^{-1}AS = T$.

$$\underset{S^{-1}}{\begin{bmatrix} 1&0&0&0&0 \\ 1&1&0&0&0 \\ 1&0&1&0&0 \\ 1&0&0&1&0 \\ 1&0&0&0&1 \end{bmatrix}} \underset{A}{\begin{bmatrix} 3&2&3&0&-3 \\ 1&-1&-2&1&4 \\ 2&-2&-2&2&5 \\ -2&-2&-3&-2&6 \\ -5&-2&-3&0&1 \end{bmatrix}} \underset{S}{\begin{bmatrix} 1&0&0&0&0 \\ -1&1&0&0&0 \\ -1&0&1&0&0 \\ -1&0&0&1&0 \\ -1&0&0&0&1 \end{bmatrix}}$$

$$= \begin{bmatrix} 1&2&3&0&-3 \\ 0&1&1&1&1 \\ 0&0&1&2&2 \\ 0&0&0&-2&3 \\ 0&0&0&0&-2 \end{bmatrix} \quad (2.2.14)$$

$$= T$$

Note that $A$ and $T$ both have trace $-1$ and determinant 4. However, if multiple roots are present, two $n \times n$ matrices with the same spectrum are not necessarily similar.

### 2.3 Similarity to triangular and spectral matrices

Eigenvalues of a triangular matrix can be determined by inspection.

**Theorem 2.3.1** The eigenvalues of an $n \times n$ triangular matrix are its $n$ diagonal entries.

*Proof:* If $T$ is any $n \times n$ triangular matrix (upper triangular, lower triangular, or diagonal) then the matrix $\lambda U - T$ is also triangular, so its determinant $D(\lambda)$ is the product

$$\prod(\lambda - t_{jj})$$

of its diagonal entries $\lambda - t_{jj}$. Hence $t_{jj}$ are the eigenvalues $\lambda_j$ of $T$.

**Definition 2.3.1** A triangular matrix is *simply ordered* if equal eigenvalues appear next to each other along its diagonal. The matrix $T$ in (2.2.14) is a simply ordered upper triangular matrix with a triple eigenvalue 1 and a double eigenvalue $-2$. A *Jacobi matrix* is a simply ordered upper triangular matrix whose $ij$ entry is zero whenever the $i$th and $j$th diagonal entries are distinct ($\lambda_i \neq \lambda_j$). A Jacobi matrix is a quasi-diagonal direct sum of $s$ blocks of respective dimensions $n_j$, each block being an upper triangular matrix with equal eigenvalues.

The following Jacobi matrix $J_1$ and Jordan matrix $J_2$ are each similar to the upper triangular matrix $T$ in (2.2.14):

$$J_1 = \begin{bmatrix} 1&2&3&0&0 \\ 0&1&1&0&0 \\ 0&0&1&0&0 \\ 0&0&0&-2&3 \\ 0&0&0&0&-2 \end{bmatrix} \qquad J_2 = \begin{bmatrix} 1&1&0&0&0 \\ 0&1&1&0&0 \\ 0&0&1&0&0 \\ 0&0&0&-2&1 \\ 0&0&0&0&-2 \end{bmatrix}$$

**Definition 2.3.2** A *Jordan matrix* $\Lambda$ is a Jacobi matrix whose $ij$ entries are 0 unless $j = i$ or $i + 1$. Its $(i, i + 1)$ entry may be 0 or 1 if $\lambda_i = \lambda_{i+1}$, but is 0 otherwise. A Jordan matrix is the direct sum of one or more simple Jordan matrices $\Lambda_j$ each having equal eigenvalues $\lambda_j$ on its diagonal, l's immediately to the right of the diagonal, and 0's elsewhere.

$$\underset{\text{Jordan}}{\Lambda = \begin{bmatrix} \Lambda_1 & 0 & \ldots & 0 \\ 0 & \Lambda_2 & \ldots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \ldots & \Lambda_k \end{bmatrix}} \qquad \underset{\text{Simple Jordan}}{\Lambda_j = \begin{bmatrix} \lambda_j & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_j & 1 & \ldots & 0 & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 0 & 0 & 0 & \ldots & \lambda_j & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_j \end{bmatrix}} \quad (2.3.1)$$

A Jordan matrix $\Lambda$ is a diagonal matrix if and only if all its simple Jordan submatrices $\Lambda_j$ are one-dimensional. A diagonable matrix may have repeated eigenvalues.

We shall prove that every $n \times n$ matrix is similar to an upper triangular matrix, indicate that it is similar to a Jacobi matrix, and merely state that it is similar to a Jordan matrix. The latter will be called its *spectral* matrix.

**Definition 2.3.3** A matrix $S$ is called *unitary* if

$$SS^* = S^*S = U \qquad \text{(with } S \text{ unitary)} \qquad (2.3.2)$$

where $S^*$ denotes the conjugate transpose of $S$.

The product of two $n \times n$ unitary matrices is easily shown to be unitary. If $S$ is unitary, the matrix $A$ is called *unitarily similar* to $S^*AS$ or to $SAS^*$. It is easier to perform a unitary change of coordinates, for which $S^{-1} = S^*$, than one in which the computation of $S^{-1}$ is more complicated.

**Theorem 2.3.2** Any $n \times n$ matrix $A$ is unitarily similar to an upper triangular matrix $T$ whose diagonal entries are the eigenvalues $\lambda_j$ of $A$ arranged in any prescribed order. That is,

$$A = STS^* \qquad (2.3.3)$$

where $SS^* = U$, $T$ is upper triangular, and $t_{jj} = \lambda_j$.

*Proof:* Since any $1 \times 1$ matrix is itself upper triangular, the theorem is true for $n = 1$ with $S = U_1 = 1$. To prove the theorem by mathematical induction, we assume its validity for square matrices of order less than $n$, and prove it for an arbitrary $n \times n$ matrix $A$ whose eigenvalues have been numbered in the prescribed order $\lambda_1, \lambda_2, \ldots, \lambda_n$, (allowing repetitions). An eigenvector of $A$ for the eigenvalue $\lambda_1$ is first reduced to a unit vector by dividing by its length; then its first component is changed to a real number $x \geq 0$, by dividing the vector (if necessary) by a complex number of absolute value 1 that preserves the length. From the partitioned unit eigenvector

$$V_1 = \begin{bmatrix} x \\ X \end{bmatrix} \begin{matrix} 1 \text{ row} \\ n-1 \text{ rows} \end{matrix}$$

we construct the $n \times n$ matrix

$$V = (V_1, V_2) = \left[\begin{array}{c|c} x & X^* \\ \hline X & -U + XX^*/(1+x) \end{array}\right] = V^* \qquad (2.3.4)$$

$$V_1^*V_1 = x^2 + X^*X = 1$$

Direct multiplication shows that $V$ is not only hermitian ($V = V^*$), but also unitary:

$$VV^* = V^*V = U_n \qquad V_2^*V_1 = 0 \qquad (2.3.5)$$

Since $AV_1 = V_1\lambda_1$, the transform of $A$ by $V$ is

$$V^*AV = \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} [V_1\lambda_1, AV_2] = \begin{bmatrix} \lambda_1 & V_1^*AV_2 \\ 0 & V_2^*AV_2 \end{bmatrix} \qquad (2.3.6)$$

Since $A$ and $V^*AV$ have equal eigenvalues, those of $V_2^*AV_2$ are $\lambda_2, \ldots \lambda_n$. By induction hypothesis the $(n-1) \times (n-1)$ matrix $V_2^*AV_2$ can be transformed into an upper triangular matrix $T_2$ with diagonal entries $\lambda_2, \ldots, \lambda_n$, by means of a unitary matrix $W$.

$$W^*(V_2^*AV_2)W = T_2 \text{ (upper triangular)} \qquad W^*W = U_{n-1} \qquad (2.3.7)$$

Construct the unitary matrix $S$ such that

$$S = V\begin{bmatrix} 1 & 0 \\ 0 & W \end{bmatrix} = (V_1, V_2W) \qquad SS^* = U \qquad (2.3.8)$$

$$S^*AS = \begin{bmatrix} 1 & 0 \\ 0 & W^* \end{bmatrix} V^*AV \begin{bmatrix} 1 & 0 \\ 0 & W \end{bmatrix} = \begin{bmatrix} \lambda_1 & V_1^*AV_2W \\ 0 & T_2 \end{bmatrix} = T \qquad (2.3.9)$$

Then $T$ is the required upper triangular matrix, unitarily similar to $A$.

**Definition 2.3.4** A matrix $A$ is called *normal* if $AA^* = A^*A$. The class of normal matrices includes hermitian and real symmetric matrices, unitary and real orthogonal matrices, real skew matrices, and several other important types.

The following normal matrices are marked $V$ if unitary ($VV^* = U$), $W$ if real orthogonal ($WW^T = U$), $H$ if hermitian ($H = H^*$), $S$ if real symmetric ($S = S^T$), $K$ if real skew ($K = -K^T$), and $E$ if idempotent ($E^2 = E$):

$$\underset{V,\ W,\ H,\ S}{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \\ 2/3 & 1/3 & -2/3 \end{bmatrix}} \qquad \underset{V,\ W,\ K}{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}$$

$$\underset{V}{\begin{bmatrix} 1/2(1+i) & 1/2(1+i) \\ -1/2(1+i) & 1/2(1-i) \end{bmatrix}} \qquad \underset{H,\ E}{\begin{bmatrix} 1/2 & 1/2\,i \\ -1/2\,i & 1/2 \end{bmatrix}}$$

**Theorem 2.3.3** Every normal matrix $A$ is unitarily similar to a diagonal matrix.

*Proof:* If $A$ is normal, so is the upper triangular matrix $T = S^*AS$ of (2.3.9), since

$$TT^* = S^*ASS^*A^*S = S^*AA^*S = S^*A^*AS = T^*T \qquad (2.3.10)$$

The squared lengths of the $j$th row and $j$th column of $T$ must be equal, since these are the $j$th diagonal entries of $TT^*$ and $T^*T$, respectively. Applied successively for $j = 1, 2, 3, \ldots, n-1$, this condition requires that all entries of $T$ to the right of the diagonal must vanish. $T$ is diagonal and the theorem is proved.

The matrix equation

$$\begin{bmatrix} \lambda_j & c_{jk} \\ 0 & \lambda_k \end{bmatrix}\begin{bmatrix} 1 & c_{jk}/(\lambda_k - \lambda_j) \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & c_{jk}/(\lambda_k - \lambda_j) \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \lambda_j & 0 \\ 0 & \lambda_k \end{bmatrix} \qquad (2.3.11)$$

suggests a proof of the fact that every simply ordered upper triangular matrix can be transformed into a similar Jacobi matrix by reducing to zero each $jk$ entry for which $\lambda_j \neq \lambda_k$.

Stated here without proof is the important fact (to be discussed in Part IV) that every Jacobi matrix, and hence every square matrix, is similar to a Jordan matrix. The eigenvalues $\lambda_j$, and the dimensions of the one or more simple Jordan blocks for each $\lambda_j$, constitute a complete set of invariants for a class of similar matrices, that are completely determined by the ranks of the matrices $(\lambda_j U - A)$ and their powers.

**Definition 2.3.5** A Jordan matrix $\Lambda = S^{-1}AS$ that is similar to a given matrix $A$ is called a *spectral matrix* for $A$, and the transforming matrix $S$ with modal columns $S_{.j}$ is called a *right modal matrix* for $A$. Its inverse, $R = S^{-1}$, with modal rows $R_{i.}$ is called a *left modal matrix* for $A$.

Further discussion of nondiagonable matrices will be postponed to Part IV. The rest of the present discussion will be devoted to diagonable matrices—that is, those whose spectral matrices are diagonal.

## 2.4 Functions of diagonable matrices

The equations

$$AS = S\Lambda \quad \text{or} \quad A = S\Lambda S^{-1} \tag{2.4.1}$$

satisfied by any two similar matrices $A$ and $\Lambda$ imply the relations

$$A^2S = AS\Lambda = S\Lambda^2 \quad \text{or} \quad A^kS = S\Lambda^k \qquad k = 1, 2, 3, \ldots \tag{2.4.2}$$

Hence if $f(\lambda)$ and $f(A)$ denote corresponding scalar and matrix polynomials

$$f(\lambda) = f_0\lambda^m + f_1\lambda^{m-1} + f_2\lambda^{m-2} + \ldots + f_{m-1}\lambda + f_m \tag{2.4.3}$$

$$f(A) = f_0A^m + f_1A^{m-1} + f_2A^{m-2} + \ldots + f_{m-1}A + f_mU \tag{2.4.4}$$

then by addition using (2.4.2) we find

$$f(A)S = Sf(\Lambda) \qquad f(A) = Sf(\Lambda)S^{-1} \tag{2.4.5}$$

Under this correspondence, sums and products of functions of $\Lambda$ are mapped into corresponding sums and products of functions of $A$. If for the spectral matrix $\Lambda$ such functions as $e^{\Lambda t}$, $\cos \Lambda^{1/2}t$, ... can be defined uniquely, then (2.4.5) defines the corresponding functions of $A$.

Let $\Lambda = S^{-1}AS$ be a diagonal matrix similar to a given diagonable matrix $A$. Then for any polynomial $f(\lambda)$ we have

$$f(\Lambda) = \begin{bmatrix} f(\lambda_1) & 0 & \ldots & 0 \\ 0 & f(\lambda_2) & \ldots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \ldots & f(\lambda_n) \end{bmatrix} = \sum_{i=1}^{n} f(\lambda_i)\epsilon_{ii} \tag{2.4.6}$$

where $\epsilon_{ii}$ are the matrix units defined in Part I.

An important special case is this:

**Theorem 2.4.1 (Hamilton-Cayley theorem)** Every square matrix satisfies its characteristic equation.

*Proof for diagonable matrices:* Taking $D(\lambda)$ for $f(\lambda)$ in (2.4.6), it follows that $D(\lambda_j) = 0$ for each $j$, so $D(\Lambda)$ is the zero matrix. By (2.4.5) it follows that $D(A)$ is also the zero matrix.

*Example 2* Taking $A$ as in Example 1, we have

$$A^2 - 4A - 5U = \begin{bmatrix} 9 & 8 \\ 16 & 17 \end{bmatrix} - 4\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} - 5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{2.4.7}$$

**Definition 2.4.1** A matrix $E$ is called *idempotent* if $E = E^2$.

Examples of $2 \times 2$ idempotent matrices are

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix} \quad \begin{bmatrix} 1/3 & -1/3 \\ -2/3 & 2/3 \end{bmatrix}$$

The first two of these are diagonal idempotents. Both the first pair and second pair have sum $U$ and product 0.

Equations (2.4.6) and (2.4.5) may be used to define $f(\Lambda)$ and $f(A)$, if $f(\lambda)$ is a function defined by a power series that converges at all the eigenvalues of $A$ (or of $\Lambda$).

