VOLUME 5 ☐ NUMBER 5 **MAY 1968**

IEEE spectrum

features

Departments: *please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

the cover

The theory of sets is becoming more prominent and vital in the ever-evolving world of computer technology; its importance is emphasized by the set-oriented "new math" now used throughout the U.S. educational system. The article beginning on page 47 describes some of the progress, problems, and paradoxes involved in this expanding field.

Ever since Marconi's name was given to the base-fed, grounded, vertical wire antenna of 1896, or Lodge's name was given to the center-fed, non-grounded or "counterpoise" version of 1897, inventors' names have been applied to their antenna creations. It is often difficult to remember an antenna configuration just from the inventor's name, so a list is given below of such antennas as are often identified by names of persons. Antennas not included are those derived from optics, as Cassegrain, Gregory, Lloyd, Luneburg, Mangin, Newton, etc., nor those having names related to the antenna excitation or radiation pattern, as Bessel, Dolph, Chebyshev, Hamming, Taylor, etc. The following list, compiled by Jack Ramsay, consultant in AIL's Department of Radar Techniques and Advanced Development, is a continuation of "The Variety of Antennas," which appeared in last month's issue. The antennas listed have evolved from electrical and electronics engineering rather than from optics or mathematical physics. In most cases a rough sketch of the antenna configuration can be made from the descriptions given; if not, reference to the textbooks will provide the information.

Proper-Name Antennas

Adcock: An antiphased pair of vertical receiving dipoles (or monopoles) for direction finding. Crossed pairs are commonly used with a goniometer.

Alexanderson: Horizontal wire VLF antenna connected to ground at equispaced multiple points by vertical wires with base-loading coils, the transmitter being coupled to an end coil.

Alford: Resonant square-loop antenna formed by four linear sides with their ends bent inwards to provide capacitive loading which equalizes the current round the loop.

Bellini-Tosi: Direction finding antenna consisting of two vertical orthogonal triangular loops with their bases over ground, used with a goniometer.

Beverage: Low height, long-wire receiving antenna over ground, having the receiver at one end, and a terminating resistor at the other end, which is nearer the remote transmitter.

Bruce: Original name of "Rhombic" end fire antenna, which consists of a diamond shaped arrangement of four wires with the feed line at one vertex and a resistive termination at the opposite vertex of the longer diagonal of the rhombus. The first published U.S. description was given by Bruce in 1931. Bruce's name is also given to a series fed array of vertically polarized, resonant rectangular loops (or half-loops over ground) each one-quarter-wave wide and one-half-wave spaced, the array having a castellated appearance.

Butler: Array antenna with feed system incorporating hybrid junctions to obtain a plurality of independent beams.

Chireix or Chireix-Mesny: Resonant series fed array of square loops of half-wave sides, feeding each other in cascade, corner to corner, and looking like a double zigzag.

Christiansen: Radiotelescope consisting of two interferometer arrays at right angles, and resembling a *Mills* cross antenna.

Covington and Broten: Compound interferometer consisting of a long line source adjacent to a two element interferometer of comparable aperture, in the same straight line.

Cutler: Rear feed for paraboloid reflector antenna, consisting of support waveguide with terminating cavity containing two resonant slots facing reflector, one on either side of support waveguide, each slot being parallel to the broad faces of the feed guide.

Foster: A mechanically variable parallel-plate prism attached to the aperture of a line source antenna to produce beam scanning. The path length prism is embodied in fractional regions of the space between two slightly separated coaxial cones, the outer cone being stationary and the inner rotating. The delimitation of the conical airspace to prismatic sections is achieved by providing non colliding short-circuits along generators of both cones, either by interleaving gratings or by choke grooves. The line source may be either outside the stator, or inside the rotor, which is usually divided and provides part of the total parallel-plate region.

Franklin: Base-fed vertical wire antenna several wavelengths high, and radiating broadside by the elimination of phase reversals by means of loading coils or wire folds.

Hertz: Elementary linear dipole radiator, with or without spherical or flat-plate ends (1887). Also loop antenna.

Kooman: Vertical array of horizontal full-wave dipoles fed by transposed two-wire line, and backed by parasitic reflecting curtain or horizontal dipoles.

Lewis: Microwave scanning antenna consisting of a (lensed) flat horn which tapers, not to a waveguide, but to a narrow rectangular opening across which a waveguide feed can be moved to scan the beam. The horn is folded by the incorporation of a 45-degree reflecting strip, and the thin rectangular end is formed into a circular annulus, around which the feed can rotate, the resulting deformed parallel-plate region taking a conical shape with the feed-circle as base.

Lodge: Counterpoise antenna, consisting of vertical dipole having horizontal top and bottom plates or screens, the lower being spaced from ground. Other versions were the "bow-tie" antenna, where two narrow triangular plates were connected at their thin ends by a coil, and the "umbrella" antenna in which the end plates were made conical and constituted by wires (1897). Lodge also originated the waveguide open-end radiator and microwave lenses in 1894, before the spark microwave technology disappeared and was replaced by the longer wave techniques introduced by Marconi in 1896.

Luneburg: Lens antenna consisting of a spherical lens illuminated by a feed at or near the end of any diameter, a beam emerging from the opposite side, and positionable at any direction in space by moving the feed around the surface. The lens is given a spherically symmetric, non uniform distribution of dielectric constant, obeying the law $2 - r^2$, and internally bends the divergent rays from the feed around curved paths so that they come out parallel.

Marconi: Original name of vertical wire antenna, base fed with a ground connection (1896).

Mills: Two independent fan-beam antennas at right angles in cross formation, with common center and common pencil-beam volume of low gain. Antennas are combined by switching from phase addition of outputs to phase opposition at constant rate to secure the angular resolution of the pencil-beam component (a technique suggested by M. Ryle).

Robinson: Microwave scanning antenna consisting of an astigmatic reflector having as feed system a parallel-plate region fed by a waveguide. The parallel-plate region is bent and configured so that the feed waveguide end is circular, to allow rotation of the feed guide, and is approximately in the same plane as the output end, or larger aperture of the parallel-plate region. (Cf. *Lewis*)

Schmidt: Microwave scanning antenna based on optical Schmidt camera. Spherical reflector has aspheric microwave lens at centre of curvature, with scanner or multifeed approximately half-way between these elements.

Schwarzschild: Microwave scanning antenna analogue of S. astronomical telescope, embodying two reflecting surfaces profiled to minimise aberrations, and illuminated in cascade by a scanner or multifeed.

Sterba: Series-fed array of adjacent, broadside-firing, transposed square loops of half-wave sides and approximately half-wave spacing. (Cf. *Chireix*).

Van Atta: A retrodirective array of antenna elements so interconnected that an incident wave produces a radiated beam from the array in the direction of the incident wave reversed over a significant range of angles of incidence, as in the case of an optical autocollimator. Versions of the array may contain amplifiers, circulators and mixers, may offset the frequency, and may direct the transmitted beam, with or without modifications, in a direction or directions other than that of the incident wave reversed.

Walmsley: An array of vertical rectangular loops of one wavelength height and half-wave spacing arranged in parallel planes, and series fed at the mid-points of the longer, vertical sides, with appropriate transpositions of the feed lines.

Windom: Horizontal half wave dipole above ground, fed by a vertical or near-vertical single wire connected at about one-twelfth wavelength from the center of the dipole.

Yagi or Yagi-Uda: End-fire antenna, consisting of driven dipole with parasitic dipole reflector and one or more parasitic dipole directors. Driven dipole is usually a folded dipole, and all the antenna elements usually lie in a plane. The parasitic elements need not be coplanar but can be distributed on both sides of the plane of symmetry.

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

Spectral lines

Evidence of intelligent life on earth. Just as there can be no doubt that a major problem facing the United States today is the economic collapse of its large cities, there can be no doubt that a major cause of that collapse has been the vehicle propelled by an internal-combustion engine—the bus, the private auto, the truck. Only in the United States has the problem now attained the dimensions of a crisis, but the problem is nonnational; major cities of Europe and Japan are rapidly on their way to membership in the club.

Full of good intentions, and backed vigorously by the voters, governments have provided ever-bigger and ever-better roads into and through the cities. Traffic noise and air pollution have induced the middle class to move to the suburbs and commute, thus creating more traffic, more freeways through residential areas, more noise, more fumes, more parking lots, more hardtop to exacerbate the summer's heat—and further evacuation to the suburbs. Eventually the roads become so good that business and industries can move to the suburbs too. Railroads, their passenger facilities used less and taxed more, raise their prices and cut their service, more commuters take to the automobile or the bus, and the vicious cycle continues.

With intercity traffic, American railroads have an equally hard time. Operating economies must be worked out within the framework of work rules dating from 1920. We can only guess at the quality of the management talent that can be recruited to man this straitjacket, but an aggressive young man may possibly feel more drawn to the airlines, for which the year 1920 never achieved acceptance as a norm.

Few will deny that the present is unsatisfactory, but where do we go from here? No good prescription can be written without taking thought about how we got where we are now. Certainly we have been operating on a false footing. It seems clear, for one thing, that a major error has been made in the allocation of costs. As an example, each suburbanite has been free to decide whether his cost of commuting (including nonmonetary costs, such as inconvenience) is minimized by driving to work in his own car—on roads paid for, in part, by taxes on the railroad that they induce him to forsake. But his decision to drive imposes costs on the city; among others, costs associated with polluted air, depressed real-estate values, and schools with no middle-class pupils. In short, a decision about a complex system has been made, without the stabilizing supply–demand feedback of a market, by a host of individuals whose concern is with only one small element of the system—individuals, indeed, who have scarcely been conscious that the system is there.

It seems certain that the "systems approach," a technique closely allied with electrical engineering, will be brought to bear on the problems of transporting people and things. Engineers will have a role in the system planning and analysis, as well as in providing the elements of the system with power, controls, and communications. It is highly appropriate, therefore, that the Proceedings of the IEEE for April is a special issue on transportation. More than most of its predecessors, this special issue deals not merely with technical problems, but also with social and economic ones. It envisions intercity ground transport of passengers at 500 km/h (300 mi/h), personal vehicles that are computer-controlled when on main arteries, sustained emphasis on minimizing door-to-door travel time.

An integrated system to replace the present congeries of subsystems will demand radical change. That such change is not impossible is demonstrated by the current revolution in oceanborne shipping. Here systems analysis of an age-old activity has generated new kinds of ships, new kinds of containers, new kinds of railroad cars, new trucks, new port facilities. Mainly by reducing mismatch at the interfaces, the systems approach has effected reductions in cost so great as to justify the expense of the new installations and new capital equipment. The same will doubtless be true of a well-planned system for travel by people, though revising the present jumble will take decades, and may well entail outlays comparable to those of the current space program.

The transportation revolution is not just a gleam in the eyes of a few men who see visions. Though there is a very long way to go, people are already at work on some of the hardware. For what they are doing, and analyses of why it should be done, give some time to the April issue of the Proceedings. It is a remarkable document.

J. J. G. McCue

Authors

Electronics in transition (page 44)

Patrick E. Haggerty (F) received the B.S.E.E. degree from Marquette University in 1936. He joined the Badger Carton Co. in 1935 as a cooperative student engineer and following his graduation became production manager, and later assistant general manager. During World War II he was a lieutenant in the U.S. Naval Reserve, in the Bureau of Aeronautics, Washington, D.C. His duties were in connection with airborne electronic equipment production. He subsequently headed the Electronic Production Branch of the Electronics Component Group of the Bureau. In 1945 he became general manager of the Laboratory and Manufacturing Division of Geophysical Service Inc., which later became Texas Instruments Incorporated. Mr. Haggerty served Texas Instruments from 1950 to 1958 as executive vice president and director, and from 1958 to 1966 as president. Since the end of 1966 he has been chairman of the Board of Directors. He served as IRE President in 1962 and is a former member of the IEEE Board of Directors.

The third great crisis in mathematics (page 47)

C. K. Gordon, Jr. (M) is a senior engineer with the Missiles and Space Systems Division of the McDonnell Douglas Corp., Santa Monica, Calif. He is also on the faculty of the Mathematics Department of the Los Angeles City College Extended Day Division. Prior to joining McDonnell Douglas he worked in the area of mathematical models of organizations in the Command Systems Department of the Systems Development Corp. Previously he directed the activities of the Bionics Group at Litton Industries' Research and Analysis Department.

Mr. Gordon received the B.S. degree from the University of California, Los Angeles, in 1952 and the M.A. degree from the University of Southern California in 1957. He also did work toward his Ph.D. degree at the University of Michigan. His interests during the past few years have centered on cybernetics and the applications of mathematics to the biological and behavioral sciences. He is currently working on a textbook dealing with mathematical techniques in the "soft sciences."

Automation comes to the printing and publishing industry (page 53)

Gordon D. Friedlander, a staff writer for IEEE SPECTRUM, has also written articles for IEEE STUDENT JOURNAL, *Civil Engineering*, *Journal of the AIA*, and *The Washington Post*. He is also the author of a series of articles on engineering and architecture in the Sunday *St. Louis Post-Dispatch*. During his engineering practice, he was affiliated with Ebasco Services; Ford, Bacon & Davis, Inc.; and P. T. Mikluchin & Associates, Toronto, as a professional consultant and project engineer for the design of steam-electric and hydroelectric power plants, and industrial and commercial buildings.

Mr. Friedlander attended R.P.I., University of Michigan, and the Università di Firenze, and holds the B.S. and M.S. degrees in civil engineering. In 1967, he served as special consultant on the Committee on the Future of *The New York Times*. He has lectured at University of Wisconsin and Industrial Education Institute engineering seminars. His biographical sketch appears in the current issue of *Who's Who in the East*.

Making integrated electronics technology work (page 63)

Robert N. Noyce (F) was one of the founders of Fairchild Semiconductor, now a division of Fairchild Camera and Instrument Corp., where he established the Research Department and directed the initial development of the silicon mesa and planar transistor lines. Since 1965 he has served as group vice president of FCI, with responsibility for both the Semiconductor and Instrumentation Divisions.

He received the B.A. degree from Grinnell College in 1949, and the Ph.D. degree in physical electronics from M.I.T. in 1953. At the Research Division of Philco Corp. he led a solid-state physics group who worked on the development of germanium and silicon high-frequency transistors. In 1956 he left Philco to join Shockley Semiconductor Laboratory, where he directed the design and development of diffused silicon transistors. Dr. Noyce holds 15 patents in the semiconductor field and is a member of Phi Beta Kappa and the American Physical Society.

Electrochemical vehicle power plants (page 71)

D. A. J. Swinkels is a supervisory research chemist with the Research Laboratories of the General Motors Corporation, where he is in charge of basic and applied electrochemical studies. He was born in Rotterdam, Netherlands, in 1935. In 1952 he migrated to Newcastle, Australia, where he worked while pursuing his education at the University of New South Wales, from which he received the B.Sc. degree in applied chemistry in 1960. He then joined the Electrochemistry Group at the University of Pennsylvania, from which he received the Ph.D. degree in 1963. He subsequently joined the Allison Division of General Motors. Early in 1964 he recommended that the $Li\text{-}Cl_2$ system be investigated for use in a high-energy-density high-power-density energy-storage device; most of his work since then has been associated with this system. Dr. Swinkels has had some ten technical papers published in various journals and is an active member of the Electrochemical Society.

Electronics and the urban crisis (page 78)

Seymour N. Siegel, director of radio communications for the City of New York, has long been prominent in national and international broadcasting circles. A member of the staff of the Municipal Broadcasting System since 1934, he was appointed its director in 1947 by Mayor William O'Dwyer. He has been reappointed five times by succeeding mayors. He has served as the City's Deputy Director of Civil Defense Communications and at present is chairman of the Mayor's Committee on Living Music. Mr. Siegel received the B.S. degree in economics from the University of Pennsylvania and the M.A. degree from Columbia University. During his many years in communications he has held numerous important posts. He is past president of the National Association of Educational Broadcasters, and was first president of the Broadcasting Foundation of America. He has received numerous national and international awards for his pioneering work in educational broadcasting.

Laser applications (page 82)

R. D. Haun, Jr. (M) received the B.S. degree in physics from the University of Kentucky in 1952 and the Ph.D. degree in physics from the Massachusetts Institute of Technology in 1957. While at M.I.T. he assisted J. R. Zacharias and J. G. Yates in the development of a cesium atomic beam frequency standard. He also carried out measurements of the effect of an electric field on hyperfine energy levels in Cs-133 atoms using atomic-beam magnetic-resonance measurement in order to measure frequency shifts of about one part in 10^{10}.

Dr. Haun joined Westinghouse Research Laboratories in 1957. His early work involved research on microwave ferromagnetic resonance, parametric amplification, and tunnel-diode circuits. Since 1962 he has directed research on lasers, luminescent materials, gas-discharge light emission, optical design, holography, and optical information processing.

Electronics in transition

Patrick E. Haggerty
Texas Instruments Incorporated

This is one of those times in the life of a man when, having been honored by his peers for some of his labors, he must, while accepting the honors with appreciation and humility, look at himself and his work against a background of the time in which he lives and the problems of the society of which he is a part. Mixed feelings of pride and humility are especially strong for me now, because among so many worthy men being honored by the Institute tonight, two of my associates, Roger Webster and James Biard, are among the newly elected Fellows of the Institute—and especially because, by one of those coincidences of life, you have chosen to bestow that senior award of the Institute, the Medal of Honor, upon another respected friend and associate, Gordon Teal.

These four threads that you have tied together tonight extend back over two decades to Gordon's inspired work with single crystal germanium and silicon. This work—together with the contribution of Morton, Brattain, Bardeen, and Schockley and many more of his then associates at the Bell Telephone Laboratories—must be evaluated as the discovery of a new technological world, the exploration and development of which has brought the four of us and a large proportion of all of you here tonight.

By their very nature, awards such as these must often represent culminations of one kind or another; and among the emotions and satisfactions the Founders Award has stirred in me, particularly in juxtaposition with the award of the Medal of Honor to Gordon Teal, a nagging question insists on intruding itself: Are these awards exclamation marks at the end of middle paragraphs in the composition of our life's work, or are they periods following the summary sentences?

We are engineers and these are engineering awards, and engineering has to do with the creation of useful products and services. Thus, any judgment of where these threads you have pulled together tonight fit into the total future scheme of our profession must begin with an evaluation of the pertinence of that profession to the needs and problems of today.

Exciting times, but troubled ones!

Myriads of new nations are beginning to find that independence is but a first step, and that the stability and affluence they seek can be found only at the end of a long and arduous road demanding ability, dedication, diligence, and responsibility from both government and citizen.

International trade is disturbed, and there is a widespread gold crisis. Britain devalues the pound and initiates an austerity program. France pursues an at-times agonizing nationalism. The United States faces a balance of payments problem and now, after decades as advocates of an increasing free trade, the administration proposes travel bans and overseas investment limitations; and a distressing number of U.S. senators and representatives are recommending quotas and limitations on that trade.

It took an estimated eight million years from man's beginnings to reach a world population of 300 million at the time of Christ. It has taken less than two thousand years to increase that 300 million tenfold, and one billion of the three billion have been added in less than 40 years. On the basis of present trends we'll have added another billion by 1975 and still another billion for a total of five billion by the early 1980s.

So the world divides into the "have" and "have-not" nations. The "have" nations pay themselves well and enjoy high standards of living. The "have-not" nations have low incomes and poor standards of living.

The United States is a great nation and for most of us Americans the benefits of citizenship are legion. Belatedly, sometimes inadequately, but still genuinely, we are trying to remove the barriers imposed by prejudice so that all of our citizens may have a full share of those benefits. Yet riots have torn our cities; and even without riots we are afraid to walk on our streets after dark. Lake Erie is becoming a Dead Sea; Los Angeles in sunny California rarely sees the sun.

This disturbing look into the world's Pandora's Box leads back to the statement I made earlier concerning the future scheme of our profession. The question is whether

On March 20, at the International Convention, P. E. Haggerty was honored with the IEEE Founders Award. His acceptance speech, presented here, stresses the vital responsibilities imposed on electronic technology by a rapidly changing world

the talents of the electronics industry and, more particularly, the skills of the scientists and engineers who have made it what it is, are pertinent to the solutions of this catalog of problems:

1. The problems of war and national defense.
2. The problem of the restoration and conservation of our environment.
3. The problems of our cities.
4. For the "have-not" nations, the problem of transformation into industrial societies.
5. For the "have" nations, the problem of sustaining and improving their already high standards of living in the face of the need of the "have-not" nations to industrialize and the relatively much lower wages and salaries they pay themselves while in the process.

Where does technology fit in?

Let's see how electronics might be pertinent to problem solutions in some of these areas.

So far as the problems of war and national defense are concerned, the needs are so obvious and the contributions made by electronics so vital that little elaboration is necessary except to observe perhaps that every new military need—from strategic missiles to antiballistic missile systems, from helicopters to supersonic aircraft, from reconnaissance to attack—demands more and better electronics.

A principal problem for any heavily populated, reasonably complex society is the orderly development, redistribution, reclamation, and use of its water resources. The hardware and software of electronics allow us to develop atmospheric models, watershed models, sediment models, surface-water quality and quantity models—in fact, models of all of the necessary major components—to simulate regional water resources. And the hardware and software of electronics can further provide the sensors, the memories, the information processes, the data displays, and the control mechanisms necessary for the orderly development, redistribution, and use of the water resources. But if these tools are to be applied, somewhere there must be a governmental unit with the comprehension, the imagination, the authority, and the funds to use them.

Let's take, for example, the problem of traffic in any major city. It would now be completely feasible technically to calculate for every artery the conditions for maximum rate of flow of traffic, and then to identify every vehicle entering that flow of traffic and direct it into arteries not yet at maximum flow and forbid it entry when its addition would simply slow the total flow. There is a saturation point for every system of traffic arteries, and electronics can insure that every traffic system carry the maximum amount of traffic it is capable of carrying under the specified circumstances. Electronics cannot make the artery carry more than that specified maximum. But it can help detect and correct breakdowns and make clear what further steps must be taken to increase that maximum flow.

To switch to another problem area, the United States is a country with an exceptionally high standard of living. We pay ourselves more for our hours of work than does any other country in the world. The United States must sell to and compete with the rest of the industrialized world. Simultaneously, it must sell to and buy from the underdeveloped world while assisting the underdeveloped nations in their own strivings toward industrialization. If the United States is to sustain its high standard of living, then its overall industrial effectiveness must be sufficiently higher than that of any other nation so that when the high wages we pay ourselves for our hours of labor are multiplied by our relative effectiveness, we still come out competitively high.

Those of us who are engaged in developing, manufacturing, and selling products and services around the world know that in many respects this has become increasingly difficult over the past decade. For example, even though the rate of increase in hourly earnings of manufacturing employees in Japan has been increasing faster than it has in the United States, the absolute difference has increased from about $1.80 per hour in 1957 to nearly $2.15 per hour in 1966—or by comparison with Italy, from a $1.72-per-hour difference in 1957 to $2.08 in 1966. To make a comparison with another Far Eastern nation, striving with considerable success to move into the modern industrial world, in 1957 Taiwan's manufacturing employee received on an average $1.98 per hour less than a similar worker in the United States; in 1966 that difference had increased to nearly $2.60. Although these illustrations are all with respect to hourly earnings of manufacturing employees, the relative differences are only moderately less if the comparison is made for professionals and managers.

Obviously, if the United States is to compete with wage and salary differences such as these, the overall effectiveness of its total industrial machine must be extraordinarily high. To the extent that we cannot meet the challenge, our standard of living must also respond by either slowing down the rate of improvement or actually dropping. Thus far the United States has, by and large, sustained the upward trend of its overall standard of living by a combination of advanced technology in its products and services and a superior way of organizing to produce those products and services. Both the technology and our way of organizing to produce our goods and services—that is, the way we manage—have been diffused throughout the world since the end of World War II.

Japan is an excellent example of rapid growth into an efficient industrial nation while still paying itself at wage and salary rates appreciably below ours. Inevitably, we are unable to compete in certain areas. The problem we face with respect to the other industrialized nations, they too face in varying degrees with respect to one another, and all of the Western industrial nations face with respect to Japan, but the whole spectrum of industrial development from the United States through Western Europe and Japan and down through the underdeveloped world is such that inevitably every industrialized nation, including Japan, already does and will continue to face similar competition from other nations struggling to evolve an industrial society.

Thus, it seems there must be an inexorable pressure, not just in the United States but, on a somewhat descending scale, throughout the entire industrialized world for a constantly improving effectiveness in every segment of each nation's industrial machine. This pressure for improvement in effectiveness must change fundamentally the way in which we design, manufacture, and market our products; but what may be even more significant, it will change the way we organize and manage our businesses.

It is difficult to draw any other conclusion except that the sensors, the logic, the memory, the displays, the controllers, and their associated software are the principal and essential tools that must be applied to increase the effectiveness of our own industrial machine sufficiently to compensate for our very high wage and salary costs. Over a relatively short span of time the same problem, for the same reasons, will exist in every industrialized nation.

These are just a few examples, but the solutions to all of these problems have major elements in common. In all of them we need to convert information into usable signals. We need to separate these useful signals from the inevitable noise that accompanies them. We need to store, process, and interpret the raw data accumulated. We need to store and display the processed and interpreted data. We need to cause action in response to the information and create feedback signals to correct the error in the actions taken; and much of the action must be in real time. Thus, there is a startling coincidence between these needs and what much of electronics is all about today. Thirty years ago we could have made only the most rudimentary kind of contribution to the solution of these problems. And it is really only in the last ten years, with the advent of solid-state components and sophisticated data-processing hardware and software, that we can truthfully say we can tackle these difficult problems with tools whose sophistication is commensurate with the difficulty of the problems. It is probably also accurate that almost every tool that is necessary is at least in sight or will be in the 1970s.

Of course, what is really common about the worldwide potential contributions of electronics is, as President Herwald has already emphasized, that in each case man is using electronics to extend his mind, his senses, and, to some degree, his spirit. The Industrial Revolution that began with the invention of the steam engine by Watt was a revolution that depended upon the extension of man's muscles. The even more significant revolution that is just beginning in these middle decades of the 20th century is an extension not just in capability of the mind and senses, but in space and time as well. With electronics, man can see, hear, and measure space and time with sensitivity, breadth of response, and precision far beyond that of his unaided senses. With the logic and the memory of electronics, man can expand his mind not just in size but in speed and apply that logic and memory at a distance. With television and radio, man can already see and be seen, and hear and be heard, with few limitations in space and time. With laser holography, in the relatively near future managers and professionals will stay in their offices and laboratories and still participate personally in distant conferences and meetings in three dimensions and in color. Since they can also manipulate and process information with electronic logic and memories at a distance, they may be more effective participants in such a meeting or conference than if they were truly present in person.

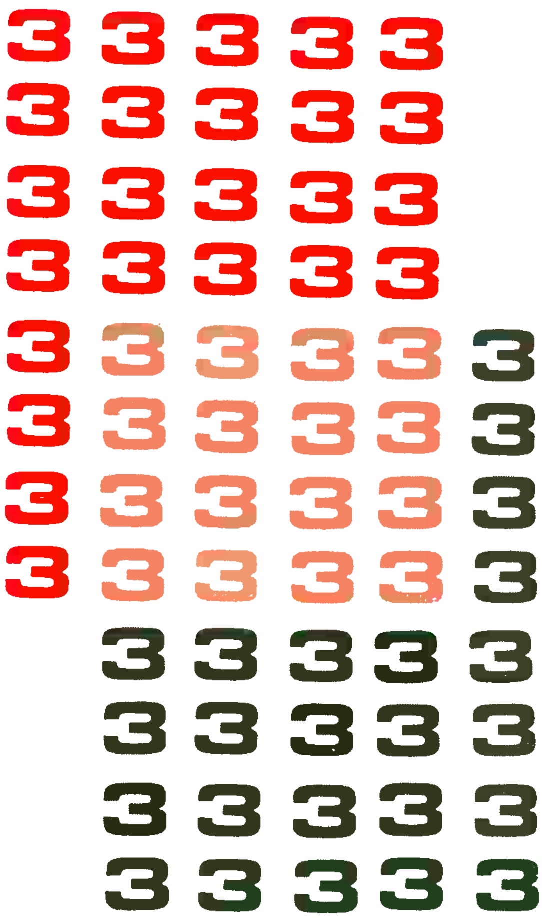
So whether these awards represent exclamation marks or periods, I have no doubt whatever that our profession has barely begun its work, and our industry is still at a relatively early stage in its growth. It may be true that the growth rates of electronics over the next five to ten years will be slowed while we develop and expand the tools on hand and in sight and while we evolve the procedures and institutions, both public and private, to apply them. After that and for as far into the future as any of us are likely to be interested, I believe that growth in quality, complexity, and size will continue at a rate well beyond that of the Gross National Product. The match between the capabilities the tools of electronics provide and the most pressing needs not just of the United States but of the entire industrialized world is too close for it to be otherwise.

As one engineer, I am sure of all this. But it is not enough. I hope that as man expands his sight, he will really see; and his hearing, he will really hear; and his touch, he will really feel; and that as his mind grows, so will his wisdom and his heart. But all this we must add as men, not as engineers.

The third great crisis in mathematics

Modern mathematics is experiencing many changes, in both approach and emphasis, to keep pace with modern technological progress. Prominent in the "new math" is increasing use of, and emphasis on, the theory of sets

C. K. Gordon, Jr.
McDonnell Douglas Corporation



Topology and measure theory, along with other branches of mathematics, are founded upon the theory of sets. The beginnings of the theory were steeped in paradoxes. One concerns the very concept of a set itself, such as proving that "0 is ordinary if and only if it is not." The concepts have been debated for more than a century and from them have evolved the axiom of choice, axiom of selection, infinite set theory, well-ordering theorem, and the Banach-Tarski paradox, among others.

Around the 5th century B.C. the discovery that all geometrical magnitudes of the same kind are not commensurable with one another precipitated the first crisis in mathematics. When it was shown that the diagonal and a side of a square contained no common unit of measure, the entire Pythagorean theory of proportion had to be abandoned as unsound. Though many attempts were made in ensuing years, this crisis was not resolved until around 310 B.C. by Eudoxus. His outstanding revision of the theory of proportion and magnitude and his brilliant treatment of incommensurables appear in the fifth, sixth, and tenth books of Euclid's *Elements*. Eudoxus' analysis of the irrational number (viz., numbers such as $2^{1/2}$ withstood the scrutiny of succeeding generations of mathematicians until the second half of the 19th century, when it received a new twist at the hands of Dedekind.

The second crisis occurred in the late 17th century around the newly created calculus, and contradictions and paradoxes appeared in increasing numbers until a crisis in the very foundations of the subject became evident. It wasn't until the early 19th century that Gauss and Cauchy took the first steps toward resolving this crisis by replacing the vague notion of infinitesimals with the precise methods of limits. Their work was followed with the "arithmetization" of analysis by Weierstrass, Bolzano, Dedekind, Cantor, and others.

The third major crisis in foundations is the one to which I want to devote this discussion, for it is still going on, and to me is the most exciting. The problem suddenly burst forth upon a startled mathematical world around the turn of this century; now, almost 70 years later, it is still not resolved to the satisfaction of all mathematicians.

This crisis was brought about by the discovery of paradoxes around the fringes of Cantor's general theory of sets. The seriousness of this is due not only to the fact that so much of modern mathematics is permeated with set-theoretic notions but that certain branches, e.g., topology and measure theory, rest upon set theory as a foundation.

The Set

According to Cantor, a *set* is "any collection into a whole, *M*, of definite and separate [i.e., distinguishable] objects *m* (which will be called 'elements' of *M*) of our intuition or thought." This seems innocent enough and, indeed, no one realized at the time that the concept of a *set*, so defined, was loaded.

Some examples of sets are the set of all musicians in a particular orchestra, e.g., the Boston Pops; the set of all people living today who are 50 years or older; the set of all people living today who are 500 years or older (a perfectly legitimate set, which just happens to be empty—it is a special set called the *empty* or *null set* and plays somewhat the same role in set theory as does zero in algebra); the set of all even whole numbers (a set containing infinitely many elements); the set of all even prime numbers (a set consisting of a single element, viz., 2); the set of all points belonging to some particular line segment (another infinite set, which, surprisingly, can be shown to contain more elements than the set of all even whole numbers). These examples all fit Cantor's notion of *set*, the last being ubiquitous in modern developments of the theory.

The first hint of trouble with this notion appeared in a paper published by Burali-Forti in 1897. This first paradox in set theory is too technical to be discussed here; it deals with Cantor's theory of transfinite ordinals and was already known to him in 1895. However, I can give a nontechnical account of a similar paradox discovered by Cantor around 1899. But first, it is necessary to consider his notion of a *transfinite number*.

Transfinite ordinal and cardinal numbers

Cantor's theory of transfinite ordinal and cardinal numbers is, I think, one of the most beautiful and original edifices of modern mathematics; it is an exciting adventure you may wish to embark upon by consulting the bibliography at the end of this article. For the present purpose, a couple of notions from this theory will suffice. The first notion concerns *cardinality*, which, roughly, represents the number of members belonging to the set. For example, the cardinality of the set of even primes is 1; the cardinality of the set of all possibilities when a coin is tossed twice in a row is 4 (viz., HH, HT, TH, TT). Whenever a set contains N members, where N is any positive integer (no matter how large) it is said to have a finite cardinal number and is called a finite set. Now, what about the set of all even positive integers? What is its cardinality? Clearly, since no positive integer N will suffice to characterize its cardinality, Cantor did a very clever thing. He invented a new number, a_0, to characterize the cardinality of this set. Actually, he used aleph, the first letter of the Hebrew alphabet, with a subscript zero—called *aleph null*—because he needed a brand-new symbol for this brand-new notion. Aleph null is the first *transfinite cardinal* (number). I say "first" because there are other, greater transfinite cardinals; for example, the cardinality of the set of all real numbers is greater than the cardinality of the set of all even positive integers. I hope you won't be disturbed by this invention of new numbers. After all, every number is an "invention," some of them of fairly recent origin. The negative numbers were invented just a few hundred years ago and used to be called "fictitious numbers" because not all mathematicians in those days believed they were "real." More recently, we have the advent of the so-called imaginary numbers—which, of course, are no more imaginary than any other numbers.