If $\Lambda$ has the $s$ distinct eigenvalues $\lambda_j$ with respective multiplicities $n_j$, we shall denote by $I_j$ the sum of the $n_j$ matrix units $\epsilon_{ii}$ for all $i$ such that $\lambda_i = \lambda_j$. This matrix $I_j$ is a diagonal idempotent matrix with $n_j$ 1's as diagonal entries, and having rank and trace $n_j$. Its transform by $S$ is a so-called *constituent idempotent* matrix $A_j$ for $A$, whose rank and trace are also $n_j$.

$$A_j = SI_jS^{-1} = S\sum_i \epsilon_{ii}R = \sum_i S_{.i}R_{i.} \quad \text{for } \lambda_i = \lambda_j \tag{2.4.8}$$

The idempotents $I_j$ and $A_j$ satisfy the relations

$$\sum I_j = \sum A_j = U \tag{2.4.9}$$

$$I_jI_k = I_j\delta_{jk} \qquad A_jA_k = A_j\delta_{jk} \tag{2.4.10}$$

**Theorem 2.4.2** An analytic function of a diagonable matrix $A$, with distinct eigenvalues $\lambda_j$, where $j = 1, 2, \ldots s$, and corresponding constituent idempotent matrices $A_j$ defined by (2.4.8), is expressed by the equation

$$f(A) = \sum_{j=1}^{s} f(\lambda_j)A_j \tag{2.4.11}$$

*Proof:* On transforming $f(\Lambda)$ in (2.4.6) by $S$ and applying (2.4.8) we obtain the required result:

$$f(A) = Sf(\Lambda)R = \sum_{i=1}^{i} f(\lambda_i)S\epsilon_{ii}R \tag{2.4.12}$$

Thus any analytic function of a diagonable matrix $A$ that has $s$ distinct eigenvalues $\lambda_j$ (of multiplicities $n_j$) is a linear combination of the $s$ constituent matrices $A_j$. The power formula

$$A^k = \sum_{j=1}^{s} \lambda_j^k A_j \tag{2.4.13}$$

can also be used for $k = -1$ to compute the inverse $A^{-1}$ of a matrix $A$ without 0 eigenvalues, since

$$\sum_{i=1}^{s} \lambda_j A_j \sum_{h=1}^{s} \lambda_h^{-1} A_h = \sum_{j,h} \frac{\lambda_j}{\lambda_h} A_jA_j = \sum_j A_j = U \tag{2.4.14}$$

For nonintegral values of $k$, the function $A^k$ may be multiple valued.

To use (2.4.11), the eigenvalues $\lambda_j$ and corresponding constituent idempotents $A_j$ must first be computed. The former are roots of the characteristic equation $D(\lambda) = 0$. The latter can be computed from the modal matrix $S$ and its inverse $R$ if these are known. For normal matrices $A$, such that $AA^* = A^*A$, including real symmetric or real orthogonal matrices, the modal matrix $S$ may be chosen unitary, so that $R = S^{-1} = S^*$ and $R_{i.} = S_{.i}^*$. Here the eigenvectors $S_{.j}$ suffice to determine the constituent idempotents from (2.4.8).

However, the constituent idempotents $A_j$ can be determined directly as polynomials in $A$ by (2.4.18), without knowing the eigenvectors. The columns of $A_j$ can then be used to compute the eigenvectors. To show this, we choose for $f(\lambda)$ in (2.4.6) the interpolatory polynomial $q_k(\lambda)$ of degree $s - 1$ defined by

$$q_k(\lambda) = \prod_{\substack{j=1 \\ j \neq k}}^{s} \frac{\lambda - \lambda_j}{\lambda_k - \lambda_j} \tag{2.4.15}$$

We note that $q_k(\lambda)$ vanishes at each root $\lambda_j$ except $\lambda_k$.

$$q_k(\lambda_j) = 0 \qquad j \neq k \qquad q_k(\lambda_k) = 1 \tag{2.4.16}$$

Hence

$$q_k(\Lambda) = I_k \tag{2.4.17}$$

$$A_k = SI_kR = q_k(A) = \prod_{\substack{j=1 \\ j \neq k}}^{s} \frac{A - \lambda_j U}{\lambda_k - \lambda_j} \tag{2.4.18}$$

Here then is an explicit formula for the constituent idempotents of a diagonable matrix $A$. In Example 1 (Section 2.2), where $\lambda_1 = -1$ and $\lambda_2 = 5$, we have

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \qquad A_1 = \frac{A - 5U}{-1 - 5} = \frac{5U - A}{6}$$

$$= \begin{bmatrix} ^2/_3 & -^1/_3 \\ -^2/_3 & ^1/_3 \end{bmatrix} \tag{2.4.19}$$

$$A_2 = \frac{A + U}{5 + 1} = \frac{U + A}{6} = \begin{bmatrix} ^1/_3 & ^1/_3 \\ ^2/_3 & ^2/_3 \end{bmatrix}$$

We note that $A_1A_2 = 0$, $A_1 + A_2 = U$, $A^k = \lambda_1{}^kA_1 + \lambda_2{}^kA_2$, tr $A_1$ = tr $A_2 = 1$, and that the columns of $A_1$ and $A_2$ are eigenvectors of $A$ for the eigenvalues $\lambda_1$, $\lambda_2$.

**Theorem 2.4.3** Eigenvectors of any square matrix that correspond to distinct eigenvalues are linearly independent.

*Proof:* Let $\lambda_1, \lambda_2, \ldots, \lambda_s$ be distinct eigenvalues of an $n \times n$ matrix $A$, and let the columns $S_{.1}, S_{.2}, \ldots, S_{.s}$ of an $n \times s$ matrix $S$ be corresponding eigenvectors. Then

$$AS_{.j} = S_{.j}\lambda_j \qquad AS = S\Lambda \tag{2.4.20}$$

where $\Lambda$ is an $s \times s$ diagonal matrix with the distinct numbers $\lambda_1 \ldots \lambda_s$ on the diagonal. Linear independence of the $s$ vectors $S_{.j}$ means that if $s$ scalars $c_j$ form an $s \times 1$ column vector $C$ such that $SC = 0$, then $C = 0$. Assuming $SC = 0$, we multiply on the left by $q_j(A)$; see (2.4.15). Then

$$0 = q_j(A)SC = Sq_j(\Lambda)C = S\epsilon_{jj}C = S_{.j}c_j \tag{2.4.21}$$

Since $S_{.j} \neq 0$, we have $c_j = 0$ for all $j$. Hence $SC = 0$ implies $C = 0$.

**Theorem 2.4.4** If $A$ is any normal matrix, then there exists a polynomial $f(\lambda)$ such that

$$A^* = f(A) \qquad \lambda_j^* = f(\lambda_j) \qquad \text{for each } j \tag{2.4.22}$$

*Proof:* Since $I_j = I_j^*$, we have

$$A = \sum_{j=1}^{s} \lambda_j A_j = \sum_{j=1}^{s} \lambda_j SI_jS^* \tag{2.4.23}$$

$$A^* = \sum_{j=1}^{s} \lambda_j^* SI_j^*S^* = \sum_{j=1}^{s} \lambda_j^* A_j = \sum_{j=1}^{s} \lambda_j^* q_j(A) \tag{2.4.24}$$

The required polynomial is

$$f(\lambda) = \sum \lambda_j^* q_j(\lambda)$$

which proves (2.4.22).

**Theorem 2.4.5** Every matrix $B$ that commutes with an $n \times n$ matrix $A$ having $n$ distinct eigenvalues is equal to a polynomial in $A$.

*Proof:* If $B$ commutes with $A$ it commutes with the idempotents $A_j = S_{.j}R_{j.}$ that are polynomials in $A$. Let $\mu_j = R_{j.}BS_{.j}$. Then

$$B = UB = \sum_{j=1}^{n} A_j{}^2B = \sum_{j=1}^{n} A_jBA_j = \sum_{j=1}^{s} S_{.j}\mu_jR_{j.} \tag{2.4.25}$$

$$= \sum_{j=1}^{n} \mu_jA_j = \sum_{j=1}^{n} \mu_jq_j(A) = f(A)$$

where $f(\lambda) = \sum \mu_jq_j(\lambda)$

A word of caution is necessary in saying that all functions of a diagonable matrix are expressible by Eq. (2.4.11). This equation evaluates single-valued analytic functions, but may not give all values of multiple-valued functions, such as $f(\lambda) = \lambda^{1/2}$. For example, the following equation shows that the $2 \times 2$ unit matrix $U$ has infinitely many square roots that are not expressible as polynomials in $U$.

$$\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.4.26}$$

### 2.5 Computation of functions of matrices

*Example 3* Find the square roots of

$$A = \begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix} \tag{2.5.1}$$

*Solution:*

$$D(\lambda) = |\lambda U - A| = \begin{vmatrix} \lambda - 8 & -4 \\ -4 & \lambda - 8 \end{vmatrix}$$

$$= \lambda^2 - 16\lambda + 48 = (\lambda - 4)(\lambda - 12)$$

The eigenvalues of $A$ are $\lambda_1 = 4$, $\lambda_2 = 12$. The constituent idempotents of $A$ are

$$A_1 = \frac{A - \lambda_2 U}{\lambda_1 - \lambda_2} = \frac{12U - A}{8} = \begin{bmatrix} ^1/_2 & -^1/_2 \\ -^1/_2 & ^1/_2 \end{bmatrix}$$

$$\tag{2.5.2}$$

$$A_2 = \frac{A - \lambda_1 U}{\lambda_2 - \lambda_1} = \frac{A - 4U}{8} = \begin{bmatrix} ^1/_2 & ^1/_2 \\ ^1/_2 & ^1/_2 \end{bmatrix}$$

Check:

$$A_1{}^2 = A_1,\ A_2{}^2 = A_2,\ A_1A_2 = 0,\ A_1 + A_2 = U$$

$$A = \lambda_1A_1 + \lambda_2A_2 = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} + \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$$

The solution to the problem includes four square roots of $A$.

$$A^{1/2} = \pm\lambda_1{}^{1/2}A_1 \pm \lambda_2{}^{1/2}A_2$$

$$= \pm\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \pm 3^{1/2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{2.5.3}$$

Columns of $A_1$ and $A_2$ are eigenvectors of the symmetric matrix $A$. A unitary modal matrix $S$ and its inverse $S^*$ are

$$S = \sqrt{2}\begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix} \quad \text{and}$$

$$S^{-1} = S^* = \sqrt{2}\begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \quad (2.5.4)$$

where we have multiplied the columns in (2.5.2) by $2^{1/2}$ to make them of unit length.

*Example 4* Evaluate $e^{J\theta}$ if

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (2.5.5)$$

where $J^2 = -U$.

*Solution:* $D(\lambda) = \begin{vmatrix} \lambda & -1 \\ 1 & \lambda \end{vmatrix} = \lambda^2 + 1$

where $\lambda_1 = j$, $\lambda_2 = -j$, $j^2 = -1$.

Constituent idempotents of $J$ are

$$J_1 = \frac{J - \lambda_2 U}{\lambda_1 - \lambda_2} = \frac{J + jU}{2j} = \frac{U - jJ}{2} = \frac{1}{2}\begin{bmatrix} 1 & -j \\ j & 1 \end{bmatrix} \quad (2.5.6)$$

$$J_2 = \frac{J - \lambda_1 U}{\lambda_2 - \lambda_1} = \frac{J - jU}{-2j} = \frac{U + jJ}{2} = \frac{1}{2}\begin{bmatrix} 1 & j \\ -j & 1 \end{bmatrix}$$

Check: $J_1^2 = J_1$, $J_2^2 = J_2$, $J_1J_2 = 0$, $J_1 + J_2 = U$.

The solution is

$$e^{J\theta} = e^{\lambda_1\theta} J_1 + e^{\lambda_2\theta} J_2 = \frac{1}{2} e^{j\theta}\begin{bmatrix} 1 & -j \\ j & 1 \end{bmatrix} + \frac{1}{2} e^{-j\theta}\begin{bmatrix} 1 & j \\ -j & 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad (2.5.7)$$

Thus $e^{J\theta}$ is a matrix that represents a rotation of the plane through the angle $\theta$.

*Example 5* These principles will now be applied to a dynamical problem. Let three equal masses be attached to a spring of negligible weight at equally spaced points dividing the spring into four equal parts when it is stretched at rest between two hooks four units apart. Let the three masses oscillate along the line between two hooks, so that at time $t$ the displacement of the $i$th mass from its equilibrium point is $x_i$. Assume that $x_i$ has the value $c_i$ and derivative $v_i$ at $t = 0$. Finally, let $mk^2$ denote the spring constant.

Then the force acting on the $i$th mass at time $t$ is

$$F_i = mk^2[(x_{i+1} - x_i) - (x_i - x_{i-1})] \quad i = 1, 2, 3 \quad (2.5.8)$$

where $x_0 = x_4 = 0$

Writing $X$ as a column vector with components $x_1$, $x_2$, $x_3$, the equations of motion take the form

$$m\frac{d^2X}{dt^2} = mk^2\begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix} X$$

or (2.5.9)

$$\frac{d^2X}{dt^2} + k^2(2U - A)X = 0$$

where the matrix $A$ and its characteristic polynomial are as follows:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (2.5.10)$$

$$D(\lambda) = \begin{vmatrix} \lambda & -1 & 0 \\ -1 & \lambda & -1 \\ 0 & -1 & \lambda \end{vmatrix} = \lambda^3 - 2\lambda$$

The eigenvalues are $\lambda_1 = 2^{1/2}$, $\lambda_2 = 0$, and $\lambda_3 = -2^{1/2}$, and the constituent idempotent matrices are

$$A_1 = \frac{(A - \lambda_2 U)(A - \lambda_3 U)}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)} = \frac{A^2 + 2^{1/2}A}{4} = \frac{1}{4}\begin{bmatrix} 1 & \sqrt{2} & 1 \\ \sqrt{2} & 2 & \sqrt{2} \\ 1 & \sqrt{2} & 1 \end{bmatrix}$$

$$A_2 = \frac{(A - \lambda_1 U)(A - \lambda_3 U)}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)} = \frac{A^2 - 2U}{-2} = \frac{1}{4}\begin{bmatrix} 2 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 2 \end{bmatrix} \quad (2.5.11)$$

$$A_3 = \frac{(A - \lambda_1 U)(A - \lambda_2 U)}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)} = \frac{A^2 - 2^{1/2}A}{4} = \frac{1}{4}\begin{bmatrix} 1 & -\sqrt{2} & 1 \\ -\sqrt{2} & 2 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 \end{bmatrix}$$

Check that $A_jA_k = A_j\delta_{jk}$, $\Sigma A_j = U$.