Cantor was able to establish, in his theory of sets, the theorem that there is no greatest transfinite cardinal number (just as there is no greatest positive integer). But then, what about the set of *all* sets? This "superset," according to his notion, is a legitimate set every bit as much as the other sets mentioned so far. Surely no set can have more members than this set of *all* sets? But then the transfinite number associated with this set (that is, its cardinality) must be greater than any other transfinite number, contradicting the theorem that there is no greatest transfinite number. Something clearly was wrong with a theory that permitted such notions as that of the set of all sets. And yet, there was nothing in the theory prohibiting such a notion. Nor is this all.

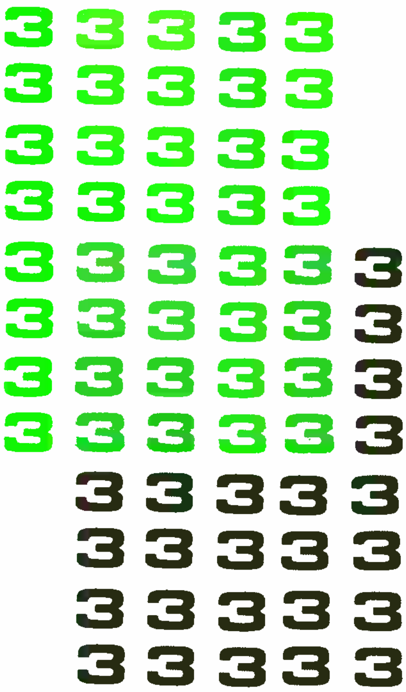
The paradox of sets

Even more serious perhaps was the discovery by Bertrand Russell (around 1902) of a paradox (discovered independently by Zermelo) depending on nothing more than the mere concept of *set*. Russell distinguished two kinds of sets: those that may contain themselves as members, called *extraordinary sets*, and those that do not, called *ordinary sets*. For example, the set of all concepts is itself a concept and so belongs to the set of all concepts; it is therefore an extraordinary set. Most sets, of course, are ordinary (for example, the set of all even numbers is not itself an even number). If this notion of extraordinary sets seems farfetched, do not be concerned, for the paradox involves the set of all ordinary sets. Call this set O. Thus, O represents all those sets that do *not* contain themselves as members. The question now is: What about O itself, is it ordinary or extraordinary? Suppose O is ordinary; then O belongs to O, the set of all ordinary sets; that is, O must be extraordinary. But if O belongs to the collection of extraordinary sets, call it E, then O belongs to E and not to O; so O must be ordinary. What it amounts to is this: O is ordinary if and only if it is not (that is, if and only if it is extraordinary). This is the theorem that can be established using a somewhat more formal proof than would be worthwhile setting down here (see the bibliography). But you can see the contradiction; it's like proving that a thing is black if and only if it is white.

To elucidate the nature of this paradox further, and to show that the problem is more general than the special set-theoretical context in which it occurs, consider the following semantic variation on the theme (first stated by Grelling around 1908). Let us classify all adjectives under two categories: those that describe themselves, such as "English" and "polysyllabic," and those that do not, such as "French" and "monosyllabic." The former we call *autological;* the latter *heterological*. What about the adjective *heterological*? Does it belong to the former category or to the latter? If "heterological" belongs to the list of heterological adjectives, then "heterological" is heterological (just as "polysyllabic" is polysyllabic) and so it must be autological. But if "heterological" is autological then (just as "monosyllabic" is polysyllabic) it doesn't describe itself and so it must be heterological, which brings us back to the beginning of the argument.

Impredicative definitions

You have, no doubt, noticed the peculiar self-referential character of the paradoxes, so that the simplest way of disposing of them would seem to be an injunction against the use of self-referential or *impredicative* (as they are called) definitions. This is precisely the attitude of Russell in enunciating his "vicious circle principle" when he says, "Whatever involves *all* of a collection must not be one of the collection." Or, alternatively, "If, provided a certain collection had a total, it would have members definable



only in terms of that total, then the said collection has no total." However, I feel that this is unsatisfactory from several standpoints. For one thing, simply banning the thing that causes the trouble does not elucidate the nature of the trouble nor does it explain how it can occur in the first place. For another, there is no guarantee that a blanket exclusion of impredicative definitions alone will prevent the occurrence of all future paradoxes. Most important, perhaps, is the fact that not all impredicative definitions cause trouble, so there must be something additionally peculiar (besides just their impredicativity) about the concepts that lead to the paradoxes. For example, to call a certain basketball player the tallest man on the team is surely completely nonparadoxical, even though the characterization is done on the basis of the totality to which the player himself belongs. Furthermore, mathematics itself needs, and has continued to use, certain impredicative definitions that appear perfectly harmless. Recall, for example, the definition of the least upper bound of a given set as the smallest member of the set of all upper bounds of the given set. Or again, recall from elementary calculus the definition of the maximum of a function in a given interval as being the greatest of the function values in the interval—clearly an impredicative definition.

It appears after almost 70 years of discussion that self-reference, in and of itself, though it may be circular is not necessarily *viciously circular*. The problem boils down to having to decide in each specific case whether to permit the use of impredicatively formed concepts. But unfortunately it can be shown that this problem is undecidable in the sense that the attempt must result in an infinite regress.

An entirely different approach to the problem of set theory consists in laying down an axiomatic foundation for the subject. The foremost example of this axiomatic approach is Zermelo's system of 1908, more recently improved by Fraenkel, Bernays, Von Neumann, and others. But, although set theory today is a perfectly respectable branch of mathematics (and the theory of numbers, classic analysis, and even Cantor's general set theory, can be based on it) there still remains a fly in the ointment. This is the now-famous Zermelo *axiom of choice*, about which almost all of the ensuing controversy has centered up to the present time. Before I turn to it, however, let me mention an axiom or two from those remaining in Zermelo's system to give you a feeling for the nature of the axiomatization.

As you are aware from the preceding discussion, the assumption that to any well-defined property $P(x)$ there corresponds a set S such that x is a member of S if, and only if, x has the property P, can lead to formal contradiction. The following example presents no problem. Let $P(x)$ be the property of being an even positive integer; that is, interpret $P(x)$ to mean "x is an even positive integer." Then S is the set of even positive integers. But suppose we take P to be the property of not belonging to itself—i.e., interpret $P(x)$ to mean "x is not a member of x" then, on substituting S for x, we get Russell's paradoxical set ("the set of all things that are not members of themselves"). To circumvent this, Zermelo introduced his *axiom of selection*. Instead of employing properties to define sets, this axiom permits their use only for carving out a subset from a given set. The axiom may be stated somewhat as follows: Given a well-defined property P and a set a there is a set b such that x is a member of b if, and only if, x is a member of a and at the same time has the property P.

Another axiom (given here in the weak form introduced by Fraenkel in 1958) states that for any two different sets R and T there exists a set S containing exactly R and T. This axiom permits one to form a new set out of a given pair of sets. Repeated application of the axiom allows the construction of more complicated sets of various kinds. Axioms such as this, and the axiom of choice to which I shall turn in a moment, are designed to facilitate the construction of new sets out of given ones. For example, Fraenkel in a recent modification of Zermelo's original axiomatic theorem is able to prove the following:

> **Theorem.** For any two sets R and S there exists the set whose members are just those that belong to both R and S.

If you have had some prior contact with set theory you will recognize this new set as being the so-called *intersection* of R and S. In an earlier version of Zermelo's set theory Fraenkel introduced this (as well as the *union* of two sets) as one of his fundamental operations. In his most recent modification, however, he postulates only the existence of the union of two sets, and proves the existence of their intersection. Incidentally, it is interesting to note that in Zermelo's original system of axioms he postulated the existence of sets containing just a single member as well as the existence of the set containing no members, the null set mentioned previously. But Fraenkel is able to *prove* their existence on the basis of his axioms.

The axiom of choice

Today, there exist a number of variants of the original 1908 Zermelo set theory. For example, there is Church's form of the Zermelo–Fraenkel–Skolem theory, there is Bernay's theory, there is Von Neumann's theory, and there is Gödel's theory—to mention the most prominent.

Each of these has certain special features; it is perhaps impossible to say which is the best. They all appear to avoid the paradoxes that shook the early foundations of the subject, but unfortunately there is one big thorn that continues to irritate a lot of today's mathematicians: Zermelo's axiom of choice. It is almost embarassing in its simplicity and intuitive appeal; yet, as you will soon see, it has the most amazing and far-reaching consequences. Before stating the axiom I must formally introduce the notion of subset. It is just what you would expect it to be: If, whenever x is a member of S and we also have x a member of T, then S is called a *subset* of T.

One form of the axiom of choice may now be stated as follows: *Given a collection of nonempty subsets of a set there is a set that contains just one member from each of these.* Let me rephrase this a bit to show you how obvious it is when the collection is finite: For any collection C of nonempty, disjoint (that is, none of the sets in the collection have any members in common) sets S, there exists at least one set T containing one and only one element from each of the sets S.

Consider the following illustration: Suppose the collection consists of just two nonempty, disjoint sets A and B. Since A is nonempty there is an element x in A; similarly, there is an element y in B (distinct from x, since A and B are disjoint). Thus, we can form the set T whose only members are x and y. Incidentally, notice that sets A and B may each obviously contain infinitely many members; we require only that the collection C be finite. In fact, when C is finite, Zermelo's axiom is not really required, since the remaining axioms of set theory are sufficient to permit the formation of T.

But the case where C consists of an infinite collection of sets $A_1, A_2, A_3, \ldots$ just doesn't come through in the same simple, apodictic way. The function of the axiom of choice is to sanction the formulation of T even in this case. That is to say, the case for the existence of a set T containing members $x_1, x_2, x_3, \ldots$, where x_1 is a member of A_1, x_2 is a member of A_2, x_3 of A_3, etc., is made by an appeal to this axiom.

Although Zermelo didn't explicitly formulate this concept as an axiom until 1904, a principle of this kind had been used regularly to guarantee the existence of certain sets that cannot be otherwise obtained. Years before, Cantor (and others) unconsciously applied such a principle as a matter of course. And today, certain branches of mathematics (e.g., topology) cannot get along without it. In spite of its long history, as well as current use by some of the world's foremost mathematicians, and in spite of its apparent credibility, many today totally reject it; others use it, but with varying degrees of skepticism. However, some equally eminent mathematicians (notably topologists) accept it without hesitation.

Before discussing how this surprising state of affairs came about, let me give you an example, again attributable to Bertrand Russell, that beautifully illustrates the difference between a situation requiring some principle or axiom of choice and one that does not. Given an infinite set of pairs of shoes it is a simple matter, *not* requiring the axiom of choice, to form a set of individual shoes using the following selection rule, say: form the set of all left shoes. That is to say, there's a predicate, "left shoe," that can be used as a basis for the selection. But given an infinite collection of pairs of socks, socks being manufactured in identical pairs, how are we to form a set containing just one sock from each pair? The axiom of choice simply sanctions the formation of such a set of individual socks from the set of sock pairs.

Infinite sets

The "choice" Russell is talking about here is not an actual, physical picking-out-by-hand of socks from a bundle of socks. For one thing, there is an infinite number of pairs of socks in the "bundle." But even if you can conceive of some infinite being doing the selecting over an infinite time, this is not the point. The "choice" of "elements" from the "infinite set" of socks is a *mathematical* (conceptual, if you will) operation, not a physical one. And for this you need some sort of mathematical rule. The axiom of choice is appealed to in just those cases where no rule can be formulated for "choosing" or "selecting" certain members out of a given set.

Notice that the root of the difficulty does not lie just with the fact that we're dealing with an infinite set, as the situation with the shoes shows. To make this clear, here is another situation that does not require the axiom of choice. Let C be an infinite collection whose members are nonempty, disjoint sets of natural numbers. Some examples of sets in this collection might be: $\{1, 2, 3\}$; $\{8, 9, 5\}$; $\{23, 7, 100\}$; and so on. Then the following selection rule may be used to form a new set containing one and only one member from each of the sets in the infinite collection C: choose the least number in each set. Because it is known that in every nonempty set of natural numbers there is a least number, there is no need for a principle of choice here.

It will be instructive to give some typical examples of highly credible theorems whose proofs use the axiom of choice, especially since I want to follow this with still another theorem that was established using this axiom but whose credibility was almost immediately challenged. You may recall the following elementary result from calculus:

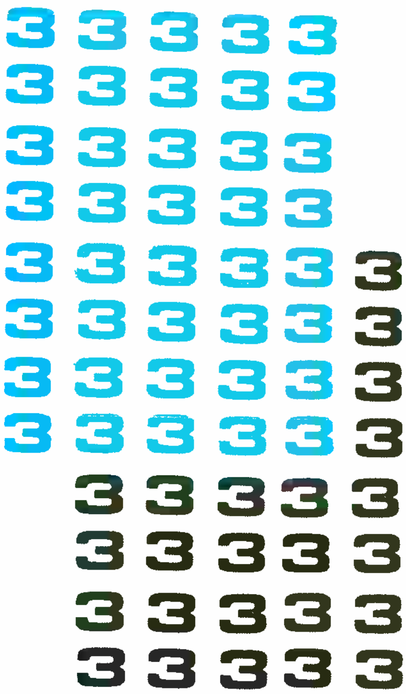
> **Theorem.** If L is the least upper bound of a set S, then there exists a sequence $a_0, a_1, a_2, \ldots, a_n, \ldots$ of members of S such that $\lim a_n = L$.

The proof of this theorem involves the right to choose an element a_n from a set S_n simultaneously for infinitely many values of n without giving any rule or property for determining which element is chosen.

The next theorem may appear somewhat novel to you but I cite it because the use of the axiom of choice in establishing it is quite straightforward.

> **Theorem.** An infinite set S has a denumerable (i.e., "countable") infinite subset.
>
> **Proof.** S is obviously not empty, otherwise it would have the finite cardinal 0. Thus, S has a member a_1. Then $S - \{a_1\}$ is not empty, for otherwise S would have the finite cardinal 1. So S has another member a_2. Continuing in this manner you can select distinct members $a_1, a_2, a_3, \ldots$ of S corresponding to the natural numbers 1, 2, 3, ... Thus, we have shown that the subset of S whose members are $a_1, a_2, a_3, \ldots$ is "countable" (by the numbers 1, 2, 3, . . .).

Notice once again that no rule is given for selecting members $a_1, a_2, a_3, \ldots$ out of S. But the proof seems reasonable enough, does it not? Incidentally, it is the corol-



lary to this theorem that says that aleph null is the first transfinite cardinal.

Well-ordering theorem

The next theorem, and the one that caused all the trouble, is as follows:

> **Zermelo's well-ordering theorem.** Every set can be well ordered.

The original reaction of a good many mathematicians to this incredible result was that there must be something wrong with the proof. Looking for a flaw in the proof, Emile Borel discovered that Zermelo had used an apparently obvious principle, one that others had been using for quite a while without ever bothering to state it explicitly—viz., the principle or axiom of choice. Even more amazing, Borel was able to show not just that Zermelo's theorem is based on the axiom of choice but that the theorem is actually equivalent to it! Since that time quite a number of theorems have been shown to be equivalent to the axiom. But let's see first why the Zermelo theorem is so amazing.

What does it mean to say that a set is *well ordered*? First, let us see what it means to say that a set is *simply ordered*. It is simple indeed. Remember the symbol $<$ (read: "less than") from algebra? Well, a set S of elements $a, b, c, \ldots$ is said to be simply ordered if: (1) whenever $a \neq b$ then either $a < b$ or $b < a$; (2) $a < b$ implies that $a \neq b$; (3) $a < b$ and $b < c$ imply that $a < c$. Now then, a set is *well ordered* if the following additional property holds: For any nonempty subset S' of S there is a member a of S' such that $a < b$, where b is any other member of S'. In other words, a simply ordered set S is well ordered if every nonempty subset of S has a "first" element. The set of natural numbers (1, 2, 3, . . .) afford an obvious example of a well-ordered set.

To repeat: Zermelo's theorem says that for *any* set there exists a relation well-ordering that set. This means that, for example, the set of points from 0 to 1 on the real line can be well ordered. Yet no one to this day has been able to show how any set of this cardinality can be well ordered. In fact, there are many sets that mathematicians have been unable not only to "well-order" but even to order at all—for example, the set of all real functions of a real variable. Furthermore, consider the ordered subset of all points to the right of 0 on the real line up to the point labeled 1. What is the *first* point in this subset? That is, what is the first point to the right of 0 on the real line?

Let me quickly point out, however, that the symbol $<$ appearing in the definition of a well ordered set is an abstract dyadic relation: It doesn't have to mean "less than," and any other symbol (such as R for "relation") could have been used in the preceding definition. In fact in the previous paragraph we interpreted $<$ to mean (with regard to points on a line) "is to the left of." The theorem only claims that some kind of well-ordering relation exists for every set.

The Banach-Tarski paradox

Once the well-ordering theorem was shown to be equivalent to the axiom of choice, you can see why a number of mathematicians began having serious doubts about the validity of the axiom. But even stranger consequences of this axiom soon appeared on the scene—for example, Hausdorff's theorem (1914) that half of a sphere's surface is congruent to a third of it or the more general Banach-Tarski paradox (1924). This latter theorem says, in effect, that given two *solid* spheres one, say, the size of the sun and one the size of a pea, then both the sun and the pea may be divided into a finite number of nonoverlapping little solid pieces so that every single part of the sun is congruent to (that is, of the same size and shape as) a unique part of the pea; moreover, after each of these small portions have been matched with each other *no portion of the sun will be left over*. To put it another way: *It is theoretically possible to partition a sphere as large as the sun into little solid pieces and then without any compression or distortion of these pieces, to rearrange them, leaving none out, into a sphere the size of a pea.* As far as I know, no flaw has been discovered in the proof of this theorem; it is a consequence of the axiom of choice.

Let me hasten to add that, paradoxical as the Banach-Tarski theorem may be, it is not a contradiction. To this day, not a single contradiction has been derived from the axiom of choice, even though a vast number of results have been derived from it—most of them not in the least paradoxical. In fact, we now know that if the remaining, generally accepted axioms of set theory are consistent, then these axioms, plus the axiom of choice, form a consistent system. This was shown by Kurt Gödel in 1938.

Perhaps at this point you are beginning to have some doubts about the axiom of choice. In that case, let me state an even more fundamental axiom, at the foundations of logic itself, that you may find it difficult to doubt. If so, then you must be prepared to also accept the axiom of choice, for it can be derived as a theorem from this logical axiom (due to Hilbert): There exists a function f associating with every property P (for which there exists at least one object having the property P) an object $f(P)$ having the property P. If *this* axiom isn't true, it is difficult to see how logic can be possible. And yet, it implies the axiom of choice.

What is the status of the axiom of choice today? As I already mentioned, topologists use it rather freely; in fact, there is little evidence that any significant portions of topology can be derived without its use. The same holds true for modern analysis, with its study of Hausdorff and Banach spaces and use of Lebesgue measure theory. (Lebesgue and Stieljes integrals, which represent a generalization of the familiar Riemann integral of elementary calculus, are finding increased application in modern probability theory.) Algebraists have mixed emotions about using the axiom; they tend to proceed as far as possible without it. When the axiom is required to establish some highly general result the custom now is specifically to call attention to the fact.

The latest word in this area consists of an ingenious proof evolved by Paul Cohen, in 1963, to the effect that if the remaining system of axioms of set theory is consistent, then the system remains consistent with the inclusion of the statement: "The axiom of choice is false." Combined with Gödel's results, this statement then means that the axiom of choice is an *independent* axiom. The situation here is rather similar to that in geometry at the time when the denial of Euclid's fifth postulate (the parallel postulate) was shown to result in a consistent (non-Euclidean) geometry. Thus, one speaks today of Cantorian and non-Cantorian set theory.

Although the axiom of choice is still controversial, I do not want to leave you with the impression that all the theorems that have been shown to be equivalent to it, or to depend upon it for their proofs, are incredible. Quite the contrary. But a further discussion of these matters would have to be prefaced by a much deeper discussion of set theory in general and Cantor's theory of transfinites in particular. Those who may wish to join in the controversy will want to consult the bibliography for further reading suggestions.

An annotated bibliography

Almost any book on topology or modern algebra devotes its first chapter to set theory, usually in a rather concentrated dose. If you have had no prior contact with the subject you might find a good starting place in my book, *Introduction to Some Mathematical Structures* (Woodland Hills, Calif.: Dickenson Publishing Company, Inc., 1967). This book has been written especially for the nonmathematician; the first four chapters, plus working through some of the exercises, will provide you with a very solid basis for pursuing more advanced topics in set theory.

Next, I recommend the magnificent book by Raymond Wilder: *Introduction to the Foundations of Mathematics* (New York: Wiley, 1965). The first two chapters have a superb discussion of axiomatic systems and the important notions of *consistency*, *independence*, and *completeness*. The third chapter introduces the reader to set theory and has a first-rate discussion of the axiom of choice and the Russell paradox. The next two chapters offer the best introduction to Cantor's theory of transfinite cardinals and ordinals that I know of, and a discussion of the well-ordering theorem appears in Chapter 5. The only prerequisite for this book is a certain amount of "mathematical maturity."

In the same category of excellence is Fraenkel and Bar-Hillel's *Foundations of Set Theory* (Amsterdam: North-Holland, 1958).

If you are mathematically inclined you will find Stephen Kleene's book, *Introduction to Metamathematics* (Princeton: Van Nostrand, 1952), invaluable (but let me warn you, it is pretty heavy stuff).

As to some of the specific topics I discussed, Russell's paradox appears in his book, *The Principles of Mathematics* (London: George Allen & Unwin Ltd., 1951); it is discussed in Section 78 and again in Chapter 10. There is an interesting little story to this. Mention of the paradox appears as a postscript in the second volume of Frege's monumental work, *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet* (Jena: H. Pohle, 1903). Just as this volume was going to press Frege received a note from Russell describing the paradox he had discovered. Frege's comment speaks for itself (and for the man): "A scientist can hardly meet with anything more undesirable than to have the foundation give way just as the work is finished. I was put in this position by a letter from Mr. Bertrand Russell as the work was nearly through the press." If you want the flavor of Frege's work you can consult Max Black's translation of Sections 86–137 of Volume 2: "Frege Against the Formalists," *Philosophical Review*, vol. 59, pp. 77 and 202, 1950.

The first proof of the well-ordering theorem is given in Zermelo's paper, "Beweis, dass jede Menge wohlgeordnet werden kann," *Mathematische Annalen*, vol. 59, pp. 514–516, 1904. You might also want to consult the discussion by Borel *et al.* appearing in volume 60 on pages 181, 194, and 465. The second proof of this theorem appears in Zermelo's paper, "Neuer Beweis für die Wohlordnung," *Mathematische Annalen*, vol. 65, pp. 107–128, 1908. Herein also appears his independent discovery of Russell's paradox. The axioms of Zermelo's set theory also appear on pages 261ff of this volume in a paper entitled, "Untersuchungen über die Grundlagen der Mengenlehre."

Someone once remarked, on being asked to what he attributed his success, "By studying the masters, not their pupils," so you may wish to consult Cantor's *Contributions to the Founding of the Theory of Transfinite Numbers*, translated by P. E. B. Jourdain (Chicago: Open Court, 1915). It is beautifully written. References to the axiom of choice are legion. A good starting place would be Fraenkel's paper, "L'axiome du choix," *Revue Philosophique de Louvain*, vol. 50, 1950.

The proof (highly technical) that the axiom of choice is consistent with the remaining axioms of set theory (provided, of course, they in turn are consistent) appears in Gödel's classic paper, "The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory," (Princeton: Princeton University Press, *Annals of Mathematic Studies*, no. 3, 1940). The first chapter lists Gödel's variant of the axioms of set theory. Incidentally, Cantor's continuum hypothesis is a fascinating chapter in the history of set theory. There is an admirable semitechnical discussion of this in Gödel's paper, "What is Cantor's continuum problem?" *American Mathematical Monthly*, vol. 54, pp. 515–525, 1947.

The Banach-Tarski theorem appears in the paper, "Sur la décomposition des ensembles de points en partie respectivement congruentes," *Fund. Math.* vol. 6, pp. 244–277, 1924. A highly entertaining and telling defense of self-referential expressions appears in Popper's paper "Self-reference and meaning in ordinary language," *Mind*, vol. 63, p. 162, 1954. Most recently, there is the lovely article by P. J. Cohen and R. Hersh in the December 1967 issue of *Scientific American*: "Non-Cantorian set theory."

Automation comes to the printing and publishing industry

Production and distribution of magazines, periodicals, and books

Exciting new methods of information acquisition, the development of fully automated presses and new binding systems, and satellite plants are molding the future shape of this vital industry

Gordon D. Friedlander *Staff Writer*

Existing information retrieval systems, hardware and software for automated publication, and newly developed press and bindery equipment* are the precursors of ever more dramatic and sophisticated systems now in the R&D stages. For example, a fully automated web offset press could be in commercial production within the next few years. At the same time, there is a quiet revolution in bookbinding, based upon plastics technology, that is rendering obsolete conventional binding methods. The day is not far distant when almost completely automatic satellite plants will be operated for producing and distributing periodicals, indexes, catalogs, and books.

The trend toward automation in the printing and publishing industry (P&PI) began several years ago and its pace is accelerating. One of the leading exemplars in the special publications field is the . . .

National Library of Medicine and MEDLARS

In 1961, the National Library of Medicine developed specifications for a computer-oriented information storage and retrieval system (and automated publication technique) called MEDLARS (Medical Literature Analysis and Retrieval System), the planned objectives of which were

1. To improve the quality of *Index Medicus*, the Library's comprehensive monthly index of medical journal literature, and simultaneously to increase the number of journals indexed in this periodical.
2. To provide for the production of compilations similar to *Index Medicus* in format and content, but devoted to specialized subject fields.
3. To provide for prompt and efficient searching of a large computer memory store of information for citations on biomedical and biomedically related literature.
4. To reduce the time required for preparing *Index Medicus*.

The system design was completed in 1962, but the preparation of computer programs, installation of equipment, training of personnel, and detailed system testing and debugging required another two years. The system has three major subdivisions: an *input subsystem*, in which the skills of professional indexers are used in conjunction with the capabilities of a large-capacity digital computer (Honeywell 800 system); a *retrieval subsystem*, in which the capabilities of professional librarians and literature searchers are also used in conjunction with the computer;

* For definitions of bindery terms, see the Appendix at the end of this article (page 62).

and a *publication subsystem*, which converts retrieved citations to a printout on photosensitive film.

The input subsystem. Citations and other literary information are fed into MEDLARS by means of this system. In the initial step, journals are analyzed by trained personnel who assign appropriate descriptors to each article from "Medical Subject Headings," which is the Library's controlled vocabulary listing.

After the journal articles have been indexed, operators using Friden Flexowriters (with specially modified keyboards to include diacritical marks required for certain foreign languages) prepare basic unit records by converting information on the indexers' data sheets to punched paper tape. The basic unit record entering the computer consists of journal references, article titles, authors, and subject descriptors.

The Flexowriter keyboards permit the input of 88 alphanumeric characters; the machines also have line-deletion capability for use by the typists in making rapid corrections. Typewritten copy, produced simultaneously with the punched paper tape, is edited and checked for accuracy by proofreaders; then the tapes are spliced together in batches and fed into the computer for batch processing. The computer configuration has seven magnetic tape units, a paper tape reader, card reader and punch, and a high-speed printer. The central processor contains 8192 words (48 kilobits) of core memory storage. In addition to the 800, a Honeywell 200 system was installed in 1965 to replace much of the former peripheral equipment that was required.

By means of the computer's input programs, the information on the punched paper tape is edited, transcribed onto magnetic tape, and then incorporated into two major data files—the *compressed citation file* and the *processed citation file*. The former contains highly coded citations that can be retrieved as "demand bibliographies," which are intended for the individual use of a requester and not for publication. The process citation file contains citations used in publishing *Index Medicus* and bibliographies that are intended for serialized publication and wide distribution.

Retrieval subsystem. Citations stored in the computer are recovered by means of the retrieval subsystem. Requests for bibliographic citations from physicians, scientists, librarians, etc., are forwarded to a staff of specialists who analyze the inquiries and formulate search statements in machine-compatible form. These requests are encoded on punch cards rather than on paper tape. The retrieved citations, machine-edited and assorted, are recorded on magnetic tape and then decoded into plain English prior to being printed either on cards or on continuously tabulating paper by the computer's high-speed printer.

System loads. The current annual volume of about 150 000 indexed papers is expected to increase to 250 000 by 1970. Also, at some future date, about 10 000 citations of monographs (books and other nonperiodical volumes) will be entered into the system annually.

Publication subsystem. This subsystem provides recurring bibliographies by processing citations for eventual printing from photo-positive transparencies. In this process, the magnetic tape files of retrieved citations are used for the preparation of printed copy by a computer-driven photocompositor (a custom-built Photon machine) called GRACE—an acronym for *GR*aphic *A*rts *C*omposing *E*quipment. Operating at a character generation speed of about 300 per second (Fig. 1), GRACE can use 226 different alphanumeric characters in preparing 23-cm-wide positive photographic film. Character sets include a 6-point font of regular and boldface upper- and lowercase characters, a 10-point font of upper case, and a 14-point font of upper case. The exposed film is developed by an automatic film processor; it is then inspected and cut into standard-page-size sheets.

FIGURE 1. Control console of the Graphic Arts Composing Equipment (GRACE) system, installed by the National Library of Medicine in 1964 to serve as part of its publications subsystem. This photocompositor operates at a speed of 3600 words per minute.

Automation: The next generation

The existing MEDLARS system is but the threshold to a remarkable coming era of drastic change in the P&PI. Automation in typesetting, photocomposition, and graphic arts processes is being sequentially followed by machine surveillance of control functions on web presses. And research is under way for similar controls on sheet-fed presses. Bindery operations for all of the large-volume binding techniques—wire stitched, adhesive, and sewn—have been developed into in-line systems.

Toward automatic control of web printing. The high-speed web press for printing newspapers and national magazines has put the ability to control the printing process beyond manual reach or rapid human intervention. At speeds of up to 60 000 copies or signatures per hour, the human eye cannot keep up with the pace. Thus emergency changes or corrections by manual means cannot be made before hundreds—or thousands—of copies have been produced. Further, pressmen cannot keep a simultaneous watch on the many press conditions that are in a constant state of flux during a large-volume run.

This inability to maintain efficient manual control has often caused huge losses in time, money, and materials to printing houses, particularly those using the web offset process. Because of this, the automated control trend has been evolving steadily over the past decade so that today we can see the ultimate goal of complete automation of the printing process.

Initially, automatic controls were applied by means of optical scanners and mechanical feeders[1] that were used for register control (alignment of color printing) and web break detection. These elementary controls at least told the operator something about on-line conditions and what corrective action to take during a press run.

In the next step, a group of "black boxes" were installed at various locations on the press to provide certain corrective actions automatically. In addition to sensors, these devices included motors and control mechanisms to replace the slow-response manual handwheels. Push buttons permitted the operator to instruct or program proper speed and pressure settings; then the black boxes took over the monitoring of the press and the maintenance of the proper settings.

The individual isolated control devices at various locations on the press units, however, presented a problem in communications. For example, when a need for adjustment in a black-box program at one end of the press produced a reaction at the other end, a "whoop and holler" vocal alarm or a mad dash down the pressroom catwalk were the conventional means of conveying the message.

The press console. The communications problem is now simplified by the central control console (Fig. 2), which contains a considerable degree of computer circuitry, visual readouts, and push-button controls that permit the press operator to change settings almost instantly.

The Goss Digital Color-Comp console[2] shown in Fig. 2 is one of the most advanced devices on the market. Although designed for newspaper presses, it will soon be applied to the commercial press field. Essentially, it provides the means, from one station, of automatically programming settings and then adjusting them during the run. This unit is primarily designed to monitor and control ink flow, but it could be applied as well to lithographic fountain (three- or four-color) control.

The Goss console permits the automatic increase or decrease of the ink flow to any column on any page of the publication. The programming of this operation before the run and the on-line ink adjustments during the run are made by means of the buttons indicated on the console diagram.

Other consoles on commercial web presses have centralized various control functions such as press speed, register, and drying oven temperatures. Telltale signal lights on some consoles have also been incorporated to pinpoint web breaks and other emergency situations.