The solution of (2.5.9) can be written in a form similar to (2.1.6), namely

$$X = (\cos k\sqrt{2U - A}\,t)C + \left(\frac{\sin k\sqrt{2U - A}\,t}{k\sqrt{2U - A}}\right)V \quad (2.5.12)$$

With the aid of the function expansion formula (2.4.11), this becomes

$$X = \sum_{j=1}^{3}\left\{[\cos k(2 - \lambda_j)^{1/2}t]A_jC + \left[\frac{\sin k(2 - \lambda_j)^{1/2}t}{k(2 - \lambda_j)^{1/2}}\right]A_jV\right\} \quad (2.5.13)$$

When we generalize the problem to $N - 1$ equal masses that divide a stretched spring into $N$ equal parts at equilibrium, then the $N - 1$ eigenvalues and constituent idempotents of $A$ are

$$\lambda_j = 2\cos\frac{j\pi}{N} \quad \text{and} \quad (2 - \lambda_j)^{1/2} = 2\sin\frac{j\pi}{2N} \quad (2.5.14)$$

where

$$(A_j)_{rs} = \frac{2}{N}\sin\frac{rj\pi}{N}\sin\frac{sj\pi}{N}$$

$r, s = 1, 2, \ldots, N - 1$

*(The third article in this series will appear in May.)*

# Heat storage for electric house heating

*Electric house heating involves short periods of high power demand. To ease production and distribution needs, a means of energy storage in the home is needed*

C. W. Bary *Consultant on Utility Economics*
R. E. Rice *Comstock and Wescott, Inc.*
J. F. Paquette, Jr. *Philadelphia Electric Company*

### *The advance of electric heating*

*The Electric Heating Association, a newly formed national trade association, recently released data that are indicative of the growing trend toward electrically heated homes. The EHA figures show that at present there are about 1.4 million electrically heated homes in the United States, or about five times as many as were in existence less than seven years ago. And over 20 per cent of all the new homes built in 1963 are electrically heated. Only four years ago the figure for new homes was less than half this number.*

*A forecast for 1980, based on the Federal Power Commission's National Power Survey Report No. 13, states that there will be 19 million electrically heated homes in 1980. At about 17 000 kWh per customer (the approximate difference between homes with and without electric space heating), this amounts to about 325 billion kWh or 12 per cent of the total energy requirements of 2.7 trillion kWh also forecast for 1980.*

*The many advantages of electric house heating have been widely publicized but little attention has been directed toward one fundamental disadvantage for the utilities—no economical means exists for storing large amounts of electricity on the customers' premises. The accompanying article reports on one possible solution to this problem.*

Electric power for house heating is being adopted at a rapidly increasing rate. Electric heat can be produced by resistance units, heat pumps, heat storage, radiant panel systems—and even by means of infrared devices. Each of these systems may be the sole supply of heat to the premises, or may be supplemented by any of the others. The installations may be at central locations on the premises, or in individual rooms, and they may be operated under fixed or variable settings of indoor thermostats. The systems can provide direct area heating or indirect heating by means of heated air circulated through ducts, or hot water or steam circulated through pipes and radiators.

The major economic handicap of electric domestic systems stems from the inherently low diversified annual load factor of house heating. This is caused by the high coincidence in demands of individual heating loads all operating at their maximum requirements during spells of extreme cold, but occurring infrequently during normal winter seasons in most areas of the country. Thus, the average heating season demands of house-heating installation groups are about one fifth the magnitude of their diversified peak demands. As electric house heating gains in popularity, it is becoming apparent that this progressively increasing load demand must be met in a manner that is of mutual benefit both to the consumer and the electric power industry.

The primary difference between electricity and combustible fuels is that electricity must be produced in exact response to the instantaneous demand—no economical method exists for storing large amounts of electricity. All combustible fuels, however, have inherent storage features between the points of production and consumption. These may be tanks, bins, or pipelines. Such fuels can be consumed for short periods at very high rates without creating correspondingly high demands on production and distribution facilities.

#### **Preliminary feasibility investigations**

As a first step toward the provision of an economical form of electric energy storage, the electric industry, through the agency of the former Joint AEIC-EEI Heat Pump Committee engaged Comstock & Wescott, Inc. (C&W) and Battelle Memorial Institute, six years ago, to investigate chemical and physical phenomena that might be employed for the storage of sizable quantities of energy on the premises of electrically heated houses. The investigations indicated the existence of promising possibilities that deserved further studies.[1] The utilities industry then authorized C&W to undertake a comprehensive research and development program.

As a result of this program, a system was selected, based on storage of sensible heat at moderately high temperatures. In this system, called the "Therm-Bank," a material is cycled through a temperature range which includes its melting point. This process employs the *heat of fusion* principle, and it uses sodium hydroxide—modified with corrosion inhibitors—as a heat storage medium.[2] The combination of chemical compounds is called "Therm-Keep."

The most recently completed phase of this effort has been the design, construction, and testing of a complete, full-sized pilot model for which the principles and materials that evolved from the preliminary work were pragmatically consolidated.

### Thermodynamic principles of operation

The heat storage material, Therm-Keep, is a solid at room temperature and a liquid at temperatures above 540°F. For application in the Therm-Bank, a maximum operating temperature of 900°F is used which is considerably below the boiling point of the material. Thus, the system does not develop pressure within the unit.

When heat is withdrawn from the Therm-Keep, sensible heat is extracted while the temperature drops from the top of 900°F down to 540°F, at which point the material solidifies, giving up its latent heat of fusion. As the temperature continues to drop, sensible heat and additional latent heat are delivered as the material undergoes a solid-phase transformation until the bottom operating temperature of 265°F is reached.

As compared to using hot water for the heat storage medium, the Therm-Keep has two distinct advantages. The sensible heat obtained from the temperature differential of 635°F between the top and bottom operating temperatures, together with the latent heat of fusion, gives the Therm-Keep a high heat capacity per unit of volume as compared to unpressurized water, which would have a much lower operating temperature differential for house-heating applications. Also, at the bottom operating temperature of 265°F for the Therm-Keep, heat may be withdrawn at a much higher rate than at the bottom temperature of an unpressurized water system. The combination of these factors reduces the volume of the Therm-Bank to a fraction of that required for an unpressurized water system of equal heat storage capacity with the minimum rate of heat withdrawal required for house heating.

### The Therm-Bank pilot model

The exterior envelope and essential components of the Therm-Bank pilot model are illustrated in Fig. 1. The Therm-Keep is contained in six welded-steel vessels called modules. The modules are equipped with electric resistors that heat the Therm-Keep. The modules are located within a steel inner tank, which, with its cover, forms a complete enclosure for these elements. Two kilowatts of electric-resistance heating are provided in each module, and 230-volt three-wire service is required for the heat storage system and its controls.

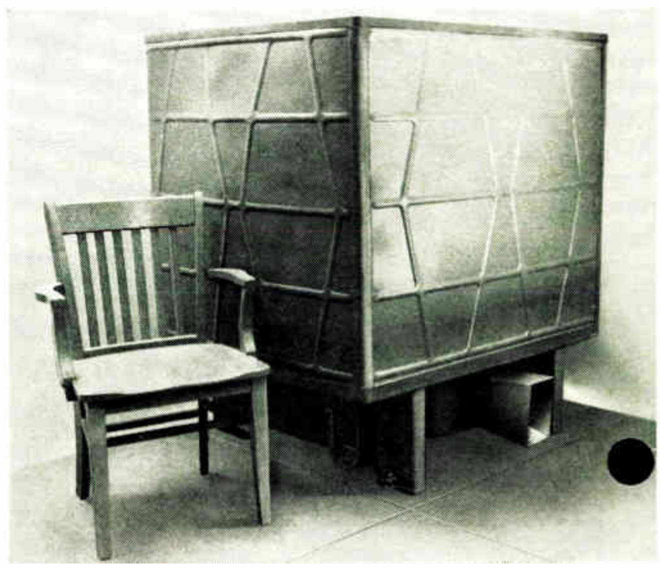
The inner tank contains two air passages that extend completely across two opposite sides, forming, respectively, the entrance and exit air passages through the thermal insulation. The tank is surrounded by a thick layer of thermal insulation to limit the heat loss from the system, and to prevent the exterior surfaces from being heated excessively. The outer sides and top of the insulation are protected by sheet-metal panels. The parts just described are supported by a base pan. Attached to the pan are right and left air ducts, an air bypass, a bypass damper, a blower for circulating air, and a control motor for operating the bypass damper.

### Mode of operation

The air to be heated enters the chamber that contains the bypass damper. To control the heat delivery rate and to maintain a uniform temperature of 155°F in the air delivered to the house-heating system, the hot air emerging from the inner tank must be mixed with cool air (except when the unit is delivering its last increment of useful heat). The bypass damper divides the entering air into two streams—one of which is admitted to the inlet air duct, travels through the air passage, enters the inner tank, and passes through the air spaces between the modules. The air becomes heated at this point, then exits through the other air passage and the outlet air duct, where it mixes with cool air from the bypass. The emerging 155°F air is drawn into the air blower, and delivered to the house-heating system carrier duct.

A thermostat is located near the blower outlet in the

Fig. 1. Fully assembled pilot model of a Therm-Bank heat storage unit. Its compact size may be noted by scale comparison to the chair. Cutaway view (right) of pilot model shows internal elements, components, and the path of air circulation throughout the unit.



air stream. The thermostat senses the air temperature, actuates the damper motor, and in turn positions the bypass damper. Thus, the ratio of air entering the inner tank to air entering the bypass is adjusted.

The maximum temperature of the modules is governed by a thermostatic control located in a well that extends into the top of one of the modules. This control is used to turn off the electric power to the modules at the top temperature of 900°F.

**Testing procedures**

An elaborate arrangement of instrumentation was utilized to take accurate measurements in the determination of the heating and heat delivery characteristics of the Therm-Bank. A total of 48 thermocouples and other devices was used to record internal and surface temperature distributions of the six modules. Measurements were made of the inlet and outlet air temperatures, humidity, pressure, and flow rate, to determine the heat delivery rates. Energy inputs were calibrated by a recording watt-hour meter.

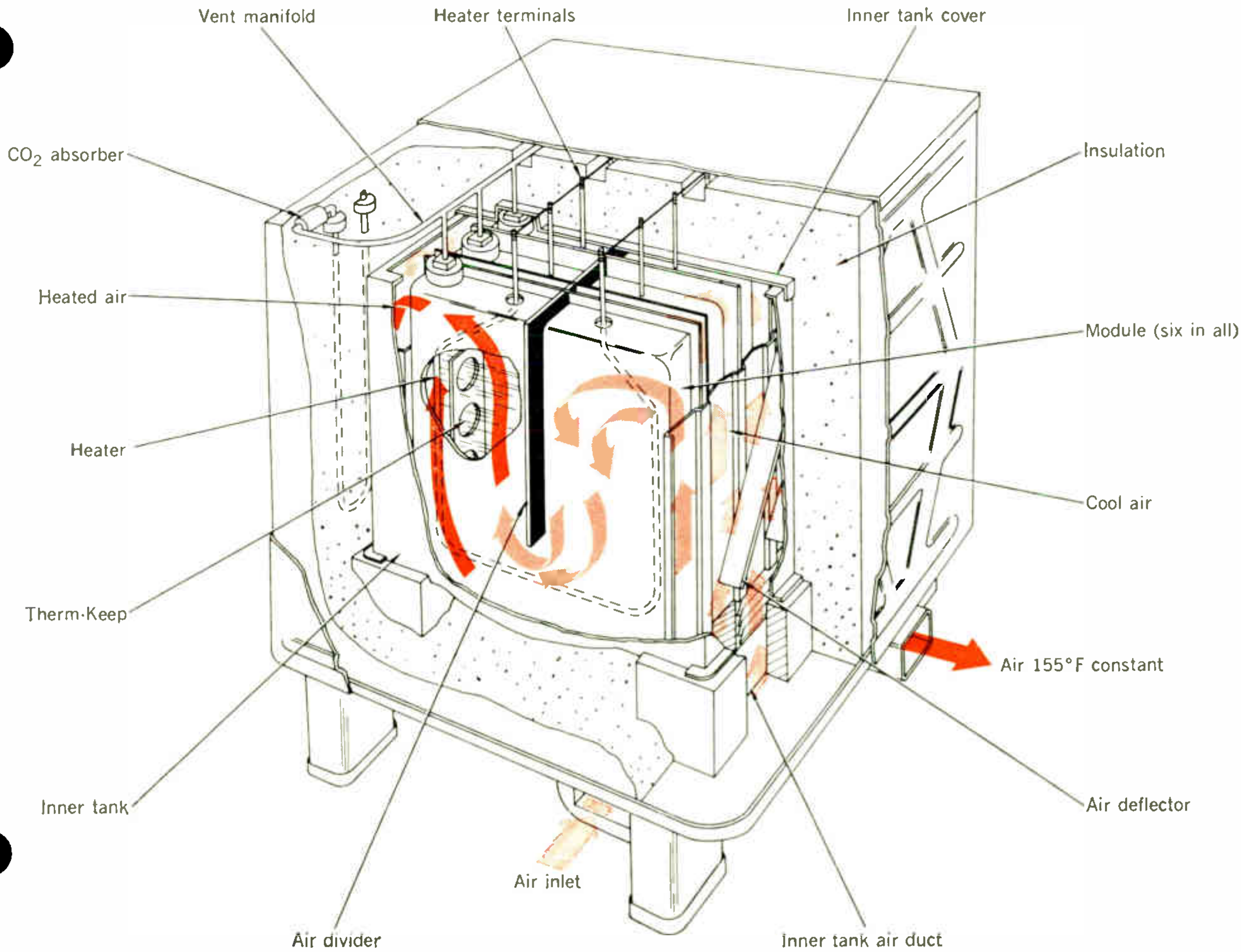
The basic experiments consisted of recharging tests from the "bottom" state, heat delivery tests from the "top" state, and determinations of stand-by heat losses.

**Test results**

The charging characteristics of the Therm-Bank pilot model, at power input of 6.0 and 11.4 kW to the entire system, are given by Fig. 2 for the two test conditions. For each test, uniform charging rates were maintained on each of the six modules of the system. The Fig. 2 graph shows that the temperature of the system increases at nearly twice the rate when charging at 11.4 kW. Based upon our experiments and those by others, we believe that a power input several times greater than 12 kW could be applied without changes in design other than substitution of heaters of appropriate wattage. Figure 3 illustrates the heat delivery rate from the top state. It may be seen that this rate remains substantially uniform at 37 500 Btu/h for approximately seven hours, then falls gradually for the next 4½ hours until the minimum specified heat delivery rate of 30 000 Btu/h is attained. During this 11½-hour period, the outlet air temperature is maintained at 155°F, and the total heat delivered amounts to 422 000 Btu.

**Safety and reliability**

From the conceptual phase through the research and development stages of the Therm-Bank, complete system safety has been a primary goal. An essential step in this achievement is the assurance that the heat storage medium will have no corrosive effects on the containers. Therefore, it was decided early in the developmental stages that the heat-storing medium–container combination must be one that had been tested and proved by actual service experience. Such ready-made proof existed



in a mixture of sodium hydroxide and sodium nitrate, chemical compounds long used by industry for cleaning and descaling metals. Essentially, this is the composition of Therm-Keep.

The corrosion resistivity afforded by the Therm-Keep assures a 20-year, trouble-free life expectancy for the Therm-Bank unit.

### Versatility of design

The basic design of the Therm-Bank permits adaptation to other heat storage requirements. For example, it can be converted into a completely self-contained electric hot-air furnace—with the heat storage feature—capable of meeting the entire house-heating requirement. This may be achieved either by direct resistance heating, storage, or a combination of both. The internal heat transfer characteristics between the electric heaters and the heat storage medium are such that heat can be charged into storage, if desired, at very high values. The Therm-Bank also can be coupled to a hydraulic system, by the delivery of its heat via hot water or steam rather than heated air.

The heat storage capacity of 400 000 Btu was selected to fit a particular set of conditions. Other parameters will require various capacities.