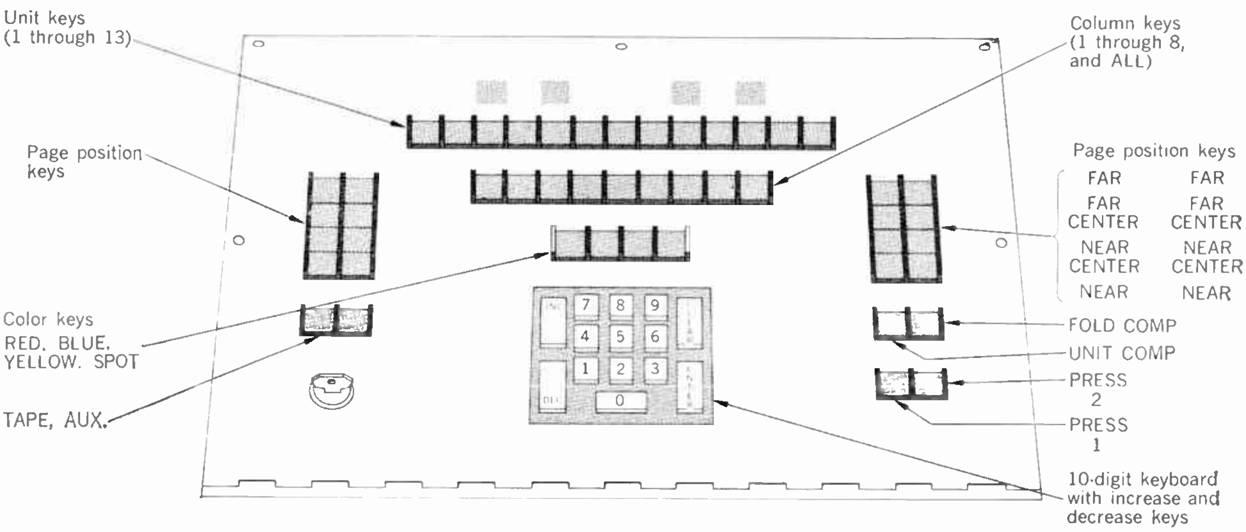
Although the present-day press consoles represent a step toward ultimate consolidation of automated controls, they are still discrete devices that function individually. In short, the operator must convert his control decisions into a number of separate control programs that will perform in parallel both on-line and off-line. But engineers and researchers are already at work on the design of integrated instruments and controls to relieve the operator of the necessity for decision translation. Thus the fully automated printing press—which may be evolved in the next 3 to 5 years—will probably be operated from a central computer and data processing room at a remote location.

The ultimate in press automation. The final stage of automation will be the preprogramming of the entire press operation, by means of analog and digital computer controls, to run off a job that is completely automated from start-up to shutdown.

In this ultimate mode, all press settings would be predetermined before a job reached the pressroom. The types of paper and inks to be used, running speeds, etc., would be programmed into a general-purpose computer by punched paper tape or cards, or by retrieval from a random-access disk file. Other off-line computer programs would handle press-run estimates, production schedules, and other aspects of the job specifications.

The '20 blocks' of press automation. Figure 3 shows

FIGURE 2. Control panel of a Goss Digital Color-Comp console. This and similar consoles will form the "decision center" for future automated press operations.



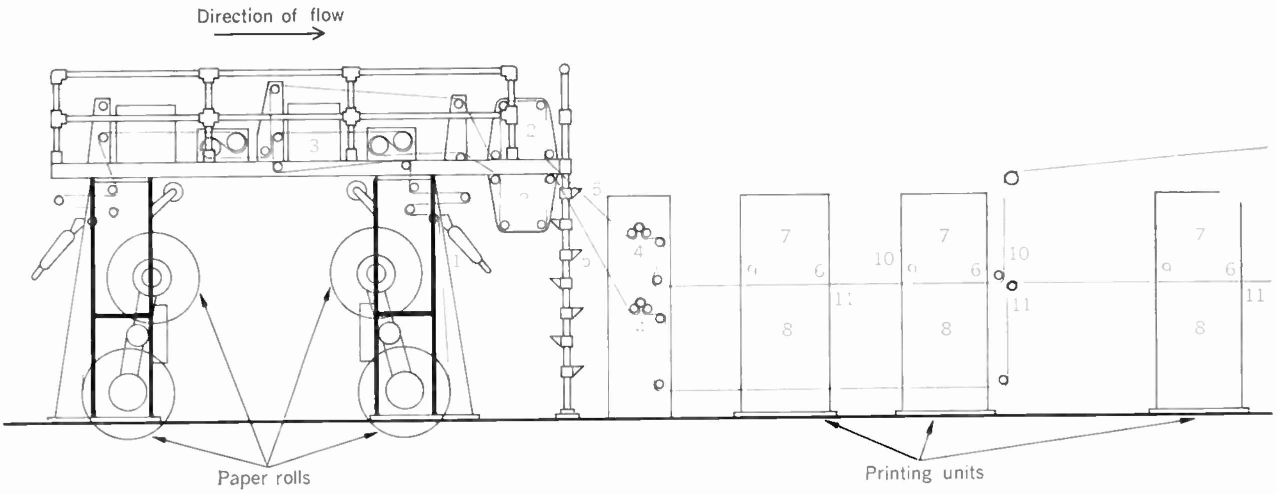


FIGURE 3. Profile of a large web offset press, showing the 20 devices or functions adaptable at the present time to automatic controls.

a profile outline of a typical large-volume, high-speed web offset press,[1] and spots 20 of the press functions that are compatible with automated control from one integrated central station. The following list of equipment is applicable to the web offset press, except as noted:

1. *Flying pasters.* These devices splice new paper rolls into the web without stopping the press, thereby preventing variations in color imprint, register, and ink flow that occur when the press is stopped for manual splicing. (A 10–15 percent increase in productivity and a 2–4 percent decrease in paper waste can be attained with this accessory.)

2. *Automatic web guides.* These devices (located ahead of the first printing unit and before the folder or stacker) monitor the web's lateral alignment by means of photocells, air resistance, or pressure sensors.

3. *Web preconditioners.* Foreign material is removed from the web by vacuum suction, and the web moisture content is stabilized and made uniform by hot air blowers. This is essential in reducing web breakage frequency.

4. *Digital speed-ratio monitor.* This equipment, by means of very accurate sensing and governing of roller under- and overspeeds, maintains uniform web tension to minimize breakage.

5. *Splice detectors.* As the name implies, such devices actuate an alarm so that the operator can remove the spoiled copy containing the web splice from the delivery.

6. *Automatic impression throwoff.* This system is employed in conjunction with item 5 to lift the printing impression automatically as a splice passes the printing unit. The impression is returned after the splice has passed. This technique prevents the tension imparted by the blanket cylinder* from pulling a splice apart, and also prevents the thickness of the splice from damaging the blanket.

7. *Ink system temperature controls, agitators, and pumping systems.* The temperature controls use water circulation through hollow rollers in the ink system to maintain roller temperature at a constant level. (Friction in ink systems can cause rollers to swell from the heat produced, thereby changing the system's ink transfer characteristics.) Agitators are used to prevent a skin of ink from forming on the surface of the fountains, and also to ensure better color consistency and the need for less ink in the fountain. Pumps (used primarily on web rotary presses) are employed when the same ink is required on a majority of press runs. Benefits are less ink spoilage and greater safety for press operators.

8. *Fountain recirculation and level control systems.* The systems (common to rotogravure presses) maintain the proper ink solution levels. Also, the pH value is stabilized by constant recirculation of the ink. With these systems, accidents are greatly reduced.

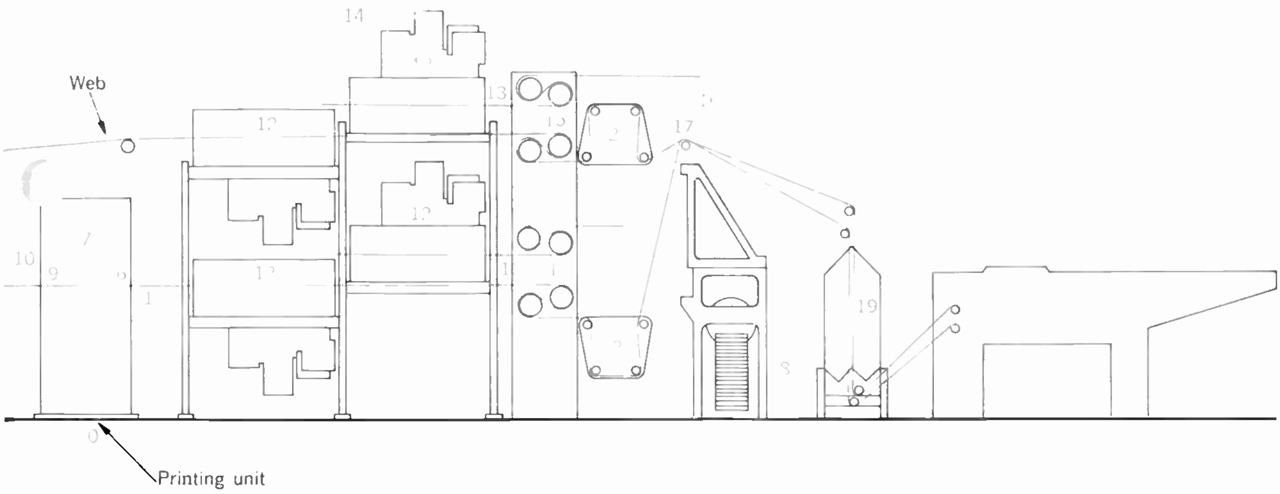
9. *Automatic blanket washers.* These devices wash the blanket cylinders "on the fly," thereby eliminating lost press time and start-up waste when done as a separate operation. The wash is usually done during a "flying paste" splice, and the web carries off any foreign material.

10. *Color register control.* In this equipment, a photocell scanner unit detects color register (alignment) marks printed on the web and corrects for any longitudinal misregister between the printing units. The scanner is synchronized to a counter that transmits a signal for each revolution of the printing unit. If the photocell "reading" of the register mark does not coincide exactly with the counter signal, corrective action will be taken automatically. Such controls reduce spoilage during the run and start-up, and ensure more consistent quality of color plates.

11. *Web break detectors, alarm systems, and severing devices.* Photocell sensors instantly detect web breaks and automatically shut down the press. Break alarm indicator lights on the console tell the operator just where the break has occurred. Severing devices are used in conjunction with the web break detectors to cut the web at strategic points (for resplicing) when a break occurs. This protects the press against paper being pulled back into the printing unit.

12. *Dryers.* These units dry the ink on the moving web by heat application. Variable temperature controls must be included for the proper processing of a wide variety

* For description of blanket cylinders, see the first installment of this article, IEEE SPECTRUM, April 1968, pages 46–62.



of paper stock and inks. While in the dryers, the web is supported by jets of air.

13. *Web temperature control.* This device monitors the temperature of the web as it emerges from the dryer; it also controls the amount of heat imparted to the web by the dryer.

14. *Catalytic smoke incinerators.* These are essentially air pollution control accessories that oxidize the ink fumes produced by the dryer.

15. *Chill-roll temperature controls.* The chill rolls cool the web after it leaves the dryers to ensure the proper "setting" of the ink. These controls maintain uniform temperatures in the chill roll units.

16. *Insert controls.* To obtain higher-quality printing of color display advertising, the insertion of preprinted material is a common practice. However, preprints require special controls to match their register with that of the pages being printed on the press. By feeding preprint rolls into the press under tension, the web is sufficiently stretched to match image length with the other pages. Photocells are used in this control process to detect variations in register and to adjust the tension on the web.

17. *Automatic cutoff registration.* This device is similar to item 10, except that the position of imprint relative to the page fold or cutoff point is controlled. Thus spoilage caused by misregister on start-up, flying paster splices, etc., is greatly reduced.

18. *Stackers.* Many devices are available for handling signatures as they emerge from the folders of web presses that are operating at runoff speeds up to 1000 copies per minute. This equipment mechanically stacks—and sometimes bundles—signatures. Usually, book-size signatures are bundled in this operation because they are easy to handle. Magazine sizes, because they are more difficult to handle, are generally stacked only.

19. *Moisture recovery chamber.* A spraying system within the chamber applies an infinitely thin film of moisture to each side of the web to eliminate static electricity in the signature folding operation.

20. *Press drive controls.* Actually, the digital speed-ratio monitor is part of the complex electronic control system that governs the synchronous rotational speed and stability of press rollers and cylinders. Control devices also include regenerative braking systems to effect emergency stops in a much shorter time, and with less mechanical shock than dynamic braking.

Automation in binding

Bindery automation has actually been under way since the end of World War II, when a few leading binderies installed automatic conveyors and machine-to-machine transfer systems in order to reduce manual handling.[3]

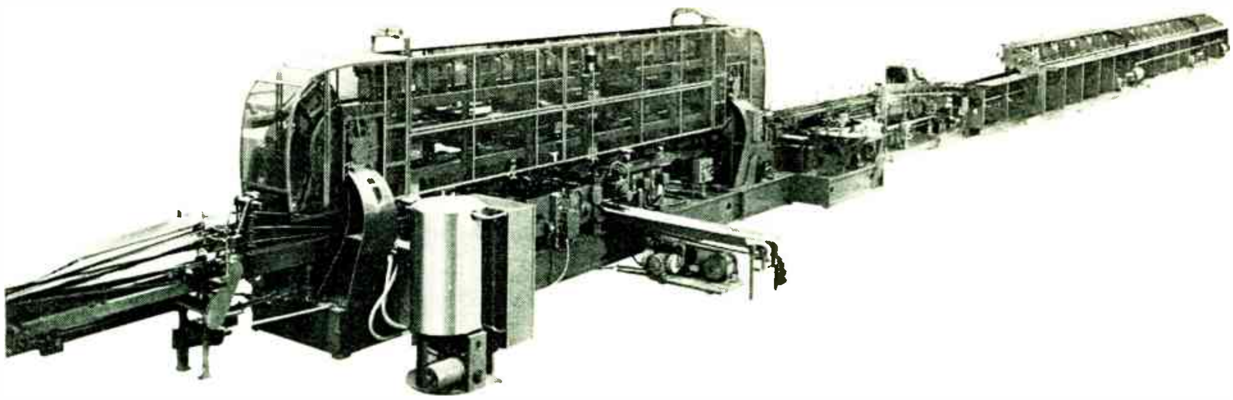
Presently available, however, are complete in-line gathering, stitching, and trimming systems (Fig. 4); fully automated adhesive binding lines (from gathering to trimming); and machine-to-machine conveyor lines that can carry hard-cover books from smashing through building-in and wrapping.

Present research efforts are concerned with developing new automatic feeders, delivery stackers, and systems to store and return books automatically to another machine line whenever one of the units is shut down for repairs or maintenance. Another trend is in the design of faster machine components to step up the hard-cover book production rate from the conventional 2000 per hour to the order of 3000 to 6000 per hour. Also, significant increases in the speed of wire stitching and adhesive binding have been achieved. Thus binding technology has been brought to that point at which complete automation is on the horizon—just as it is for typesetting, photocomposition, and press operations.

Among the most apparent technological advances required on the road to complete automation are—

Machine design improvement. Although most machine adjustments and maintenance operations are mechanically simple, they usually require time-consuming machine shutdowns. For fully automated control, these resetting operations would have to be on-line "rolling adjustments," in which error-sensing servomechanisms would actuate telemotors to do the jobs formerly accomplished manually. And, whenever possible, adjustments set to common specifications should be coupled either mechanically or electronically so that they can be made with one input signal at the master console.