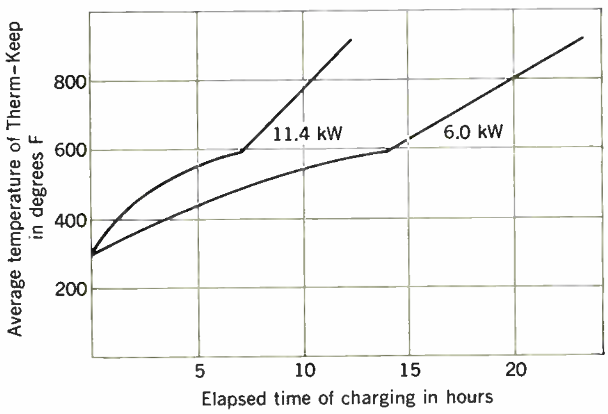
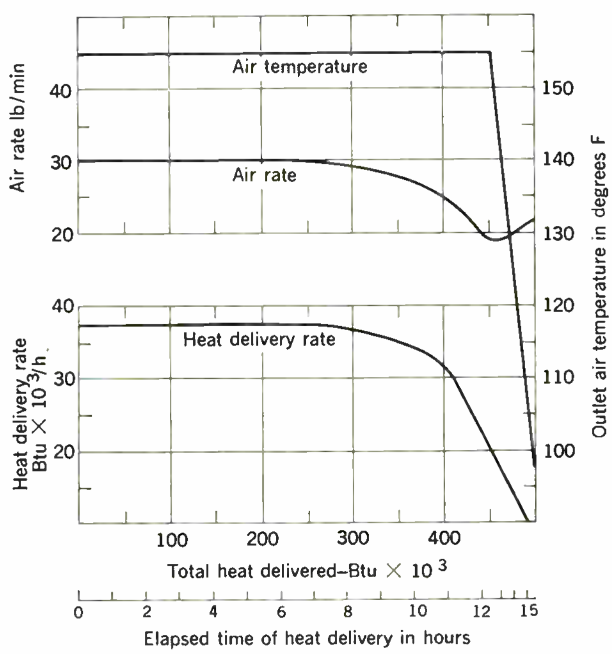



Fig. 2. Graph of performance characteristics for Therm-Bank unit during heat recharging tests at 11.4- and 6.0-kW power inputs. Average temperature (degrees F) of Therm-Keep is plotted against elapsed time of charging.

Fig. 3. Plot of performance characteristics of Therm-Bank unit pilot model during heat delivery tests.




### Summary of the R&D and testing programs

The overall results of the research and development, and the testing program have been very satisfactory. The Therm-Bank unit has emerged as a practical piece of equipment with these advantages:

1. It is completely self-contained, easily installed in any dwelling, and requires only electric and air duct connections.
2. It is of acceptable size, with overall dimensions of 54⅞ inches in height, 48⅛ inches in width, and 41⁷⁄₃₂ inches in depth.
3. It is readily adaptable to the need for varying heat storage capacities corresponding to different climatic conditions and house sizes. This versatility is afforded by the modular design of the Therm-Keep containers.
4. The unit is charged with heat electrically; it is flameless, and produces no by-products or waste materials.
5. The estimated installed cost to the consumer is not more than $1.50 per thousand Btu of heat storage capacity.
6. It employs a heat storage medium that will never lose its ability to store and deliver heat equivalent to its full-rated storage capacity.
7. The unit is subject to heat losses, at its maximum operating temperature, that are acceptable from the viewpoints of economical space heating and operation.
8. The device can deliver its entire useful heat capacity of 420 000 Btu at the specified rate of not less than 30 000 Btu/h, and can deliver an additional 78 000 Btu at the rate of about 20 000 Btu/h.

### Evaluation of Therm-Bank

In its various modifications the Therm-Bank can be a useful device for the transformation of the steadily increasing electric house-heating load into a beneficial new market for electric power. As a supplement to a basic electric heating system, it permits a reduction in the connected load of the base system—plus the use of regulated amounts of energy for storage during "valleys" of the electric system's daily load curves. When modified for use as an electric furnace with storage, it permits the complete restriction, for several hours each day, of all electric power for house heating as a means of "peak shaving."


REFERENCES

1. "First Step on Heat Storage Research Program Completed," Prog. Rept., Joint AEIC-EEI Heat Pump Committee. *Bull.*, Edison Electric Institute, New York, June 1959, p. 250.
2. "Promising Heat Storage Material Found Through Research," *Ibid.*, Feb. 1962, pp. 49–50.



(This article is based on CP64-121 and CP64-122, presented at the IEEE Winter Power Meeting, New York, N.Y., Feb. 2–7, 1964.)

# To all IEEE Members

It is my privilege to present to the IEEE membership a policy statement, recently approved by the Board of Directors, which deals with the relations between IEEE and other technical and professional organizations. It establishes a position that defines the right of free choice of every member and recognizes the benefits of joint endeavors.

The policy was developed by a special committee appointed last June, whose members were L. V. Berkner, C. V. Carlson, W. H. Chase, Seymour Herwald, F. K. Willenbrock, and the undersigned. In reporting to the Board, this committee presented comments which are printed here at the conclusion of the policy statement.

This subject deserves the attention of all IEEE members who are concerned with the complex interrelations of our Institute with other societies and agencies throughout the world. IEEE has no desire to live in a vacuum. Joint responsibilities with others in science, education, and professional development will, I feel sure, be upheld by actions based on this policy.

Clarence H. Linder
*President*

# IEEE policy regarding professional relations

IEEE is a nonnational society devoted to the extension and dissemination of technical and scientific knowledge in the science and technology of electricity, electronics, and related fields. Members of IEEE have equal privileges, and the policies of IEEE should reflect this equality of recognition and opportunity. Unlike an international society whose members are represented through national agencies or nationally organized committees, the nonnational membership character of IEEE involves representation through arbitrarily designated "regions" of the world not coinciding with national boundaries, and through directly elected officers and directors. Its activities are world-wide.

In addition to the dissemination of technical information to its members, the IEEE will benefit its membership and the profession by taking an active role in association with other organizations in activities which relate to its fields of interest. These activities can be categorized as those concerning technical, educational, and professional matters. The Board of Directors has adopted the following broad policies for the guidance of the Executive Committee; component organizations of the IEEE, such as Groups and Sections; and the Headquarters Staff.

There are two general groupings of organizations whose activities relate to the interests of the IEEE: (A) international or nonnational groups whose activities extend across national boundaries, and (B) national groups.

A. International groups fall generally into two categories:

**a.** Intergovernmental groups, e.g., (1) United Nations and its specialized agencies, such as United Nations Educational, Scientific and Cultural Organization (UNESCO); World Meteorological Organization (WMO); and International Telecommunication Union (ITU), including International Telegraph and Telephone Consultative Committee (CCITT), International Radio Consultative Committee (CCIR), and International

Frequency Registration Board (IFRB); and (2) Science Committee of North American Treaty Organization (NATO).

**b.** Nongovernmental international groups, e.g., (1) International Council of Scientific Unions (ICSU), and its affiliated activities such as International Scientific Radio Union (URSI), International Union of Geodesy and Geophysics (IUGG), Scientific Committee on Space Research (COSPAR), Scientific Committee for Antarctic Research (SCAR), International Geophysical Committee (CIG), Federation of Astronomical and Geophysical Permanent Services (FAGS); (2) Conférence des Sociétés d'Ingénieurs de l'Europe Occidentale et des Etats-Unis d'Amérique (EUSEC); (3) Union Panamericana de Asociaciones de Ingenieros (UPADI); (4) International Electro-Technical Commission (IEC); and (5) Union of International Engineering Organizations (UIEO).

B. National groups represented by national societies, associations, and organizations, both governmental and nongovernmental:

**a.** Major technical and scientific societies such as the American Society of Mechanical Engineers (ASME), American Society of Civil Engineers (ASCE), American Physical Society (APS), American Chemical Society (ACS), the British Institution of Electrical Engineers, and the Canadian ECC.

**b.** National academies such as the United States National Academy of Science–National Research Council, the Royal Society of Great Britain, and the National Research Council of Canada.

**c.** Engineering Councils and central committees such as Engineers Joint Council (EJC), and Engineers' Council for Professional Development (ECPD).

**d.** Regional associations of local engineering societies.

**e.** Industrial associations, both national and regional.

**f.** National committees of international organizations such as URSI, IUGG, etc.

### General policy concerning relations with other organizations

1. In furtherance of IEEE objectives, and with the approval of the Executive Committee, the IEEE and its component organizations can have membership in, representation on, or association with international, national, or local groups of recognized standing concerned with technical or educational matters related to IEEE fields of interest or professional matters consistent with the nonnational character of the IEEE.

2. In fulfillment of its responsibility to society or in response to requests for information, the IEEE shall take public positions on those technical or educational questions in which it has recognized competence. It is essential that all important viewpoints, including minority opinions, be represented in such statements of public positions.

3. All public statements of IEEE positions shall be approved in advance by the Executive Committee.

4. The representation of the IEEE or its component organizations in any councils or central committees of which it is a member shall be effective in protecting and forwarding IEEE policies. In cases where public statements are issued to which the IEEE is opposed, the IEEE shall publicize its position if appropriate.

5. The IEEE and its component organizations shall not participate directly in activities primarily directed for political objectives or the influencing of international, national, or local legislation.

6. The IEEE and its component organizations shall be selective in participating in activities with other organizations to avoid a diversion of manpower and financial resources from the primary objectives of the Institute.

### Statement of policy concerning technical activities

Technical activities relate to publications, meetings, symposia, exhibits, nomenclature, symbols, and standards.

Policies governing IEEE technical activities which involve other organizations are:

1. Wherever such action will advance the purposes of the IEEE, it should initiate or cooperate actively in conjunction with national or international groups of recognized standing, in sponsoring national or international meetings, exhibits, symposia, or other informational activities.

2. Component organizations of the IEEE should actively cooperate for the advancement of IEEE objectives with other associations having related technical interests. Joint membership or affiliation with such associations shall be approved by the Executive Committee.

3. Where appropriate, and subject to the approval of the Executive Committee, the IEEE and its component organizations shall take the initiative in collaboration with national and international groups to further standardization of nomenclature, symbols, measurements, and techniques.

### Statement of policy concerning educational activities

Since one of the major functions of the IEEE is the improvement of the technical competence of its members, it is appropriate for the Institute to cooperate with those agencies whose activities relate to educational matters in electrical engineering, electronics, and related fields. These agencies may be national, international, or nonnational organizations of a governmental or nongovernmental character.

Policies governing IEEE educational activities with other organizations are:

1. The IEEE shall actively participate in any national or international organization concerned with technical or scientific education in fields of IEEE interest. For example, in the USA, the IEEE shall be active in the ECPD since it accredits engineering curricula in academic institutions.

2. The IEEE and its component organizations shall aid those academic institutions offering instruction in IEEE fields of interest in keeping abreast of the rapid advances in science and technology in any appropriate fashion.

3. With the approval of the Executive Committee, component organizations of the IEEE, such as the Group on Education, shall cooperate with or join, as appropriate, corresponding groups having related interests in education, such as the American Society for Engineering Education.

4. The standing committee on Education (Bylaw

409.4) shall formulate statements of the IEEE position on educational matters and with the approval of the Executive Committee these statements shall be publicized in the promotion of the interests of the IEEE and its members.

5. In cooperation with other organizations, or on its own initiative, the IEEE should supply career information in its fields of interests to students, teachers, and guidance counselors at the preparatory school level.

6. In conjunction with other technical or scientific organizations, or on its own initiative, the IEEE shall be active in providing for its members education in addition to the formal courses of instruction offered by academic institutions. This continuing education shall be directed toward the prevention of technical obsolescence and shall be designed to assist the members to keep abreast of the rapid advances in science and technology.

**Statement of policy concerning professional activities**

Professional activities are those which relate to engineering as a profession such as registration as a professional engineer, codes of ethics, and professional conduct. These activities can be separated into internal and external categories. An example of an internal professional activity would be the definition of professional conduct within the IEEE. Internal activities are the concern of the Professional Relations Committee. An example of external professional activity is cooperation with State Boards of Registration in the development of appropriate technical examinations. The requirements of the C-3 tax status of the IEEE may prevent it from taking direct action in many areas of external professional activity.

Policies relating to professional activities are:

1. The Executive Committee will be responsible for keeping the IEEE membership informed on the legal, ethical, and other professional aspects of the practice of engineering which are of importance to any major segment of the membership.

2. With the approval of the Executive Committee, the Professional Relations Committee shall draft statements of the IEEE position in external professional matters, but these position statements shall not be transmitted to the membership without prior approval of the Board of Directors.

3. The Professional Relations Committee (Bylaw 409.12) shall maintain channels of information open with such organizations as the NSPE and NCSBEE in the USA and equivalent organizations in other countries which are directly concerned with external professional matters. In response to questions or on its own initiative, the Executive Committee shall inform these organizations about the position of IEEE.

# Report of the Ad Hoc Committee on Professional Relations Policy

Dr. Ernst Weber, President
The Institute of Electrical and Electronics Engineers, Inc.
Box A, Lenox Hill Station
New York 21, New York

Dear Dr. Weber:

In response to your letter of June 3, 1963, an Ad Hoc Committee on Professional Relations Policy consisting of C. Linder (Chairman), L. V. Berkner, C. Carlson, W. Chase, R. Chipp, S. Herwald, H. Lowe, F. K. Willenbrock (Secretary), with C. Savage as a consultant, has met and wishes to submit its report in three parts. In this covering letter, we shall include those recommendations which are of a current nature and which, in our opinion, require immediate action. The second part is our formal report to the Executive Committee which includes as its first recommendation that the Board of Directors adopt a series of policies governing its relationship with other organizations and that these policies be set forth in a unified document and be made available to the IEEE membership. The third part is the proposed statement of policies and it is our recommendation that these policies be adopted by the Board of Directors and be promulgated to the membership. Consequently, the major recommendations of this Ad Hoc Committee are to be found in this policy statement.

You will recall that on May 15, 1963, the Professional Relations Committee submitted to the Executive Committee a policy statement in connection with the registration of engineers and that this statement was referred to this Ad Hoc Committee for consideration. The broad policy statements included in our report discuss the question of professional relations. However, we feel that the Professional Relations Committee should be requested to revise the statement of May 15, and prepare for Board approval early in 1964 a detailed statement of the IEEE's position on legislation concerning the practice of engineering.

Our recommendations for immediate action are as follows:

1. That legal counsel review the proposed policy statement in the light of the current c-3 tax status of the IEEE to see that none of the policies set forth endanger this status.

2. That legal counsel review the IEEE's present affiliations with other societies such as the National Council of State Boards of Engineering Examiners and

determine whether any of these relationships endanger the c-3 tax status.

3. That the distinction between activities appropriate to a technical society such as IEEE and the activities appropriate to organizations primarily concerned with the legal, economic, and public relations status of engineering as a profession be clearly recognized and preserved. For those of its members who are actively interested in these latter aspects of engineering as a profession and in professional registration, the IEEE should recommend that they also consider membership in such professional organizations.

4. That the Board of Directors recognize the need for legislation concerning the practice of engineering when such practice relates to the health, welfare, and safety of the public. Furthermore, that such legislation should be as uniform as possible in all political subdivisions to permit the maximum mobility for members of the engineering profession.

5. That the IEEE make a public statement relevant to the formation of a National Academy of Engineering and specifying what principles should be basic in its selection of members.

6. That the IEEE, in view of the recent consolidation of its student branches, broaden its relations with academic institutions offering instruction in fields of IEEE interest and be of greater assistance to academic administrators in the advancement of scientific and technical education.

7. That the IEEE recognize the vital need of its members for help in keeping abreast of the rapid advances in science and technology in their fields of interest and, in conjunction with other organizations or on its own initiative, be active in the development of new methods of providing for their continuing education.

8. That the Intersociety Relations Committee review all external relationships of the IEEE* and withdraw from those associations which have proved ineffective and strengthen its representation in those organizations which relate to vital interests of the IEEE and its membership. In particular, the IEEE representatives to the ECPD in its activity of academic cirricula accreditation should be selected from the most forward-looking members of the Institute for their technical competence and effectiveness in forwarding IEEE objectives.

Sincerely yours,
/s/ Clarence H. Linder
Chairman, Ad Hoc Committee
on Professional Relations Policy

* Lists of organizations to which AIEE and IRE made financial contributions in 1962 and also a list of those organizations in which they have had representation are being forwarded in a supplementary mailing.

The Ad Hoc Committee on Professional Relations Policy was appointed by President Weber at the request of the Executive Committee in June 1963, to study the relations of the IEEE with various other professional societies and organizations and to recommend a position which recognizes the free choice of each member and also the useful aspects of group endeavors.

The Committee has discussed the question of the extent to which the Board of Directors can speak for the membership. The Board, as a duly elected body, represents the consolidated voice of the membership and, while it cannot speak for an individual member, it is the only body which is representative of the membership as a whole. While the Board should be sensitive to the thinking of large segments of the membership on major questions, it has the responsibility to rise above the parochial opinions of particular individuals or groups and act in the broadest interest of the Institute. The Board should be assisted and guided by the various committees in the Institute's structure whose members are selected for their competence in particular specialties and their representation of the various segments of the membership. In taking a position on a question on which there is considerable difference of opinion, it is recommended that the Board use the terminology "The Board of Directors of the IEEE endorses . . ." rather than "The IEEE endorses . . . ."

The Committee recommends to the Executive Committee:

1. That the Board of Directors adopt a series of policies governing its relationships with other professional societies and organizations, and that these policies be set forth in a unified document available to the IEEE membership. A proposed statement of these policies is attached to this report.

2. That the Executive Committee shall develop methods for implementing the IEEE policies on technical, educational, and professional matters which relate to the interests of the IEEE. Then by public statements, by representation on councils, or central committees, or by other appropriate means, the position of the IEEE should be made known and advanced.

3. That the Intersociety Relations Committee (Bylaw 409.3) be requested to study the representation of IEEE on various councils and central committees and make specific recommendations to the Executive Committee as to the most appropriate form of representation. The representation of the IEEE must be effective in protecting and forwarding the interests of IEEE and its membership.

4. That any action in connection with relationships with other organizations in exception to the policies enumerated in the attached statement must be approved by the Board of Directors.

# Authors

**E. C. Jordan** (F) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Alberta in 1934 and 1936, respectively, and the Ph.D. degree from Ohio State University in 1940. He was control operator at radio station CKUA for seven years, electrical engineer for International Nickel Company for two years, and since 1940 has taught electrical engineering at Worcester Polytechnic Institute, Ohio State University, and the University of Illinois. At the latter two universities he supervised research on antennas and radio direction finding. At present he is a professor and head of the Department of Electrical Engineering at the University of Illinois. Dr. Jordan is author, coauthor, or editor of several books on antennas, electromagnetic theory, and radio and electronics. He is a member of the U.S. National Committee of URSI and past chairman of U.S. Commission VI on Radio Waves and Circuits.

**G. A. Deschamps** (F) was born and educated in France. He was graduated from the Ecole Normale Supérieure, Paris, in 1934, and received advanced degrees in mathematics and physics from the Sorbonne. He taught mathematics and physics for about ten years at the Lycée Francais de New York. In 1947 he joined the Federal Telecommunication Laboratories, where he worked as a project engineer on direction-finding systems, design of high-frequency and microwave antennas, microstrip development, and radio and inertial navigation. In 1956 he was appointed a senior scientist of the ITT Laboratories. He joined the University of Illinois in 1958 as professor of electrical engineering and director of the Antenna Laboratory.

Prof. Deschamps is a member of the American Physical Society and the International Scientific Radio Union (URSI). He was chairman of the IRE Committee on Antennas and Waveguides from 1957 to 1958 and editor of the IRE TRANSACTIONS ON INFORMATION THEORY from 1958 to 1960.

**J. D. Dyson** (SM) received the B.S. degree in economics from South Dakota State College in 1940, was employed for one year as a statistician, and then served on active duty with the U.S. Army from 1941 to 1946. He received the B.S. degree in electrical engineering from South Dakota State College, where he also was a part-time instructor, in 1949, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois in 1950 and 1957, respectively. He was on the research staff of the Sandia Corporation, Albuquerque, N. Mex., from 1951 to 1952. Since October 1952 he has been on the faculty of the Electrical Engineering Department of the University of Illinois, where he is now a research associate professor, devoting half time to research work in the Antenna Laboratory and half time to teaching courses in electromagnetic theory and microwave measurements. He is a member of Sigma Xi, Eta Kappa Nu, Sigma Tau, and Pi Mu Epsilon.

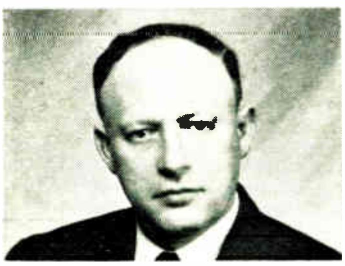
**P. E. Mayes** (M) joined the electrical engineering faculty at the University of Illinois in 1954 and is now a professor, teaching graduate courses in electromagnetic theory and supervising research in the Antenna Laboratory. He received the B.S.E.E. degree from the University of Oklahoma in 1950. He was employed as a graduate assistant and research associate in the Microwave Laboratory at Northwestern University while a graduate student there from 1950 to 1954. He received the M.S. degree in 1952 and the Ph.D. degree in 1955, both from Northwestern. His graduate research work was related to electromagnetic wave propagation along open waveguides and reflection from curved surfaces. At the University of Illinois he has worked on slot antennas, pattern synthesis, and several kinds of frequency-independent antennas. Dr. Mayes has served as consultant to a number of antenna firms and holds several patents in the antenna field.

**A. E. Lessor** received the B.S. degree in chemistry from Union College in 1949 and the Ph.D. degree in chemistry from Indiana University in 1955. After working as a consultant on analytical and physical chemical problems at General Electric Co., he joined IBM in 1959 as manager of crystallographic services. He conducted X-ray diffraction and optical measurements and also directed independent studies of tin and indium films. He was later appointed manager of thin-film materials development, and subsequently of evaporated film development. As manager of film electronics development he is responsible for thin-film network materials and processes.

**L. I. Maissel** (M) was born in Cape Town, South Africa, in 1930. He received the B.Sc. degree in physics and chemistry and the M.Sc. degree in physics in 1949 and 1951, respectively, from the University of Cape Town. In 1955 he received the Ph.D. degree from the Imperial College of Science and Technology, London, for work in optical spectroscopy. During 1956–1960, he was a project physicist with the Philco Corporation, where he worked on semiconductor materials and microminiaturization. He joined IBM in 1960, and is now a senior physicist, directing work on tantalum integrated circuits and exploring new types of sputtered films and techniques.

**R. E. Thun** (SM), manager of components development at the IBM Space Guidance Center, Owego, N.Y., is responsible for the development of thin-film technologies, magnetic-film storage devices, and integrated circuits. Since joining IBM in 1959, he has worked on problems related to physics of thin films, vacuum technology, electron optics, and computers. He received the Ph.D. in physics from the University of Frankfort-on-the-Main, Germany, and was subsequently engaged in research in metal physics and electron diffraction in Germany. He later worked as a research physicist at the U.S. Army Research and Development Laboratories.

**S. M. Fine** (M) is an associate professor of electrical engineering at Northeastern University. He received the B.A.Sc. degree from the University of Toronto in 1946 and the S.M. degree from the Massachusetts Institute of Technology in 1953, both in electrical engineering, and the M.D. degree from the University of Toronto in 1957. He interned at the Edward J. Meyer Memorial Hospital, Buffalo, N.Y. He has been associated with MIT's Research Laboratory of Electronics, the National Institutes of Health, and Brookhaven National Laboratory. His interests are in biomedical engineering and effects of radiation on biological systems. He is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and the Society of Nuclear Medicine.



**E. Klein** received the B.A. degree in physiology and biochemistry from University College, Toronto, in 1947 and the M.D. degree from the University of Toronto in 1951. He was awarded a two-year National Research Council fellowship at the Laboratories for Physical Chemistry at Harvard University and the Cilhdrens' Cancer Research Foundation, Boston. He was a research associate at that institution from 1953 to 1961 and has been a consultant since then. He has been associated with Harvard Medical School, Massachusetts General Hospital, and Tufts University. He is now chief of dermatology, Roswell Park Memorial Institute, and associate professor of experimental pathology at the N.Y. State University at Buffalo.

**R. E. Scott** (M) received the B.A.Sc. degree in 1943 and the M.A.Sc. degree in 1946, both from the University of Toronto, and the Sc.D. degree from MIT in 1950. From 1943 to 1945 he served as a radar officer in the Royal Canadian Navy. He has held the positions of instructor at the University of Toronto, and research assistant, research associate, and assistant professor at MIT. From 1954 to 1955 he was employed at Trans-Sonics, Inc., and then returned to the field of education to become an associate professor, later a professor, at Northeastern University. At present he is dean of the College of Engineering at Northeastern. Dr. Scott is a member of Beta Gamma Epsilon, Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

**Earl Ewald** (SM) received the B.E.E. degree from the University of Minnesota in 1930. After graduation, he began his employment with the Northern States Power Company, Minneapolis, and during the next 24 years progressed through a number of engineering and management positions within the company. In 1954 he was elected vice president in charge of operation and in 1962 was elected executive vice president. Mr. Ewald is a member of the National Society of Professional Engineers, and is a registered professional engineer in the states of Minnesota, Wisconsin, and South Dakota. He also is a member of the board of directors of the Northern States Power Company, a member of the advisory committee of the Columbia University Utility Management Workshop, and a vice president of the Minnesota Safety Council.

**D. W. Angland** (SM) was born in Minneapolis, Minn., on March 2, 1922. He attended the University of Minnesota, from which he received the B.S. degree in electrical engineering in 1945 and the M.S. degree in electrical engineering in 1951.

He was an instructor in the Electrical Engineering Department at the University of Minnesota from 1946 to 1951 and has since been with the Northern States Power Company, where he has held various positions in the Engineering Department and Operating Department. He is now manager of planning and is responsible for the development of electric system expansion plans and for the general supervision of the company's construction budget. Mr. Angland is a member of Eta Kappa Nu and Tau Beta Pi.

**J. S. Frame.** A biographical sketch of Dr. Frame appears on page 239 of the March issue.

**C. W. Bary** (F, L) is an independent consultant on economics of public utilities, and a member of the economics committee of Atomic Power Development Associates, Inc. Prior to his recent retirement from Philadelphia Electric Company, he was its director of economic, cost, and rate analysis, having been affiliated with the company for over 40 years in various positions. He received the B.S. degree in electrical engineering in 1927 from the Massachusetts Institute of Technology. He has made many significant contributions to the electrical industry. In 1937, for example, he formulated the second law of load diversity, now known as the "Bary curve," which is used as an element in the cost bases for designing rates. He has served the U.S. government in numerous capacities. He is the author of a book and many technical papers.

**R. E. Rice** received the S.B. degree in chemical engineering from MIT in 1935 and subsequently attended Northeastern University Graduate School. He joined Comstock & Wescott, Inc., in 1935 to work on natural color printing processes, and invented and patented a process for printing color pictures from transparencies. Projects in which he has participated include development of color cameras, studies of metal corrosion and tarnishing, and development of explosion-actuated tools. He is now vice president and director of research at Comstock & Wescott and since 1958 has directed a program on the development of a practical method for the storage of heat in electrically heated homes. He is a member of the National Academy of Science, National Research Council, and the Committee on Personnel Armor.

**J. F. Paquette, Jr.** is a graduate of Yale University, from which he received the bachelor's degree in civil engineering in 1956. He has been taking courses and is at present preparing his thesis for a master's degree in business administration from the evening school of business at Temple University, Philadelphia.

Since 1956 Mr. Paquette has been associated with The Philadelphia Electric Company, where his principal activities have been in the fields of load research studies, cost-to-serve determinations, and economic analysis. He is currently supervisor of the Supply Contracts Section in the Rate Division of that company. He is also vice president of the Philadelphia section of the Yale Engineering Association.

# IEEE publications

**scanning the issues**
**advance abstracts**
**translated journals**
**special publications**

## Scanning the issues

**Ferment in Education.** Those who view the present academic scene as "unchanging" should give some thought to the contents of the most recent IEEE TRANSACTIONS ON EDUCATION.

The first section of the issue is comprised of four papers which consider, from different vantage points, some questions pertinent to the problems of "continuing education" (not to be confused with graduate education) for engineers several years beyond their last academic degree. Another section presents two papers which deal with off-campus programs in graduate engineering education from diametrically opposite viewpoints—just how opposite is suggested by the title of the first paper, "Parricidal Implications of Off-Campus Engineering Education."

A third section of this volume is concerned with some of the accomplishments to date in a relatively new aspect of engineering education—that of programmed teaching. And finally, the contributions section introduces a number of interesting new approaches to the more effective presentation of subjects which have been taught long, but very differently, at both the undergraduate and graduate levels.

Returning to the first subject, "continuing education," the implications are so broad and far-reaching for our profession that it is appropriate to repeat here the introduction to the papers on this topic.

The onrush of technology, the rapid emergence of new science, and the transition of this science to new engineering has placed no small segment of the engineering profession in an anachronistic dilemma. Many of the 300 000 pre-1952 engineering graduates are finding that the monumental changes which have occurred in engineering education over the last decade have also introduced new values, literally changing the "rules of the game." Intuition, experience, and professional maturity, which were once at a premium, are fast losing ground to those skills which, stemming from modern engineering knowledge, are providing radically new and different devices and systems—even new industries.