One special problem peculiar to binding equipment is the need for interchangeable components that have to be manually replaced to accommodate different book sizes. To overcome this, specially designed universal com-



A

B

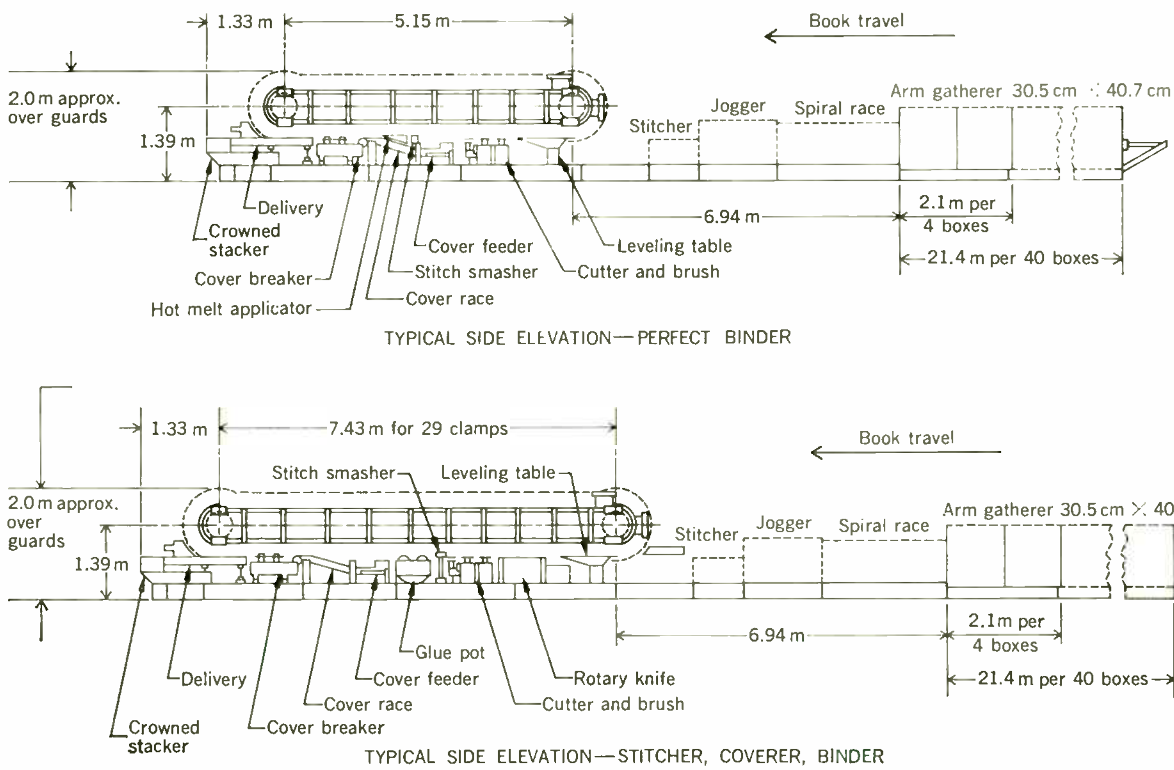
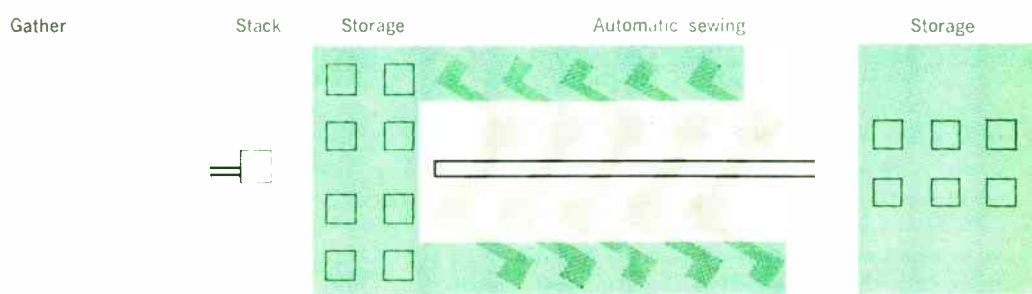


FIGURE 4. A—Dexter bindery line. Length from gatherer at extreme right to stacker at left can be more than 21 meters. B—Side elevations of Dexter in-line machines.

FIGURE 5. Block diagram of an in-line binding system. Shaded areas are those that can now be fully automated; colored areas still require manual handling of materials. Alternatives A and B are two solutions presently being considered for integrating the presently separate operations of sewing and gathering.



Alternative A

ponents for a variety of book sizes will be required.

System controls and communications. A closed-loop control system is one answer to automating the bindery line. In such a system, machine or line operation data are sensed by instruments, transmitted to the computer control center where they are translated into control instructions, and relayed back to a device that adjusts machine settings.

Modern binding lines are becoming so long that closed-circuit television is now being employed to speed communications to operators stationed along the line.

The tough problems of bindery automation

As previously mentioned, adhesive binding and saddle and wire stitching are all done on in-line equipment that can presently be fully automated. And, as indicated in Fig. 5, many steps in hard-cover sewn binding are also in-line. All of the operations shown in the shaded portion of the diagram are those now subject to automation; the colored areas still require the manual handling of materials. The primary problem here is that books, once gathered, must be rehandled signature by signature before being sent to the smashing unit.

Ideally, sewing and gathering should be consolidated into one process, with both functions performed simultaneously on each signature. Alternative A indicates a theoretical machine that might accomplish this by providing a series of sewing heads that would travel in an oval to pass each station on the gathering line. As each signature is extracted from the gathering machine pocket, it would be sewn into the already gathered signatures.

Another solution—based upon existing equipment—would be Alternative B, a combination of conveyors between gathering and automatic saddle sewers. In this case, an oval, continuous-loop conveyor might pick up collated books and deliver them to the input feeders of a line of automatic saddle-sewing machines, dropping a book at each machine in succession. The sewing machines would then feed the completed books into a conveyer for transmission to the smashing operation.

Thus far, bindery automation has been built up by the conveyor linkage of existing machines. Therefore, the key to complete automation lies in the expensive process of machine redesign from the ground up. This will be the trend of the future.

The one-piece hard-cover book

The conventional three-piece hard-cover book casing is actually one of the bottlenecks in the way of simplified and more rapid automated binding. Now, however, there are two types of "one-piece" covers—the former in experimental use and the latter in current commercial use—available:

1. A Tenite polyallomer plastic, manufactured by Eastman Chemical, Inc., and Union Carbide's Ucar plastic sheet, both with scored hinges for backing.
2. The prefabricated case made of conventional heavy-gauge cardboards, with a cloth or paper hinge.

The plastic covers have exhibited a remarkable durability and resistance to abuse in handling. Binders have reported a 15 percent increase in production by switching over to the new materials and binding techniques.

The publishing process—present and future

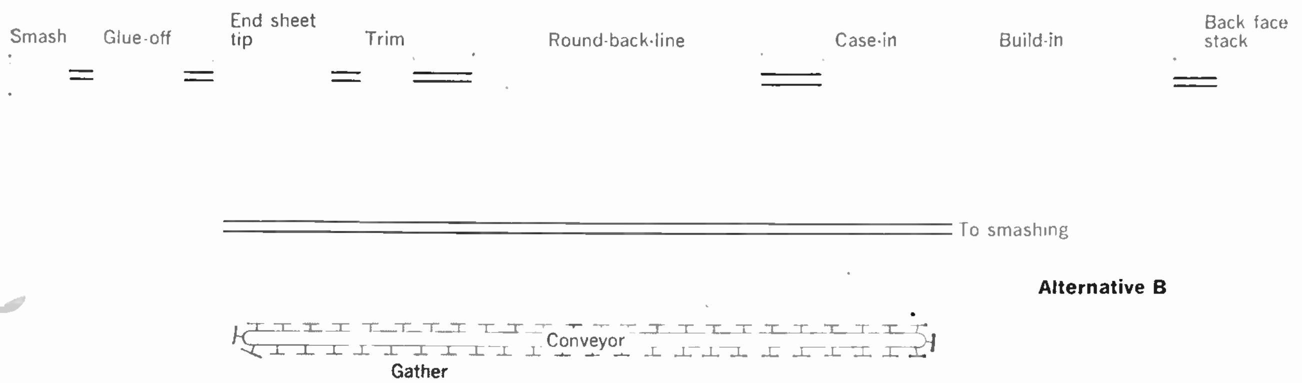
In the conventional publishing process today,[4] the products—newspapers, magazines, or books—are manufactured *first* and then distributed over long distances, either through the mails or by commercial carriers. This entails the time-consuming physical handling and transporting of thousands of tons of published materials. Much of the traditional publishing industry is engaged in a sequential physical manufacturing process in which type, pages, forms, and press plates are created. And each of these steps must be perfected and checked out before the next stage can be undertaken.

The manufacture of publications by conventional printing and binding techniques is most economically accomplished in large lots that are run off in one central plant. But extensive storage facilities and preshipping operations are often necessary for the acquisition of raw materials and the warehousing of the finished product before distribution to the consumer. All of this is very costly.

These problems have led to a revolutionary concept, which essentially says—

First distribute, then manufacture. The new communications technology, with the digital computer as its nucleus, indicates a reversal of the traditional manufacturing–distribution process for quicker and more efficient dissemination of literature. The scheme of what the future may be like in the P&PI is shown in Fig. 6.

Digital coding is the strategic tool in this radical



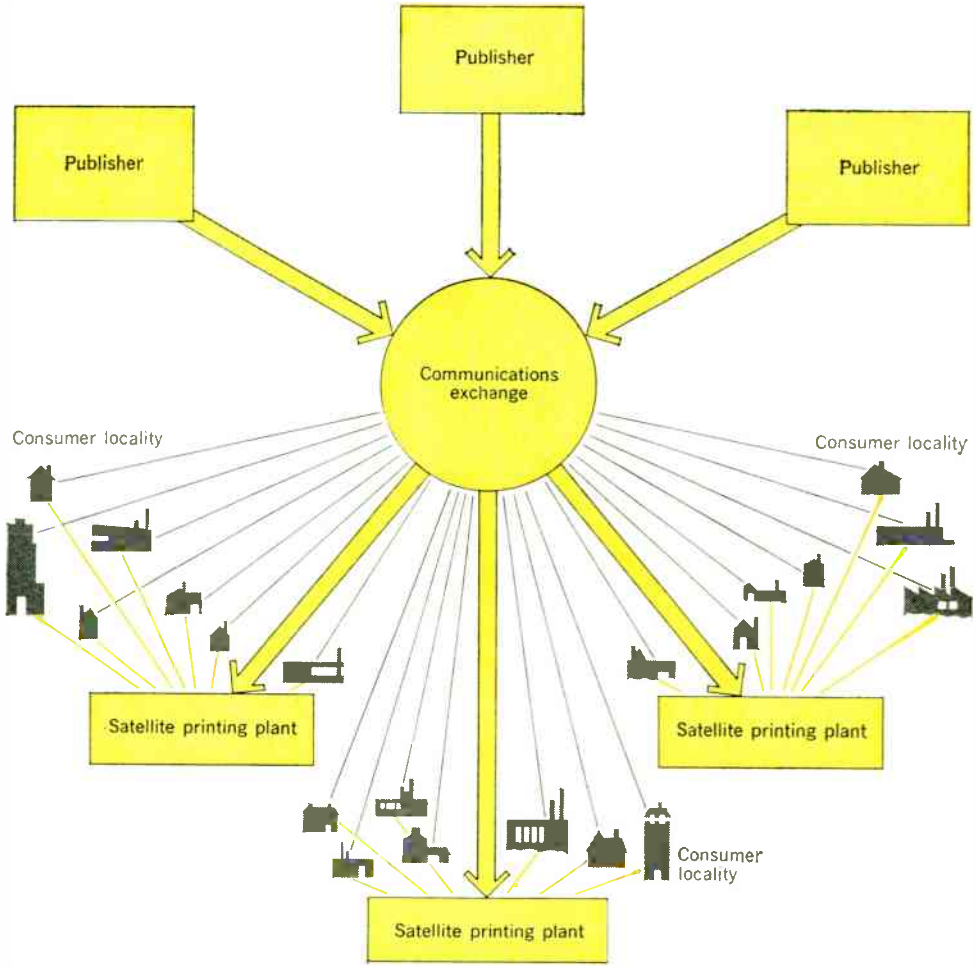
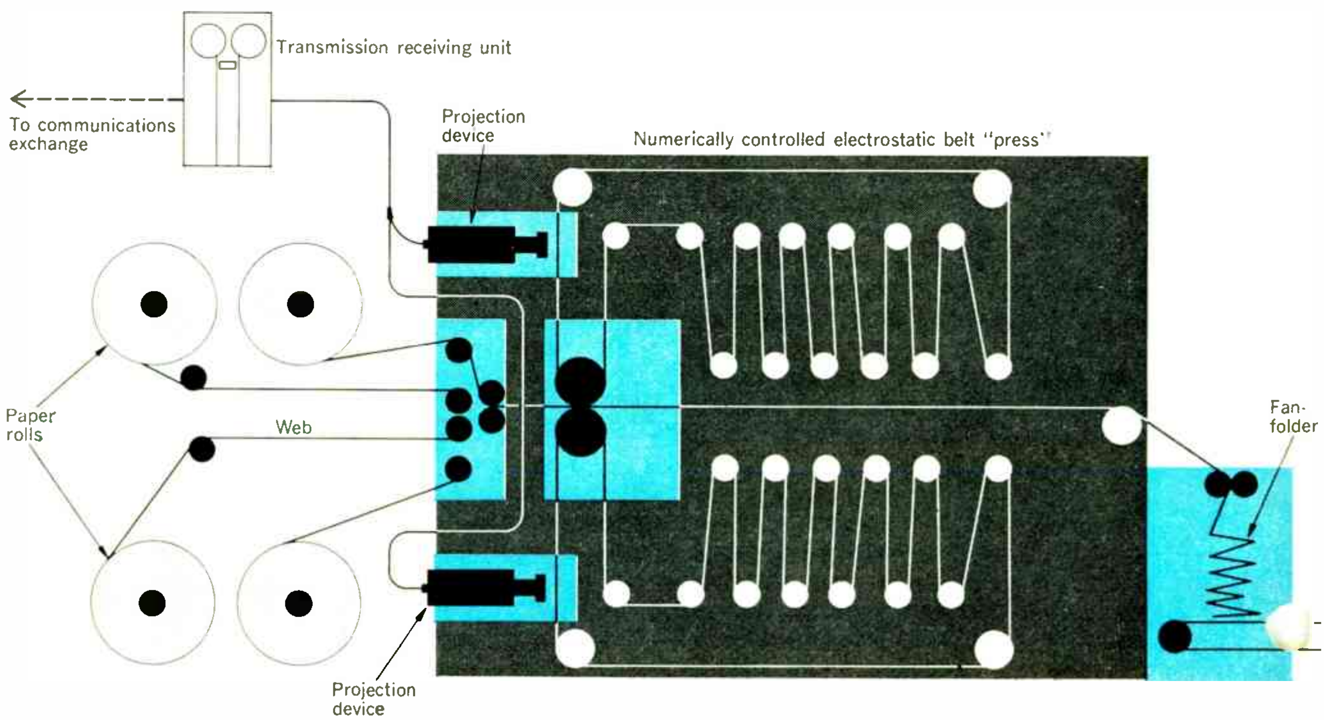


FIGURE 6. Block diagram of the future publishing process, in which electronic information distribution and automated satellite plants will be key features.

FIGURE 7. Process flow diagram of a proposed satellite plant, which will contain automated electrostatic belt presses and a numerically controlled binding line.



approach in which information is first distributed, then converted to physical form. In the block diagram, one or more publishers of magazines or books would simultaneously prepare hard copy and punched-paper-tape output for any of a number of telecommunications devices. The hard copy could be used for the publishers' records and, if desired, for facsimile copy transmission. All information, however, would be transmitted to a central communications exchange or distribution utility. This computerized center could be operated by communications firms such as AT&T, Western Union, or RCA.

When a subscriber wishes information on any subject, he will be able to obtain it in two ways, depending upon his requirements:

1. By using a Touch-Tone telephone code, an abstract printout may be obtained by mail from the communications exchange, or an instant readout response image may be flashed on a home closed-circuit television screen.

2. An order may be placed, in the subscriber's community, with a printing firm that holds a franchise to operate as the distributor of full-text printed outputs.

The local plant (a unit in a chain of automated satellite printing plants) would call the central communications exchange to obtain the required encoded material for the operation of an automatic production line, similar to that shown in Fig. 7. The book, pamphlet, magazine, or other periodical, made specifically to order, would then be quickly delivered to the local subscriber.

Full automation for satellite production. A number of experts in the P&PI feel that present-day experimental electrostatic line printers (operating similar to the principle of xerography) will soon be commercially available to produce text at speeds of more than 4000 lines per minute, and that these units will eventually replace a large portion of our present photocomposition equipment. Thus the electrostatic belt may be the key to the satellite plant of the future.

Figure 7 will give the reader some idea of the possible production flow sequences in such a plant. First, coded instructions from the data line terminal receiving unit of a computer telecommunications system would operate numerically controlled devices (shown in color) to adjust settings for size of publication, choice of paper, type faces and fonts, etc. Next, video signals would actuate projection devices to flash images onto the electrostatic belt, which would print pages sequentially. The printed web would then be fan-folded, trimmed, and cut into signatures for delivery to the adhesive binding line.

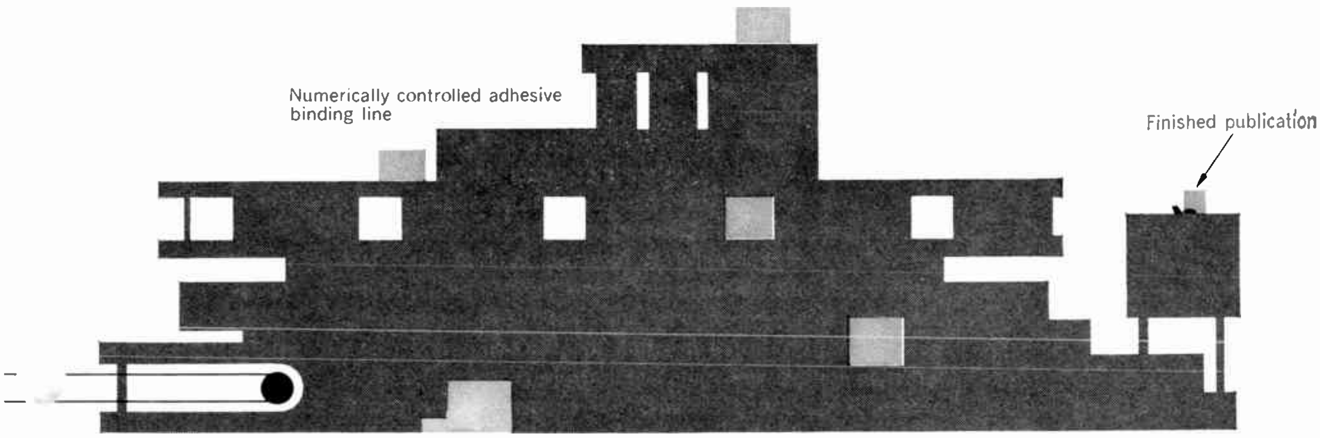
The in-line system could operate continuously without shutdown for changeovers to accommodate new formats, book sizes, and job specifications. As the final page of one book is printed, the first page of a new volume could be projected onto the electrostatic belt. Machine settings would be adjusted successively as the first copy of the next book moves along the production line.