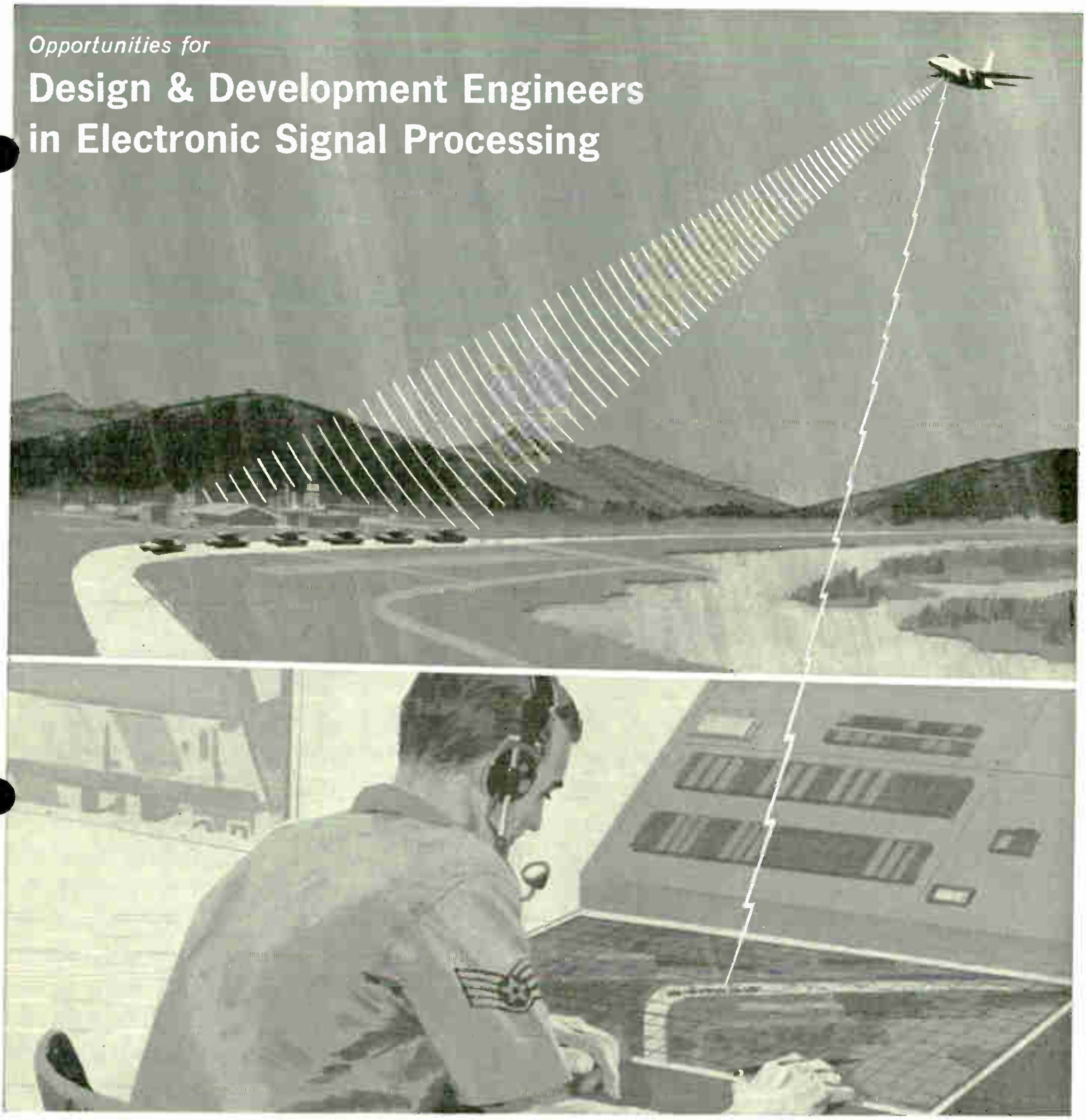
Continuing education, insofar as its effect on and relationship to practicing professionals is concerned, is being regarded in a new light. Suggested here is not a "watered-down" version of existing graduate courses, but a new educational domain concerned with the recurrent updating, modernizing, and strengthening of the technical backgrounds of the whole manpower continuum, i.e., the engineering profession. The success of this effort will, to a large degree, depend upon our ability to keep a clear focus on the continuing professional study needs of experienced engineers. These needs have not and will not be satisfied by the "credit-sequence" approach, common to well-established degree programs.

There is growing evidence that self-contained blocks or modules of modern scientific and engineering knowledge can be presented effectively within practical economic and time constraints. Encouraging results have been realized from the General Electric and U.C.L.A. programs aimed at engineering executives. The Polytechnic Institute of Brooklyn is also devising an Executive Technical Development program for industrial personnel and, in addition, has instituted this fall a program of somewhat greater depth for practicing engineers with 10 to 15 years of experience.

The possible "payoff" from these beginning efforts to industry, education, and members of the profession is of no mean proportion. The more effective utilization of engineering talent; a resurgence of engineering contributions from experienced members of the profession; new teaching methods and structures adapted to the specific needs of mature individuals; new textbooks, teaching aids, and monographs all embodying self-learning—these are but a few of the realizable results. (G.E. Moore, "The importance of continuing education," *IEEE Trans. on Education*, December 1963.)

**Computers the World Over.** As a nonnational organization, it is not unusual that the IEEE should devote a considerable number of its publication pages to developments outside the North American continent. The December issue of IEEE TRANSACTIONS ON ELECTRONIC COMPUTERS is nevertheless worthy of special note. The 370-page issue features a truly international collection of 28 invited and contributed papers on computer systems. The papers come from Poland, Germany, Israel, Denmark, Sweden, Australia, Italy, France, Czechoslovakia, and Japan, as well as the United States, providing a unique picture of the latest developments around the world not only in conventional computers, but also in multicomputer systems, list processors, hybrids, variable structure systems, pattern recognition systems, and many other fields of current interest not hitherto described in the open literature.

**Time-Frequency Duality.** The concept of duality has proved quite useful in the analysis and synthesis of lumped electric networks. This usefulness stems from the fact that two situations, which are entirely analogous on a current and voltage basis, respectively, have identical behavior patterns, except for an interchange of the roles played by voltage and current, while physically and geometrically they are distinctly different. Thus, a means is provided for




Design and development activities in the field of Electronic Signal Processing are rapidly expanding today at HUGHES Aerospace Divisions.

Development of systems utilizing advanced correlation and matched filter techniques for *High-Resolution Radar*, *Acoustic Detection & Classification* and *Pulse Doppler Radar* is being accelerated.

Specialists in Signal Processing, Circuit Design, Mechanical Design, Packaging Design, Performance Analysis and Project Engineering will be interested in the outstanding assignments now available.

Graduate engineers with experience in wide-band video amplifiers; high-resolution cathode ray tube circuits and applications (including ultra-linear sweep, gamma correction and dynamic focus); high-voltage power supplies; low-jitter timing circuitry; high-speed analog sampling circuitry; precision film transports; ultra-high speed film development; scan conversion systems; synthetic array radar systems; imagery recording, or similar fields—are invited to submit resumes.

For immediate consideration please write:

**Mr. Robert A. Martin**
**Head of Employment**
**HUGHES Aerospace Divisions**
**11940 W. Jefferson Blvd.**
**Culver City 47, Calif.**

*Creating a new world with electronics*

**HUGHES**

HUGHES AIRCRAFT COMPANY
AEROSPACE DIVISIONS

An equal opportunity employer

U. S. CITIZENSHIP REQUIRED

# Focal points

**industry**
**education**
**government**

## Promising one-step conversion system uses electrically charged fission-fragments directly

The Rand Corporation, of Santa Monica, California, and Dr. George Safonov of Advanced Concepts Technology, Inc., have described the results of seven years of experimentation on the direct conversion to electricity of energy from nuclear fission.

The announcement revealed work on the new atomic power concept that holds promise for unique future applications, particularly in space.

Conventional atomic power plants employ a three-step scheme for converting fissile fuel energy into electric energy. Fission-fragment energy is first converted to heat energy, which produces high-temperature steam. Second, a steam turbine converts the heat to mechanical energy. Third, the turbine drives a generator, and mechanical energy is converted to electricity. Two-step schemes, under active development by several industrial laboratories, first degrade fragment energy to heat and thence convert the heat into electricity. One-step conversion would aim to use electrically charged fission fragments directly.

Early design studies in 1956 led to a first, small-scale experiment in 1958 that encouraged the further exploration of the possibilities of the one-step energy conversion process. While subsequent experimental research has demonstrated progressively improved performance, the recent announcement emphasizes the elementary state of the new concept's technology. Thus, the work at this time is entering into a new experimental and administrative phase devoted to the achievement of full-scale direct electricity reactors.

One-step, "one-temperature," and temperature-independent conversion of fissile to electric energy may be achieved by the immediate use of newly born fission fragments. This most direct means of conversion has been under investigation theoretically and experimentally since the inception of the program more than seven years ago. The basic mechanism of one-step conversion is illustrated by considering a system wherein one sphere, coated with fissile material, is contained within the vacuum envelope of a second sphere. Fragments born from fission on the inner sphere are collected by the outer sphere. Since the fragments are electrically charged, the two spheres develop an electric potential between them. And because of their high "birth" energy, potentials up to 4 million volts may result from fragment charging. The electric field achieved from charge accumulation on the spheres acts to decelerate fragments as they move from the emitter to the collector. Consequently, a portion of the fragment kinetic energy is converted, in a single step, to the electrostatic energy stored by the spherical condenser system. Essentially, the one-step converter is self-charging. It may discharge its electricity into an external load at a rate commensurate with the internal charging maintained by new-born fission fragments.

Since the conversion mechanism of the one-step scheme is uniquely independent of system temperatures, this converter may be operated substantially as a one-temperature machine without compromising conversion performance. For example, in space applications, where all types of converters must reject their waste heat by thermal radiation, the one-step system would be operated as a very high one-temperature machine since radiation cooling improves sharply when radiator temperature is increased. By contrast, the multi-step—or conventional—machines must necessarily reject their waste heat at the lower of their two basic temperatures.

Several other features, peculiar to the one-step system, make for ready adaptations to the accomplishment of space missions. Among these are the system's natural requirement for the vacuum of space as an insulating medium, its minimal requirements for vital working fluids that might leak to space, its promise of reliability by virtue of ultimate simplicity in conversion machinery, and its characteristically high-voltage electricity that lends itself naturally to the advanced ion-propulsion techniques under development for deep-space missions.

Since waste heat from the one-step converter can be rejected at high temperatures without affecting the conversion mechanism involved, it may form a valuable by-product of one-step converters. For instance, it may be used as feed energy by the less direct heat-to-electricity converters. Thus, the one-temperature aspect of the one-step scheme permits its complementary operation with the less direct systems, and with the conversion efficiency from the pair exceeding that obtained by each unit. The heat by-product from fragment converters may be used also in processing operations which require high-temperature heat sources.

The alternative possibility of operating at low temperatures without compromising conversion performances promises electric power plants of unusually high longevity, since the useful lifetimes of converters, in general, are extended when operating temperatures are lowered to values more gentle to their materials. It was also noted that the production of electric power at unprecedented extremes of low temperature—near absolute zero—became possible for the first time because of the natural "marriage" of fission to the one-step converter. Therefore, the individually peculiar features of fission and one-step conversion permit an exploration of the unknown implications of controlled power production near that point where, theoretically, all molecular motion ceases.

Design studies undertaken by Rand late in 1956 indicated the possibility that complete fragment-electricity reactors of practical size and promising performance could be achieved, and led to the creation of the first small-scale fission-electric cell at the Argonne National Laboratory in 1958.

The purpose of the Argonne and subsequent experiments, employing small and subcritical cells, was to demonstrate the validity of the concept and to acquire the necessary technology for the development of full-scale and self-critical fission-electric reactors.

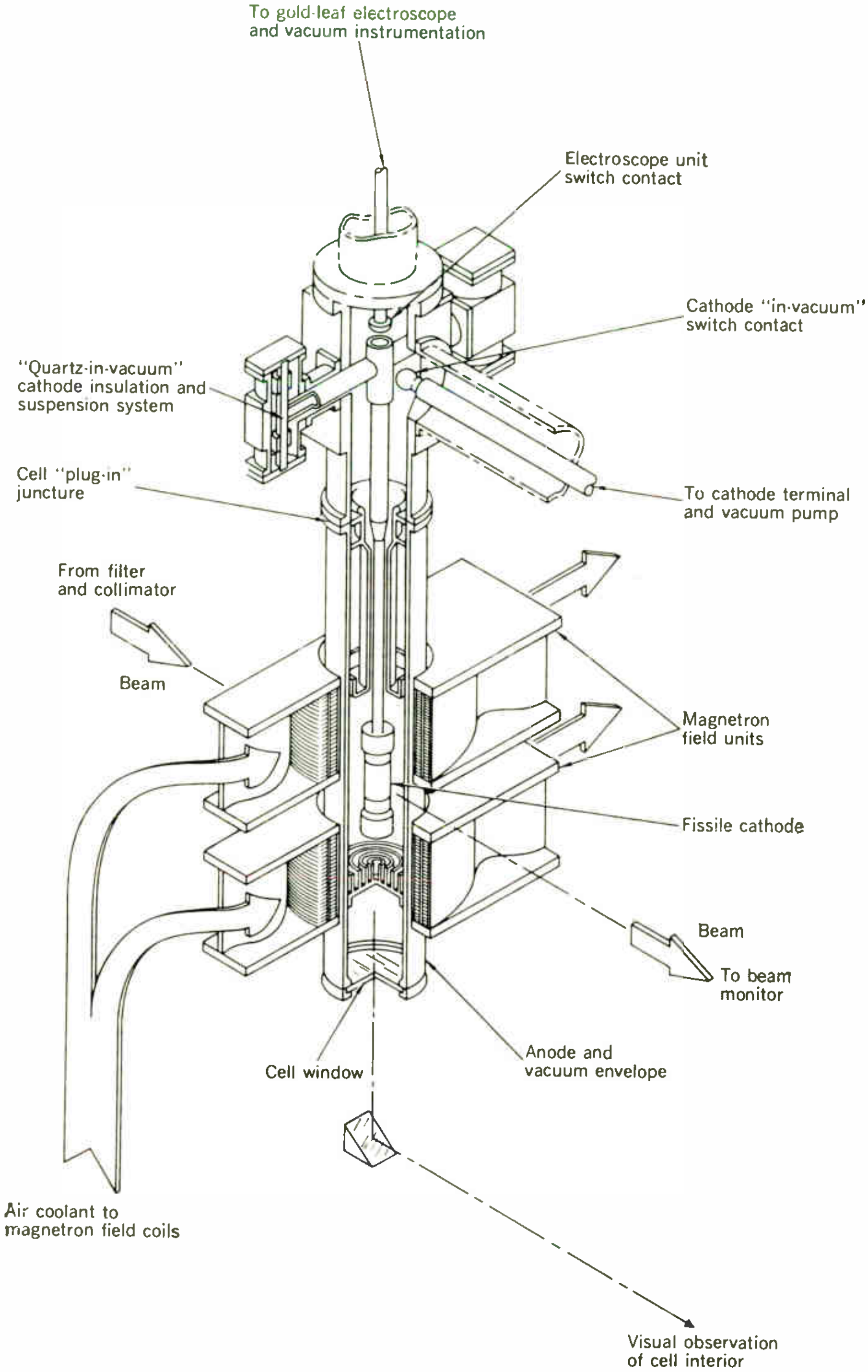
Essentially, a cell may be regarded as a representative sample of the large assembly of cells that would make up the core of a complete fission-electric reactor. This assembly would be placed in a configuration by use of the concept of the externally moderated reactor that was described in the proceedings of the 1958 UN International Conference on the Peaceful Uses of Atomic Energy.

The cells consist of two basic electrodes made of magnesium. A cigar-sized central electrode is coated with a small amount of uranium. This is centered in a quart-sized vacuum chamber which constitutes the second electrode. Fission occurs when the uranium on the central electrode is exposed to neutrons generated by a radiation source reactor. The electrically charged fragments yielded by fission speed from the central electrode and are intercepted by the vacuum-chamber electrode. In this manner, the two-electrode system is internally charged by fragment electricity which may be discharged into electric machines for the creation of useful work.

During the Argonne experiment of September 12, 1958, the fission-electric cells generated in token amounts the first electricity from fissile fuels produced by an unconventional process. The one-step conversion with fragments was verified at extremely low voltages.

Although potential voltages as high as 68 000 were indicated in these exploratory experiments, attendant problems in discharging the cell rendered these voltage figures less than definitive. The Argonne experiments, however, were sufficiently encouraging to justify additional research by use of a more carefully engineered test facility.

Early in 1959, equipment was moved to the General Electric Company's Atomic Power and Equipment Department in San Jose, California, where a small base laboratory was constructed to handle all operations not involving radioactivity. Operations at this laboratory supported irradiation experiments that were conducted in a second, controlled-environment laboratory especially constructed over the beam-port facility, within the containment vessel, of the GE test reactor at nearby Vallecitos.




Sketch of the basic cell and magnetron field system tested in 1962. The fissile cathode, made of pure magnesium, has its control zone covered with fissile material.

The cell program conducted at these GE sites was called the "Fission-Electric Cell Project." Irradiations employing nominal four-inch diameter cells of pure magnesium, and vacuum pumps of the oil diffusion type, yielded several important results and revealed the following conversion possibilities by mid-1962:

1. Routine proof indicated that one-step fragment conversion is valid with magnetic suppression or electrical suppression, or both, of fragment-produced secondary electricity.

2. Cell performance is most critically

linked to electrode surface conditions. Both current and voltage improve sharply when the cell surfaces of the magnesium cell and oil pump system are glow-cleaned. To generate voltages significantly above the 250-volt value achieved with the small-cell facility employed prior to mid-1962, advanced surface and vacuum techniques will be required.

3. In the systems tested to date, maximum performance is independent of vacuum for pressures of $10^{-5}$ mm Hg, or better.

4. Cell currents increase by several orders of magnitude when vacuums in the micron Hg range are attained. The possibility of a "hybrid" converter, with characteristically lower voltages and higher currents, is therefore recognized.

5. Voltages up to 18 000 volts have been generated by nonfragment currents—most likely by Compton electrons. The Compton electron generated, therefore, may be a possible source of high-voltage electricity.

6. A considerably greater understanding of the particles yielded at electrode surfaces is necessary for a more efficient development of the basic concept. The coarse mass-spectrograph features of previous cells have indicated the possibilities of high yields of secondary-emission ions, and the existence of unusual particles that may be new and "fundamental" rather than secondary.

As the result of these conclusions, the project group designed and constructed an ion-pump facility at the end of 1962. This cleaner vacuum pump services cells that are fabricated from an alloy of reactor-grade magnesium. This, in turn, is treated by more refined surface techniques. The ion-pump, plus the advanced-grade magnesium cell system is to be irradiated in the continuing program of the Fission-Electric Cell Project at present under prime contract with the AEC.

The continuous research and development effort was conducted for the USAF under the Project Rand contract until 1961, at which time sponsorship was assumed by the AEC. In 1963, the experiments were transferred from the AEC's contract with Rand to the Commission's contract with Dr. George Safonov, formerly with the Rand Corp., and at present with Advance Concepts Technology, Inc.

## Silicate and crystals of silver halide cause light to travel 'through a glass darkly'

Photochromism is the phenomenon involving the change of color of a substance on exposure to light. For many years scientists have known that ultraviolet light or other high-energy radiation can cause color changes in many materials, including glass. In general, the change takes a long time and usually the material cannot be reversed.

Corning Glass Works has developed a new series of inorganic materials that quickly change color on exposure to light and are completely reversible. They are silicate glasses containing dispersed submicroscopic crystals of silver halide. The crystals are precipitated in the manufacturing process during the cooling and reheating of the glass. The submicroscopic crystals darken under exposure to light and the glass changes color. When the light is removed, the color disappears. The darkening and clearing cycle can be repeated indefinitely. The crystals are a photochromic substance that remains permanently active.

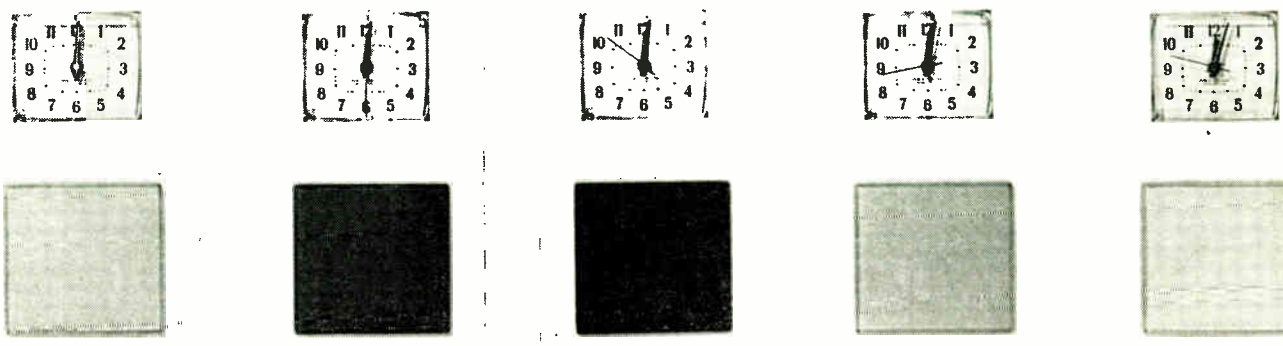
The wavelengths that produce darkening are typically the near ultraviolet. Dominant wave length and spectral range depend on the chemical composition of the glass. A typical glass darkens in daylight but stays clear under normal indoor lighting. The darkened color of the glass is usually gray, sometimes tending toward brown or purple.

In sunlight, the darkening generally approaches maximum intensity in seconds. The degree of darkness increases with intensity of light and the reaction to light can be almost instantaneous. A single one-millisecond flash of a photographer's electronic flash gun, having a capacity of 40 watt-seconds, decreased optical transmittance of a piece of glass to 25 per cent of the original value.

The time of return to the original transmittance ranges from minutes to hours, depending on composition of the glass and the heat.

The work on photochromic glasses is still in the laboratory, but expected future products include windows and sunglasses, optical memory and self-erasing display devices, and "light valves" in many types of new optical systems.

In this series, the clock indicates the time of complete darkness and clearing. Left to right: First, clear glass registers 86 per cent visible light transmission. Second, after 30 seconds' exposure to ultraviolet, it transmits only 50 per cent. Third, after 52 seconds, 28 per cent is transmitted. The light source is removed and the glass begins to clear. Fourth, after 52 seconds without the ultraviolet, the glass is clear. The full cycle takes less than three minutes. The darkening glass turns grayish. Thousands of photochromic glasses were tested to obtain the right darkening and clearing qualities



its associated digital electronics, the control system includes an acquisition sun sensor, a strap-down inertial reference consisting of gyros that integrate the rate of roll, pitch, and yaw; a Canopus star tracker; three orthogonally mounted inertia wheels; a cold gas reaction jet system; and control electronics.

The AOSO satellite is to study the sun for a full year, carrying 250 pounds of scientific instruments in a near-polar orbit 300 nautical miles above the earth. The orbit will allow all experiments on board to be constantly exposed to the sun for some nine months. During the final three months, the sun may be occulted (obstructed) by the earth during part of each orbit.

The fine sun sensor concept involves a series of optical wedges to refract sunlight and a closed-loop digital control system to position the wedges accurately. After refraction by the wedges, sunlight falls on a critical angle prism-detecting device that produces an error signal proportional to the angle of refraction. The error signal is used by the stabilization and control system to track the apparent motion of the sun.

The sun sensor is being developed by Honeywell (for Republic Aviation, which holds a NASA contract), at its Aeronautical Division facility in Boston, Mass.

## Vacuum deposition film production described

Development of a practical method for the commercial production of thin, superconducting niobium films has been announced.

The electron beam, vacuum deposition process that is used produces niobium layers at the rate of 1200 Å/s, and is readily applicable to continuous production operation. The use of a vacuum provides for precise control of layer thickness and an uncontaminated environment. Both of these characteristics are essential in microminiature electronics circuits and solid-state research.

Niobium is a hard, metallic element that becomes superconducting only at cryogenic temperatures. Niobium films are advantageous for microminiature computers, superconducting gyroscopes, and fundamental research. As an example, the new production capability will permit millions of superconducting films of only one millionth of an inch thickness to be used in a computer the size of a human brain. And these thin films would require only minute power inputs. Conventional computers performing comparable functions take up very large areas and need tremendous power inputs.

Niobium wires have been used in computers, but films have been found to be ten thousand times faster, yielding switching times on the order of one nanosecond.

In the new approach, niobium contained in a water-cooled copper crucible is bombarded by 20-kV electrons with power inputs up to 13 kW to attain the desired niobium vapor density. A pressure of about one torr is maintained to create a niobium vapor shield, which prevents residual gases from penetrating the vapor stream and contaminating the material.

The research and development work was by the National Research Corporation under a NASA contract from the California Institute of Technology Jet Propulsion Laboratory.

## Practical superconducting magnet of 100-kG is developed

A major advance in the creation of extremely intense magnetic fields, by the use of a superconducting magnet, or solenoid, said to be the first to achieve repeatedly a magnetic field of 100 000 gauss, has been announced.

This is about 200 000 times the average magnetic field strength of the earth, and about five times the field strength at which the iron core of a conventional electromagnet saturates.

The superstrength solenoid was made possible by a new superconducting material, HI-120, a niobium-titanium alloy.

The 100-kG magnet has been cycled sequentially from full strength to zero field and back again without the slightest damage to its structure. As part of the evaluation, the magnetic field was collapsed repeatedly, and the tremendous energy stored in the coil was released and dissipated. This process generated very high mechanical stresses that tended to rupture the coil. Such severe cycling, however, is a prime requirement of any practical magnet.

The new superconducting magnet is made in three sections, placed concentrically. In all, the magnet contains more than 20 miles of superconducting wire of a thickness comparable to sewing thread. The three-part construction is necessary because superconducting materials differ in their ability to remain superconducting under the intense magnetic fields they create. The middle and inside sections are wound from the new HI-120 alloy that is capable of carrying high-density electric currents in magnetic fields of more than 100 kG.

The outside layer experiences the lowest magnetic field and is the largest single section of the new magnet. It is wound from niobium-zirconium.

The new magnet is about seven inches in outside diameter, and six inches long. It has a one-eighth-inch hole through the center of the inside section where the 100-kG field exists. The coil is operated in a vessel of liquid helium at −452°F to maintain it in the superconducting state.

The new magnet was developed by Westinghouse research engineers, who believe that the device will be an important practical tool in the field of magnetic research.

### Supersensitive radar antenna developed

Scientists of the Air Force Office of Aerospace Research (OAR) have developed a radar antenna design which will be 100 times more sensitive than the world's largest radio telescope in Arecibo, Puerto Rico.

The multiplate antenna was designed by the OAR Air Force Cambridge Research Laboratories as a space surveillance and tracking radar system, and for use as an extremely powerful radio telescope. Theoretical work began in 1960, and construction of a test section, covering an area of 70 by 120 feet, was completed in 1962. Evaluation of the test section over a one-year period has afforded quantitative data that have verified the predicted performance characteristics of the full-sized antenna.

The projected multiplate antenna will consist of about 5000 flat metal plates, each 20 feet square, arranged in four elliptical areas around a 1000-foot-high tower. The tower will support a servo-controlled platform that holds the feeds for 25 beams. Each plate is adjustable in height and orientation to redirect energy from an arbitrary direction to the focus with the correct phase. A computer will be used to determine the proper tilt angle and center height of each plate.

The multiplate antenna concept resulted from the need for greater antenna gain and resolution. This can be achieved only at a specified frequency by a large collection area. For movable dishes, engineering and economic considerations

# two good reasons why you'll like Sparton's long-range job security

As an engineer . . . and as a family man, you'll appreciate not only the exceptional opportunity and challenge offered by Sparton Electronics, but also the security.

Sparton Electronics has long been a recognized pioneer in the electronics field and has a stable and varied production of military hardware that helps pay the rent for your personal job security. Engineering oriented programs specializing in high-reliability, high-precision electronics equipment will help you meet your present and future employment needs, and afford you the exciting engineering challenges you've been looking for.

**Right now, we have immediate openings for experienced and recently graduated engineers in oceanography, anti-submarine warfare, communications and navigation, and electro-mechanical devices. Steady, progressive employment that offers good opportunities for early responsibilities with a long established company within minutes of advanced educational and post-graduate facilities.**

Please submit your resume in confidence, stating salary requirements to—Personnel Director, Sparton Electronics, Division of Sparton Corporation, Jackson, Michigan.

**SPARTON ELECTRONICS**
**DIVISION/SPARTON CORPORATION • JACKSON, MICH.**
AN EQUAL OPPORTUNITY EMPLOYER

---

# Pre-Specification Selling

**in "Proceedings of the IEEE" serves the advertiser better and more effectively than in other publications**

Because "Proceedings of the IEEE" publishes basic research and engineering application articles, it is closely studied by those design engineers who must "keep ahead of the market." Often, the research published leads the industry production by from two to six years! The alert engineer, who needs this amount of time for his own design applications and for tooling up and changing the production in his own factory, studies the "Proceedings of the IEEE" carefully. Its articles keep him abreast with the changes that will inevitably come and make it possible for him to be ready for them. This is the time for you to sell the unknown engineer designing the never-dreamed-of-product "before it happens!" It is an insurance that you will be ahead of the market instead of behind it!

**The Institute of Electrical and Electronics Engineers**
Advertising Department

72 West 45th Street New York, N.Y. 10036

MUrray Hill 2-6606

---

*Explore Communications and Electronic Frontiers at*

## JANSKY & BAILEY

*rapidly expanding division of*
*ATLANTIC RESEARCH CORPORATION*

Jansky & Bailey, a trusted name in electronic communications for over 30 years, is expanding rapidly world-wide. We now offer several unusual and outstanding career opportunities to engineers who have the experience and creative ability to direct new programs and projects.

**ELECTRONIC DEVELOPMENT ENGINEER**

To participate in research and development programs in high atmosphere physics, meteorology and oceanography. Involves design, development, and test of electronic instruments and equipment applicable to these fields. Requires specific experience in field of physical measurements, electronic circuitry, electromechanical devices and transducers, and data-handling equipment. B.S., M.S. electrical engineering or physics with major in electronics or communications.

**RF ELECTRICAL ENGINEER**

To work in R&D performing theoretical and experimental work in shipboard high frequency communication systems. Work involves measurement of antenna impedances and field strengths. Requires knowledge of radio frequency measurements. B.S. engineering or physics and 2 years experience in RF circuit design as required in the design of broadband impedance matching networks and other RF circuits.

Send resume to Director,
Professional Employment, Dept. 457

*ATLANTIC RESEARCH CORPORATION*
Alexandria, Virginia A residential suburb of Washington, D. C.
An equal opportunity employer.

# Correspondence

**Invention of solid-state amplifier**—In an article in the February issue of *Physics Today*, Professor Virgil E. Bottom calls attention to several patents of Julius E. Lilienfeld (recently deceased) relating to devices for "amplifying and controlling electric currents." The remarkable thing about these patents is the dates—filed in 1926 and 1928 and issued in 1930–1933. Prof. Bottom incorrectly considers the devices to be n-p-n transistors. There is no doubt that they are field-effect transistors using n-type $Cu_2S$ as the semiconductor.