The system just described could respond quickly to instructions for changing the length of any press run. Further, the equipment would be compact enough to be installed in a relatively small plant and, therefore, would probably be within the financial means of smaller printing plants. These salient advantages would permit local firms throughout a country to become satellite distribution outlets for printed materials in their respective consumer localities. If sufficient operating versatility could be built into the plant equipment, many types of published products—newspapers, pamphlets, magazines, catalogs, and hard-cover and paperback books—could be manufactured and distributed.

The 'total system' approach

Thinking in terms of an overall system, the satellite establishments could also become an order-filling center in which orders received during the day could be machine-tabulated. By calling the communications exchange early in the evening, the printer could request the required publication tapes from the storage bank, then produce the exact number of copies at night to fill each order. Early the next morning, the packaged lots could be delivered either to local book dealers or directly to the subscribers' homes.

Billing and payment accounting could also be integrated into each local system operation. In this way, invoice receipts could be processed at the satellite plant, and receipts reported to the publishers on a scheduled basis—perhaps during a telecommunications request for additional sets of book tapes.



A solution to present-day problems?

By this comprehensive system approach, some major problems in the conventional publishing process, such as warehousing, circulation forecasting, maintenance of retail stock inventories, and unsold returns, might be eliminated. In the publication of educational materials, timely, updated texts could be obtained within hours after the authors or editors completed the latest revised edition.

Nonprinted formats. Although the printed formats of the future may be in even greater demand than they are today, some information now contained in printed media will be translated to other media. For example, the coming electronic distribution techniques also provide instantaneous two-way communication to permit intimate man-machine interaction. Therefore, future users or subscribers will be able to query a central information storage and retrieval system and receive almost instantaneous response. And, with the advent of direct address, the user may even be able to conduct a more detailed search dialogue with the system.

Encyclopedias, directories, catalogs, listings, and other reference works requiring continuous updating may be more suited for visual readout or other nonprint media. Encyclopedias are particularly vulnerable in the present scientific and technological information explosion, because the standard set of volumes is obsolescent before it is published. In book form, it cannot be readily or quickly updated to incorporate new information.

Once separated from book format, these reference works could be made compatible with—

Microfilm and microfiche. By employing microimage compression techniques, in which many pages of text can be mounted on a small card, or microfiche, quick access to a reader or facsimile printout unit could be obtained at a central reference library by several persons simultaneously. Encyclopedias in this format could be easily and quickly updated.

By using home video tape recorders (which might be incorporated as attachments to standard television receivers) in a second-generation application of the basic technique, the necessity for keeping a conventional set of encyclopedias in the home might be eliminated. In such a system, a set of prerecorded tapes would be sent to the subscriber and, at subsequent intervals, additional tapes of new or updated material would be supplied.

But now, let's be practical...

What we have been discussing in this projection is the shape of the P&PI as it may be 15 to 20 years hence. The electrostatic belt, the concept of pressless printing by laser and holography techniques (described in the first installment of this article), etc., are radical departures from present systems and hardware. Typesetting machines, photocompositors, printing presses, and binding machinery represent very heavy capital expenditures. Therefore, until the investments in existing equipment are amortized, the P&PI will be obliged to optimize its present equipment by making the fullest use of available automation systems and devices.

As of now, a number of scientific and engineering societies have introduced photocomposition techniques in the publication of their literature. For example, the American Chemical Society has been experimenting for a number of years with the typesetting of its magazines by the use of a Photon photocompositor; and, for the past two years has employed computer-based photocomposition for producing the *Journal of Chemical Documentation*. The text of this journal, however, contains no chemical or mathematical symbols.

The American Institute of Physics has embarked upon a two-year program of developing a computer-driven photocomposition system capable of producing journals that contain complex mathematical equations. If this project succeeds, AIP will be able to produce abstract journals and indexes automatically as by-products of its journal production process. The American Mathematical Society is currently working on the computerized production of *Mathematical Reviews*, and in attempting to set complex equations by electronic character generation.

The IEEE has used photocomposition to produce STUDENT JOURNAL since 1965, and for PROCEEDINGS since 1966; but the Institute feels that computer-based photocomposition for its purposes will have to await further progress in the handling of equations in its literature. The 1968 Membership Directory (to be published soon) was produced on a computerized basis by the use of a Photon ZIP. The year-end indexes in all IEEE journals will probably be produced by computer-assisted photocomposition.

Appendix

Glossary of P&PI terms. To assist the reader who is now puzzled by some of the jargon of the trade, the following may be timely:

1. *Build-in.* An operation in which machines apply heat and pressure to the covers and joint of a newly cased volume to ensure that the glue is dried and the moisture removed from the pages before jacketing and shipment.

2. *Casing-in.* A machine process that applies the cover or case to the sewn pages of the book.

3. *End sheet tip.* The process of inserting a folded sheet (in a book being bound), one leaf of which forms the paste-down on the cover and the other forms the flyleaf.

4. *Fanfold.* A collection of similar sheets, leaves, or signatures where multiple copies are required.

5. *Gather, or gatherer.* A machine that gathers, or collates, sheets, leaves, or signatures for binding.

6. *Glue-off.* The application of glue to the spine of a book during the binding process.

7. *Signature, or form.* A printed sheet, which, when folded, becomes a unit of a book or magazine.

8. *Smash, or smasher.* A machine that compresses folded book sections, or assembled books, to give firmness and uniform bulk, and eliminate the tendency to produce a wedge-shaped back caused by the threads used in sewing the binding.

Figure 1 by courtesy of the National Library of Medicine.

REFERENCES

1. Olerich, R. B., and Anderson, B. D., "Within two years: complete automation for web offset?" *Book Production Industry*, vol. 43, pp. 58–62, Oct. 1967.
2. Dunbar, R. V., "Web press design: New principles for precision control," *Book Production Industry*, vol. 43, pp. 62–64, Oct. 1967.
3. De Florez, P., "Automatic set-up: vast opportunity for savings," *Book Production Industry*, vol. 43, pp. 65–69, Oct. 1967.
4. Doebler, P. D., and Tewlow, J. S., "The shape of the future: knowledge—its distribution and display," *Book Production Industry*, vol. 43, pp. 42–49, Jan. 1967.

Making integrated electronics technology work

The success story of integrated electronics is that of digital circuits; it is in the fields of microwaves and consumer products that the challenges to the technology still lie

Robert N. Noyce
Fairchild Camera and Instrument Corporation

The keynote speaker at the recent International Solid State Circuits Conference implements the basic theme of the conference by presenting an overview of the success achieved in the field of integrated electronics—and the problems remaining to be solved. He concentrates on three areas—digital, microwave, and consumer applications—because these best illustrate the successful implementation of the technology in the first case, and its possibilities, and problems, in the others.

In past years, this solid-state conference was small and specialized; participants explored such "weighty" matters as the applications of audio-frequency transistors. Well, by this field's standards that was eons ago, and solid-state electronics has now made its way into all aspects of electronics. We have only to thumb through this year's program to realize the incredible range that the field now covers. There are sessions on microwave circuits, consumer electronics, memory techniques, linear amplifiers, digital techniques, large-scale integration, and computer-aided design, as well as a catch-all session on "new techniques and devices."

To provide a guide to all of these frontiers is nearly impossible so I shall confine myself to integrated electronics, an area that touches on nearly all the other topics. It is a field with a great need for new devices and techniques, and one that, at the same time, is poised to make major inroads in microwave and consumer applications. Within the broader category of integrated electronics, I have chosen to discuss microwave integrated circuits, consumer integrated circuits, and digital integrated circuits because they best illustrate the successful implementation of integrated electronics in one case, and the promise (and disappointment) of their application in the other two.

The success story of integrated electronics, of course, is that of digital circuits, and in this instance we can easily see the performance and cost advantages over other approaches. In the United States last year sales of digital integrated circuits totaled some $200 million; this year, digital circuit sales should approach $300 million. I mention these figures because it is important to recognize that money spent on technology must seek the highest return. Because the market is large, the investments in satisfying the market also have been large. This factor has been a major contribution to the rapid development of the digital integrated circuit market. However, there are also other reasons for this development. For one, a potentially large savings in cost existed. Second, the performance of the digital circuits was competitive with that of the early integrated circuits, and in a short time became superior. Third, there was a potentially great improvement in reliability, which we have since realized. And last, there existed a ready market that was growing and accommodating changes rapidly. In retrospect we can say that, for digital applications at least, the integrated circuit was simply "the better mousetrap."

Microwave applications

In contrast to the digital applications, some of the elements of motivation for the use of integrated circuits have been missing in microwave and consumer electronics. For microwave applications, it appears that the cost advantages of integrated electronics will be substantial but that adequate performance levels will be difficult to achieve.

Solid-state microwave circuits now appear to be at the same state of development that very-high-frequency circuits were at about ten years ago. Individual, highly selective, specialized semiconductor devices are now available, which, with adequate coaxing, are able to perform the desired functions at a marginal level. But production yields on these individual devices are extremely low and manufacturers do not look with favor on compounding their problems with multiple devices at this time. In spite of this, integrated circuits do hold great promise for microwaves. Photoengraving eventually must provide a more economical way to achieve high precision than is possible by conventional machining techniques. By shrinking physical dimensions an order of magnitude or more, and eliminating package and wiring reactances, lumped-constant circuit design techniques can be extended into the microwave region. The most exciting happening in the field today is the emergence of new devices that already are being incorporated into microwave systems, namely, Gunn-effect devices, avalanche or Read diodes, and transistors displaying reasonable gain and power output in the low-microwave regions.

Integration of these new devices into monolithic structures still seems some time away. Working with low-yield specialized devices, the hybrid approach makes more sense from an economic viewpoint. After the hybrids, we will see fully developed monolithic structures, perhaps built on a substrate of semi-insulating gallium arsenide.

Consumer electronics provides a perfect example of a market facing an economic rather than a performance obstacle. The market is large and waiting, and solid-state electronics offers everything the market wants—everything, that is, except an attractive price tag. Here the problem is to develop new techniques and innovations that will bring costs down to a level that will make the applications worthwhile.

A danger that we encounter is underrating the progress-mindedness of the consumer market. This market will respond to change rapidly, as was evidenced by the switch from black-and-white to color television sets, and in the annual model changes found in most consumer goods. However, the consumer electronics manufacturer must have some motivation to make such changes. No great performance advantage will accrue to him today from converting to integrated circuits, and the reliability of his present equipment is adequate—but the market is very sensitive to cost and it will respond rapidly to advantage in this area.

I. Economic models

$$\text{Chip cost} = \frac{\text{Statistical silicon cost} \times \text{chip area}}{\text{Chip yield} \times \text{package yield}}$$

$$\text{Package cost} = \frac{\text{Technology const.} + \text{complexity const.}}{(G^{1/2})\ \text{package yield}}$$

$$\text{Chip yield} = [(1+0.57\,GDo)(1+0.285\,GDo)(1+0.143\,GDo)]^{-1}$$

Package yield = Constant (time dependent)

G = Number of gates

Do = Average defect density per gate

Of course, cost is not the only barrier to success in the consumer electronics market, although it is the most important one. There are some technical problems to solve as well; ultrahigh-frequency tuners cannot be produced in integrated form today nor are high-Q frequency-selective circuits yet available. Still, the time is approaching when integrated circuits will provide a significant cost advantage to the consumer goods manufacturer. I believe that if we look at the cost history of any class of integrated circuits we can predict with confidence that costs will become low enough to offer real incentive for use in consumer products.

At present, integrated circuits are used in small numbers in the simplest (for integrated circuits, that is) parts of a television set. The set's front end and output still present problems, but we should be able to solve these by utilizing the technology developed for digital circuits, and the next two or three years should see a rapid expansion in the use of integrated circuits for microwave applications.

LSI enters the picture

The kind of expansion and sophistication yet to come for microwave and consumer devices is now commonplace in digital circuitry. The success of digital circuits has caused the industry to address itself to the whole new set of problems and opportunities found at higher levels of integration, namely, large-scale integration (LSI). There is still work to be done, of course, toward improving the cost and performance of integrated circuits as we know them today—but LSI is where the action is. This year will see heavy sales of medium-scale integration, or complex integrated circuit devices, and this represents the proverbial "foot in the door." These devices will introduce us formally to the problems and benefits of LSI.

The question to be answered in LSI is not "Can it be made to work?" but rather, "Can it be effectively used?" The positive answer to this question depends on the system designer as much as on the manufacturer of the LSI device. And I am sure most of you know that LSI will pose new problems in the organization of our industry, will call for different interfaces between customer and vendor, and will point out a need for better definition and specifications of test programs.

Why does anyone want to open this "Pandora's box"? The answer is simply that we think that the achievement of higher levels of integration will bring great advantages. The foremost of these is lower cost; but there will also be improvements in performance and reliability. To achieve better performance we can take advantage of smaller dimensions, which lead to less propagation delay because of the shorter distances, and to fewer problems of wiring and package reactances. Reliability is improved by having

II. Data for models of Table I

Year	Gate Area, μm^2	Defect Density, μm^{-2}	Silicon Cost, $/cm^2	Package Formula
1968	130 000	0.203	3.10	$\frac{\$0.25 + \$0.10\, G^{1/2}}{0.7}$
1970	65 000	0.102	2.32	$\frac{\$0.15 + \$0.05\, G^{1/2}}{0.8}$
1974	32 000	0.05	1.55	$\frac{\$1.10 + \$0.04\, G^{1/2}}{0.85}$

fewer connections in the system. Again, however, cost is the primary motivation.

The cost advantages of LSI are further defined in a paper by R. B. Seeds,[1] who did a quantitative prediction of the production cost versus complexity of arrays of gates, extrapolating today's laboratory achievements to a production environment for 1970, and slight improvements beyond that to 1974. The basic economic model he used is straightforward (Table I). Production cost is the sum of the chip cost and the package cost, both divided by the yield. The chip cost is the processing cost of the silicon wafer multiplied by the chip area, divided by the yield of the chip at first sort and at final test. The cost of the package has been assumed to be a constant plus factor related to the number of leads on the package, here proportional to the square root of the number of gates in the chip. Package yield is assumed to be constant, but time dependent. The formula given for the chip yield results from an investigation of the area correlation between defects in ICs in manufacture, and is well founded both empirically and theoretically.

Using this model and the data shown in Table II, production costs versus complexity of the array were calculated. The data shown for the year 1968 represent today's production technology. Data shown for 1970 assume a 2:1 improvement in gate area, defect density, and package costs. Another factor of 2 is predicted for 1974.

The costs calculated from these assumptions are shown in Fig. 1—the 1968 minimum occurring at about a 20-gate array costing 9 cents per gate. The drops in 1970 and 1974 are dramatic: 60 gates at 2 cents each and 200 gates at 0.7 cent each. Actually, I realize this prediction may be too conservative.[2] If we assume gate areas of 6500 μm^2, the minimum costs per gate at 1968 technology occurs at arrays of about 1000 gates with a cost of less than 0.7 cent per gate.

Although one may argue with the specific assumptions made, this does represent an attempt to calculate costs in a straightforward manner from what we know today and what we can predict for tomorrow. And the conclusion is clear: Cost reduction of gates of 10:1 is within the realm of feasibility if we can make use of large complex arrays, and gate costs become negligible as compared with other costs. It is the decrease in costs in other areas that the system designer largely controls. The other costs incurred in going to LSI are the engineering design and start-up costs, and these include artwork layout and generation, test sequence generation, and setup of production control. These costs, of course, are amortized over an entire production run, but easily may be the dominant cost factor in the future.

One study has shown, for example, that the layout cost per gate is about $10. If an entire production run of arrays amounted to 1000 pieces, amortized design costs would be larger than the gate costs for these arrays. But more important than costs per gate, what is the lowest cost for the system? This question is difficult to answer.

In considering the total system costs, the costs of the silicon processing quickly drop out of the picture. The costs that must be considered include the costs of the LSI package and silicon chip as well as the costs of the printed circuit board, the contacts, and the back panel wiring, or "printed circuit mother board," in addition to the total design costs, which must be amortized over the production run.