The device of 1 745 175 has a source and drain joined by a thin layer of copper converted to the sulfide, all upon a glass substrate. Midway between source and drain is placed a gate of aluminum foil, edgewise in a crack in the glass, so that the gate is flush with the glass. This gate forms a reverse-biased rectifier with respect to the $Cu_2S$. The device is operated in the enhancement mode. It is not likely that minority carriers play any part in the operation since the material is polycrystalline. Electron mobility is probably between 10 and 100 cm² volt-seconds. CdS as presently prepared for thin-film transistors has a mobility of 100 to 200.

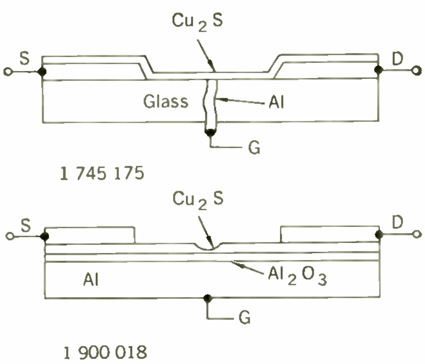



Fig. 1. Field-effect (not p-n-p) transistors by Lilienfeld. These transistors were patented in the early 1930s.

The second device, 1 900 018, describes an insulated gate field-effect transistor. The glass substrate above is replaced by an aluminum plate with 1000 Å of oxide over it. The source, drain, and $Cu_2S$ film are deposited upon the oxide. A thin region in the $Cu_2S$ is provided to enhance the pinch-off. The capacitance of the gate (aluminum plate) to source and drain must have limited operation to low frequencies.

A third patent, 1 877 140, describes a sandwich of two back-to-back $Cu_2S$-Mg rectifiers with the thin (2000 Å) Mg layer common to both junctions serving as the control electrode. One cannot accept the minority carrier hypothesis here either. Transport is most likely via grain boundaries in the manner of some grid type MIAs. This device would seem to be much more difficult to fabricate in a reproducible way than the first two devices.

Lilienfeld seems to have understood the effects of source-drain spacing and film thickness. Specific reference is made to modifying conductivity by electrostatic forces.

*Robert W. Hull*
*General Instrument Corp.*
*Newark, N.J.*

**More on engineer and scientist.** On page 172 of the February issue, Mr. Griffin and Mr. Warren, in clarifying the distinction between scientists and engineers, note that "Those who take the results of this research and apply it, at a professional level, to produce a desired result are engineers. The engineer, therefore, requires sufficient scientific training to enable him to understand the work of the scientist."

I submit that this description of an engineer, while necessary, is not sufficient, and that the emphasis on an insufficient criterion is responsible for the problem itself.

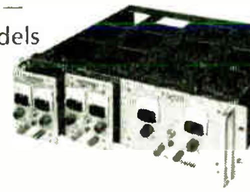
To arrive at a necessary and sufficient criterion, define an explorer as a person who seeks to extend knowledge; a scientist as one who seeks to increase understanding; and an engineer as one who seeks to apply knowledge and understanding. In using these definitions it is important to make a clear distinction between knowledge and understanding; perusal of an adequate dictionary is recommended.

It seems clear that an engineer would be remiss if he did not utilize the understanding gained by the scientist—a

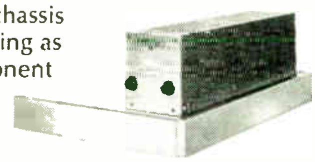
# NEW LAMBDA MODULAR-SUBRACK

## All Silicon, 0-20 and 0-40 VDC, Power Supplies Offer Highest Performance AT LOWEST PRICES

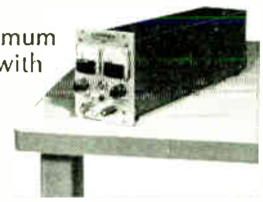
RACK MOUNTED—
¼ and ½ rack models can be mounted in unique Lambda rack adapters



CHASSIS MOUNTED—
Flush chassis mounting as component supply



BENCH USE—
Compact, minimum bench width, with protective feet



TEMPERATURE COEFFICIENT .015%/°C

LH SERIES
¼ AND ½ RACK MODELS
TO SERVE ALL YOUR NEEDS

REMOTELY
PROGRAMMABLE AND
CONTINUOUSLY VARIABLE

PERFORMANCE-
GUARANTEED FOR 5 YEARS

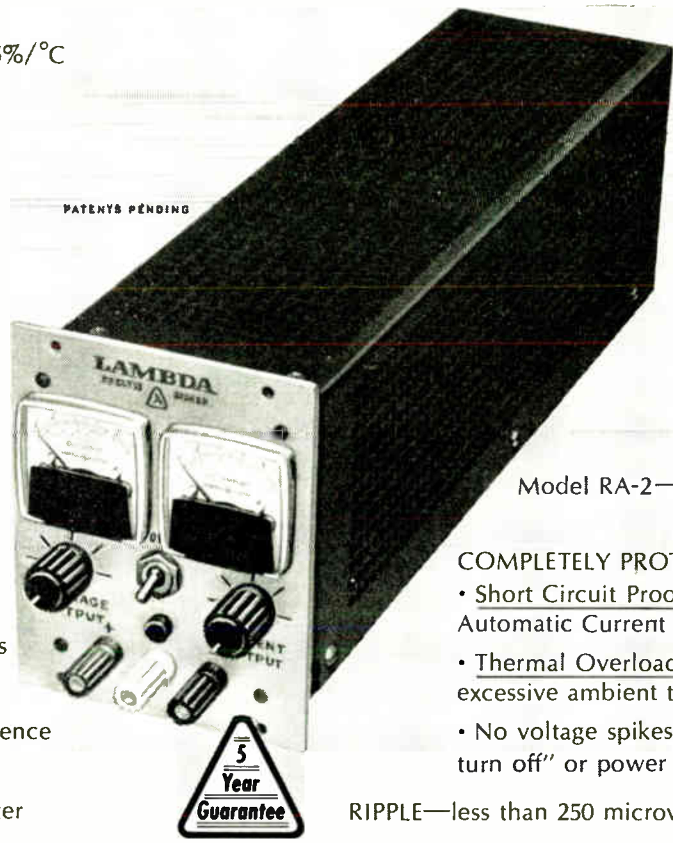
MEETS RFI SPECIFICATIONS-
MIL-I-26600, Class 3

WIDE VOLTAGE
AND FREQUENCY RANGE—
105-135 VAC, 45-480 cps

CONVECTION-COOLED—
no blowers, no external heat sinks

SERIES/PARALLEL OPERATION
and ± output to a common reference

REGULATION—Line or Load—
.015% or 1 MV whichever is greater




RACK ADAPTERS

Model RA-1—for ruggedized mounting and use with chassis slides. Features front panel insertion
Price $60.00

Model RA-2—for conventional rack mounting. Price $25.00

COMPLETELY PROTECTED

- Short Circuit Proof—Continuously Adjustable Automatic Current Limiting
- Thermal Overload Protection against excessive ambient temperatures
- No voltage spikes due to "turn on, turn off" or power failure.

RIPPLE—less than 250 microvolts rms and 1 millivolt P-P

### CONDENSED TENTATIVE DATA

**DC OUTPUT—VOLTAGE REGULATED FOR LINE AND LOAD**

| MODEL | VOLTAGE RANGE | CURRENT RANGE AT AMBIENT OF:(1) 30°C | 50°C | 60°C | 71°C | SIZE H W D | PRICES CHASSIS (2) MOUNTING | NON-METERED | METERED (3) |
|---|---|---|---|---|---|---|---|---|---|
| LH 121 | 0-20V | 0-2.5A | 0-2.0A | 0-1.7A | 0-1.5A | 5-3/16" x 4-3/16" x 15-5/16" | $154 | $159 | $184 |
| LH 122 | 0-20V | 0-4.5A | 0-3.5A | 0-3.0A | 0-2.5A | 5-3/16" x 8-3/8" x 15-7/16" | $220 | $225 | $250 |
| LH 124 | 0-40V | 0-1.25A | 0-1.1A | 0-1.0A | 0-0.75A | 5-3/16" x 4-3/16" x 15-5/16" | $149 | $154 | $179 |
| LH 125 | 0-40V | 0-2.5A | 0-2.1A | 0-1.8A | 0-1.5A | 5-3/16" x 8-3/8" x 15-7/16" | $229 | $234 | $259 |

(1) Current rating applies over entire voltage range

(2) Non-metered models with flush panel (add suffix S to model number)
LH 121-S and 124-S: 4-5/16" × 3-13/16" × 15-5/16"
LH 122-S and 125-S: 4-5/16" × 8" × 15-7/16"

(3) Metered models with front panel controls (add suffix FM to model number)

SEND FOR COMPLETE LAMBDA CATALOG.

**LAMBDA ELECTRONICS CORP.**
515 BROAD HOLLOW ROAD • MELVILLE, L. I., NEW YORK • 516 MYRTLE 4-4200

SALES OFFICES AND REPRESENTATIVES CONVENIENTLY LOCATED IN MAJOR CITIES point adequately covered by Griffin and Warren. However, it seems equally clear (at least to the writer), that the engineer is also remiss if he does not utilize the knowledge discovered by explorers, *even though this knowledge is not understood, or is imperfectly understood.* This last point is one of the keys to professionalism in engineering; among other things, a professional engineer is "licensed" to exercise his best discretion in the use of areas where understanding is lacking. The most common exercise of this discretion is in the choice of "safety factor," introduced to compensate for "uncertainties" in design; of course the existence of such uncertainties implies incomplete understanding.

The engineer in scientific research is perhaps more often engaged in exploration, certainly the case when he experiments, and draws "empirical results." The scientist may also be exploring, but often by accident, as a by-product to effort on another line (serendipity).

Concerning the lack of formal courses on the use of knowledge, probably this is a proper field for industrial training and for experience. But this seems to imply that formal engineering education must be regarded as preparatory, and that engineering requires an "internship." It might be well to inform the students of this fact, and of its consequences. (Similar thoughts also apply to engineering publications.)

The engineer may be closer to the natural philosophers of yesterday than to the scientist of today. The philosophers were concerned with both knowledge and understanding, with both exploring and explaining—a duality that seems to be rare today. Perhaps we can place this matter of science and engineering in better perspective by remembering these differences.

*R. P. Haviland*
*General Electric Co.*
*Philadelphia, Pa.*

**What's in a name?** The "Engineer or Scientist?" question, like most important questions is complex and involves variables. Every engineer has his (or her) own definition of what an engineer is, or should be. The difficulty arises in ascertaining commonalities without generalizing to the extent that a solution lacks utility. Even if a common ground among engineers were attained, there would exist the problem of changing the


# *The tape with the built-in duster!*

1000 times greater conductivity than ordinary tapes! That's how "SCOTCH" BRAND Heavy Duty Tapes drain off static charges before they can attract dust. That's the built-in duster that flicks away the growing danger of dust-caused dropout errors . . . a danger greater than ever as higher and higher recorder speeds and tape tensions generate more and more static.

Electrical resistance of the oxide coating of "SCOTCH" Heavy Duty tapes is 50 megohms per square or less. The resulting conductivity, unusual in magnetic tape, not only avoids dust contamination, it minimizes such other static problems as tape drag and skew, noise induced by arcing.


Heavy duty formulation of binder and high potency oxides withstands temperatures from −40 to as high as +250°F to conquer high head heat, assures tapes that outlast standard tapes by at least 15 times. Exclusive Silicone lubrication reduces recorder head and tape wear. And "SCOTCH" Heavy Duty Tapes are offered for all high-speed applications, even for extreme high frequency and short wavelength requirements. 16 different constructions include a variety of backing and coating thicknesses.

*TECHNICAL TALK Bulletin No. 4* provides detailed discussion of the effects of static electricity on instrumentation recording, offers helpful information in solving static-caused problems. It's free. Write 3M Magnetic Products Division, Dept. MEC-44, St. Paul 19, Minn.



**Magnetic Products Division 3M COMPANY**

nature of the general public's stereotype of an engineer, and this stereotype is only now changing from that of a man squinting at a transit.

The differentiation between "engineer" and "scientist" seems to be a spurious one. The term "engineer" itself lacks clarity and, disregarding "nonprofessionals" who identify themselves as engineers, a vast meaning difference still exists. Graduate engineers become administrators, technical recruiters, marketing specialists, manufacturers' representatives, etc., and still maintain after several years in their specialties that they are "engineers."

Obviously, engineers must agree on definition standards before an effective approach can be made to the public. However, adaptation of standards implies a mature and stabilized environment, which is clearly impossible where the art advances so rapidly. For instance, 10 to 20 years ago an electronics engineer who specialized in microwaves could be fairly identified as microwave engineer, but today he might be a specialist in solid-state physics, electron or ion beam dynamics, signal sources, millimeter-wave components, plasmas, coherent light, systems design, or packaging.

However, the variance in definition does not account for the tendency of professional publications to cater to reports of engineers engaged in research. I submit that these are the engineers, by the very nature of their work, who dominate in discovering new techniques, approaches, or ideas of general interest to engineering. And certainly, the commonweal organizations that sponsor the predominant amount of research are actively interested in having significant findings disseminated to a wide audience.

But what's in a name? Why call someone "engineer" if he wants to be called "scientist," or vice versa? Our culture has made both terms respected titles that many people aspire to. Perhaps if eminent status were to be given to "janitor," we wouldn't have the problem of restricting those who can rightfully call themselves "engineers"—but then maybe janitors would have similar problems with engineers who insisted upon being janitors!

*Woodrow W. Everett, Jr.*
*Electronics Engineer, RADC*
*Groton, N. Y.*

**Optimystically speaking:** Thanks to Fein and SPECTRUM for the pithy and

humorous article on artificial intelligentsia. Despite Mr. Bubble Piercer and the Pessimyths, the author appears to be rather safely in the middle of the road between utter radical and stodgy conservative.

The à la carte menu which he gives us lists all kinds of "processes." Perhaps he would like a proposal from Consensus Associates on the linking and organizing of the separate processes, or "units." At this stage of research we are perhaps tending to "overperfect" separate processes. The skeleton and interconnective parts, and their combined work are important. We have become too used to single-path signal flow through separate stages.

The stand against premature development contracts is defensible. On the other hand, this should not discourage research *experimentation* and its funding, provided that this is not aimed at a useless box of superb crackle-gray finish. The feedback and stimulus of a reasonably tangible production, toward the researcher, must not be overlooked.* We cannot assume that, just because we have digital computers and analog simulators designed for rather special fields of work, we are therefore well equipped with material. Form follows function, and we are trying for new functions.

*Otis N. Minot*
*Lexington, Mass.*

* Consider the effect of Mark I as compared with all the writings that preceded it.

Whynot Informatic Devices
Optimystica, USA

Dear Mr. Whynot:

Please accept our admiring congratulations. The idiom of your splendid expression in evaluating J. R. Bubble Piercer's evaluation of the proposals...is in the best tradition of Optimystica.

*Optimystically yours,*
*Homer O. Stasis*
*Optimystica International*

P.S. I personally appreciate your deep concern for the financial well-being of Optimystica USA members. I refer to the clever inclusion of the last phrase in the clause "this should not discourage research *experimentation* (your italics) and its *funding*" (my italics). You certainly keep your eye on the ball!