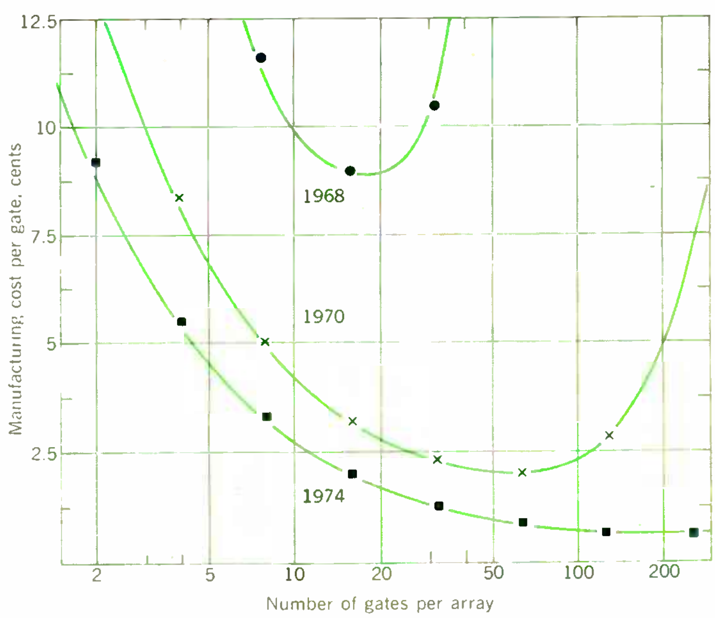
In considering the cost of layout and artwork generation, the cost of layout of the metal interconnection pattern is approximately the same as the layout today of a printed circuit board for integrated circuits. Assuming that equivalent levels of design aids are available to the vendors of tomorrow's arrays as are available to today's integrated circuit user, the layout cost will be the same. There may be some argument about which part of the industry should do this job, but the total product cost will not be affected. You may ask, "Since total cost has not increased in going to a higher level of integration, why worry about it?" The answer, of course, is that the active elements have become so inexpensive that other costs become dominant. The design costs are vulnerable to reduction by computer-aided design. For large enough production runs, layout costs can be comparable to the cost per gate that has been projected.

Some other costs

In cost projections, Dr. Seeds has assumed that the number of pins on the package is proportional to the square root of the number of gates in the LSI array. Such an assumption may be too optimistic. A recent study of system architecture for LSI[3] indicates significantly higher ratios than this; see Fig. 2.

For 100 gates, the number of pins required was between

FIGURE 1. Calculation of costs based upon data of Table II.



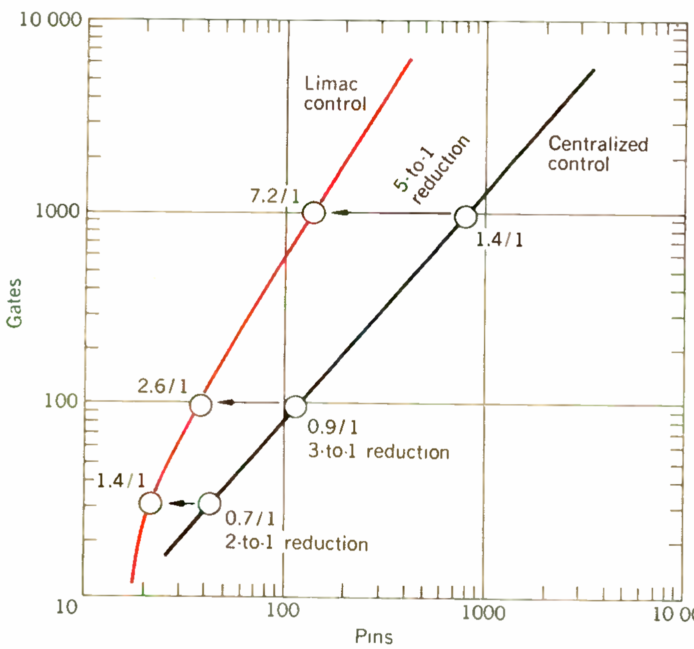
40 and 110. For gate arrays of 1100 more than 100 pins are needed for the architecture proposed in the cited study, or 800 pins for conventional design. Let me remind you that, for the small arrays, the package cost is dominant, and for the large arrays package costs are finally overcome by increasing silicon yield costs. If the number of pins required of the package rises proportionally to the number of gates included, this minimum cost point moves to larger arrays as packaging costs increase.

At some point in the system there will be pluggable contacts, either at the array or for an assemblage of arrays. At 10 to 25 cents per pluggable contact, this cost becomes dominant unless the number can be reduced. We are led then to the fact that, with the improvements in costs of gates that we foresee, the cost of the active elements essentially drops out of the total cost considerations and other hardware costs take over. This leads to a new compromise in optimal system design, in which three new rules become important:

1. One must design for a minimum number of interconnections and contacts between the various functional units or arrays.
2. The optimal use of active elements no longer is desirable if it entails higher design or hardware costs.
3. As a corollary, the partial use of a functional unit may be desirable to decrease the numbers of diverse parts.

The problems represented by these rules present a challenge to the system designer—one that I think will push him to further innovations. The problems I have been discussing are not present in highly regular integrated arrays such as registers, memories, and the like. Here the number of leads can be held to a small number, and one particular design can satisfy many requirements. For that reason, I believe that active memories using semiconductor flip-flops will challenge other techniques in providing a high-speed random-access memory.

FIGURE 2. Cost projection based upon study by Beelitz *et al.*[3] of system architecture for LSI.



Conclusion

In the foregoing, I have touched on some of the problems involved in making integrated electronics work. You will see that consumer product manufacturers will convert to integrated circuits only when the semiconductor industry demonstrates that they cannot afford not to. The challenge is to bring prices down to the point where the consumer goods manufacturer sees a way to improve his product and at the same time maintain or increase his profit levels. When our innovations make attractive prices possible, the consumer product will go the way of the computer, to integrated circuits—and I believe this conversion is near at hand.

The microwave field, on the other hand, presents yet another frontier. We cannot afford the luxury of discussing pricing for microwave integrated circuits. First, we have to demonstrate to the potential buyer that solid-state components can do the required job. Next, we will see the increasing use of these components in hybrid circuits, and finally, perhaps, in monolithic structures. It is my belief that integrated circuits eventually will dominate microwave electronics. How they will do it is the question.

Perhaps the most portentous topic of this conference is LSI, not only because it presents momentous technical and economic challenges, but because it will have a profound effect on the business organization aspects of both supplier and user companies. Like its parent, the integrated circuit, LSI is a problem not only to the scientist and the engineer, but to the manager as well. It means a long, hard look at organization, interface, cost effectiveness, and rate of return on invested capital.

The manager—the businessman—must come to know the implications of LSI even as will his engineer. He must know, as do his engineers, that LSI is a functional unit, not a device. He must know, as do his engineers, that in order to make LSI work for him, he must help create the climate, and the means, for effective continuous communication between designer and user.

The LSI designer will find his job largely determining what can be done within certain numbers of pins. He must develop a new philosophy of design. Rather than deciding merely just how much logic is necessary, he will have to determine how much surrounding equipment is necessary—with cost a critical factor. For this reason, the sessions on computer-aided design and computer organization, which, at first glance, might have been considered out of place at a solid-state conference in past years, may prove to be pivotal this year.

The challenges to us are monumental—and numerous. But if there is one trait that has characterized the electronic industry through the years, it has been its eagerness to accept challenges, to try new things, to turn the seemingly impossible into routine.

Essentially full text of the keynote address presented at the International Solid State Circuits Conference, Philadelphia, Pa., February 14, 1968.

REFERENCES

1. Seeds, R. B., "Yield and cost analysis of bipolar LSI," presented at 1967 Fall Electron Devices Meeting.

2. Murphy, B. T., and Glinski, V. J., "Transistor-transistor logic with high packing density and optimum performance at high inverse gain," presented at 1968 International Solid-State Circuits Conference.

3. Beelitz, H. R., Levy, S. Y., Linhardt, R. J., and Miller, H. S., "System architecture for large-scale integration," *Proc. Fall Joint Computer Conf.*, AFIPS, vol. 31, p. 193, 1967.

Electrochemical vehicle power plants

Work is currently under way on high-power and high-energy-density batteries that ultimately may serve as acceptable electrochemical energy storage devices for use in propulsion of electric vehicles

D. A. J. Swinkels *General Motors Corporation*

The requirement of high energy density for vehicle applications leads to the selection of battery systems made up of light, highly reactive metals found in the upper-left-hand corner of the periodic table for one electrode and the light oxidizers found in the top-right-hand corner of the periodic table for the other electrode. Many of these reactive materials will react with water and hence water-based electrolytes cannot be used in such systems. Electric vehicles also require the battery to have a high power density, which further limits the use of water-based electrolytes. The parameters relevant to achieving high-energy, high-power batteries are discussed in general terms and several systems presently being studied are reviewed.

Many books and articles[1-3] have been written about batteries. These generally give a rather detailed description of well-developed and usually commercially available systems, which often originated in the days when electrochemistry was more an art than a science. For example, who in his right mind would try to make a battery using lead or lead compounds for both electrodes, when lead is one of the heaviest elements in the periodic table? Yet the lead–acid battery has been in use for more than a hundred years and even today most of the rechargeable batteries manufactured are lead–acid batteries.

Around the turn of the century more electric cars were produced than gasoline-engine cars, but battery development remained stagnant whereas internal combustion engines were improved, and soon the limited energy and power capabilities of batteries made a museum piece of what had been a competitive item of machinery. However, current interest in the economics of gasoline versus electric vehicles, and concern about the constant emission of tons of foreign materials into the air we breathe, have resulted in a worldwide re-examination of the technical feasibility of electric vehicles.

In the following, we shall examine the properties of batteries with regard to the basis on which they are selected for various applications, and take a closer look at several systems presently under development for use as the prime power source of electric vehicles.

It is probably useful at this point to destroy the mental picture one has of a battery as consisting of two dissimilar plates placed in an aqueous solution. Although this is the form that most conventional batteries have taken, it is unlikely that this type of electric power plant will drive anything but a low-performance, limited-range vehicle. To achieve performance and range approaching those of today's gasoline-driven cars, electric power plants with an energy capacity in excess of 220 Wh/kg and a power capacity in excess of 130 W/kg will be required.[4] No battery using an aqueous (i.e., water-based) electrolyte can meet both of these requirements, although the zinc–air system comes close.[5]

The electrochemical converters have been divided, according to their function,[6] into *substance producers*, in which electric energy and materials are fed in and a new material is produced, e.g., $2\ NaCl + \text{energy} \rightarrow 2\ NaOH + Cl_2 + H_2$, and *energy producers*, in which materials are fed in and energy plus waste materials are produced, e.g., $2\ H_2 + O_2 \rightarrow \text{energy} + 2\ H_2O$.

Neither the substance producers nor the energy producers (i.e., fuel cells and primary batteries) are of interest here but rather a combination of the two with the primary function of storing energy. We shall therefore define a battery as a black box into which we can put a given amount of electric energy, store it in the form of chemical energy, and retrieve a smaller amount of electric energy at a later time. The term "electrochemical energy storer" has been suggested,[6] which is clearly much more descriptive of the function of a secondary battery.

Relevant qualities

The main qualities of electrochemical energy storage devices that determine their usefulness are the energy capacity—the amount of energy that can be stored per unit weight or sometimes per unit volume; the power capacity—the rate at which the stored energy can be taken out and restored; and the energy efficiency—the amount of useful energy that can be taken out at the required rate divided by the amount of energy that had to be put in at a given rate. These quantities decide the technical feasibility of using a given battery for a particular job. To determine its economic competitiveness with other approaches, the cycle life (the number of times the system can be usefully charged and discharged) and the inherent cost of the materials used in the system are, of course, of prime importance.

Some of these quantities are discussed in detail below. The energy capacity of a battery is usually expressed in watthours per kilogram (watthours per pound). If we accept the Pb–acid battery as typical present-day technology, then about 22 Wh/kg is the energy capacity of today's battery technology. To achieve high energy per unit weight of battery it is necessary to choose reactants of high reactivity—from opposite ends of the electronegativity scale—so that a high open-circuit voltage (OCV) is obtained. At the same time the equivalent weight of the reactants must be kept to a minimum to keep the battery weight low. These considerations lead naturally to systems made up of elements from the top left corner of the periodic table (H_2, Li, Na, Mg, Al) coupled with elements from the top right corner (F_2, Cl_2, O_2, S) or with compounds containing a large proportion of these elements, such as $CuCl_2$ and NiF_2. The electrolyte for such high OCV systems must have a decomposition voltage equal to or greater than the OCV and must not react with either reactant.

Aqueous electrolytes often will not be satisfactory since they react chemically with the high-energy metals. Organic solvents, fused salts, and ionically conductive solid membranes are all under consideration for use as electrolytes in various types of high-energy batteries. If the reaction product of the electrochemical reaction can serve as the electrolyte rather than being dissolved in another fluid, a simpler and potentially lighter system will result. This often occurs with fused-salt electrolytes.

Theoretical energy densities can be calculated for any proposed combination of reactants by multiplying the open-circuit voltage by the Faraday constant and dividing by the equivalent weight of the reactants for the reactions shown. Care has to be taken to include all reactants in such calculations since the electrolyte may participate in the reaction. Table I shows some theoretical energy densities for high-energy-density systems now under development and for some of the more conventional batteries. If these energy storage capabilities could be achieved in practice, then even the Pb–acid battery would go a long way toward meeting electric car requirements. However, only 15–40 percent of the theoretical energy storage capacity actually is achieved because reactants alone do not make a battery. Leads, containers, and electrolyte add to the total weight without contributing to the energy stored. In addition, the energy stored in a battery cannot be withdrawn with 100 percent efficiency since some is always lost as heat at practical rates of discharge.

The power capacity of a battery is to a large extent determined by the ratio of the OCV to the resistance of the electrolyte.[7] The higher the OCV and the lower the electrolyte resistance, the higher the power density that can be attained, which leads to the selection of very active electrode materials to obtain the high OCV and to fused-salt electrolytes because of their low resistivities. Typical resistivities are 0.1–1.0 ohm-cm for fused salts, 1–10 ohm-cm for aqueous electrolytes, and 100 ohm-cm and greater for organic electrolytes and solid electrolytes. Electrolyte resistance is given by $\rho l/A$ where ρ = electrolyte resistivity, l = electrolyte thickness, and A = electrode area, so that low resistances can be obtained even with high-resistivity electrolytes if they can be made sufficiently thin. Glass can be an Na^+ ion conductor with a resistivity of about 100 ohm-cm at 300°C. However, by making the glass membrane thin (say 10^{-3} cm) and using a large area the internal resistance of a battery can still be kept low.

I. Conventional and proposed battery systems

Battery	Cell Reaction (discharge →)	OCV	Temperature	Electrolyte	Theoretical Energy Capacity, Wh/kg	Practical Energy Capacity, Wh/kg
Lead–acid	$Pb + PbO_2 + 2\ H_2SO_4 \rightleftarrows 2\ PbSO_4 + 2\ H_2O$	2.1	Ambient	Aqueous H_2SO_4	176	18–33
Silver–zinc	$AgO + Zn \rightleftarrows ZnO + Ag$	1.8	Ambient	Aqueous KOH	492	65–130
Zinc–air	$Zn + 1/2\ O_2 \rightleftarrows ZnO$	1.6	Ambient	Aqueous KOH	1056	110–170
Lithium–NiF_2	$2\ Li + NiF_2 \rightleftarrows 2\ LiF + Ni$	2.8	Ambient	KPF_6 in propylene carbonate	1364	110–220 (at low rate)
Lithium–chlorine	$Li + 1/2\ Cl_2 \rightleftarrows LiCl$	3.5	650°C	LiCl	2178	330–440 (estimated)
Sodium–sulfur	$2\ Na + 3\ S \rightleftarrows Na_2S_3$	2.1	300°C	β-Al_2O_3 + sodium sulfides	792	220–330 (estimated)
Lithium–tellurium	$Li \rightleftarrows Li\ (Te)$	1.9	480°C	LiF-LiCl-LiI	572	130–200 (estimated)

Finally, the voltage losses associated with the electrode reactions should be minimized to obtain high power densities. To accomplish this, the electrode reactions should be simple, that is, they should involve a low number of electrons. Also, an increase in temperature increases reaction rates and hence reduces the voltage loss at a given current density.

In view of these considerations it is clear why the many high-energy batteries being developed using organic electrolytes are not high-power batteries. Organic electrolytes are generally limited to low-temperature operation (< 100°C) and the electrolyte resistivities are typically 100 ohm-cm or more.

The energy efficiency of a rechargeable battery, defined as the ratio of electric energy taken out during discharge to electric energy put in during the charging process, can be expressed as the product of a current efficiency and a voltage efficiency. The voltage efficiency is determined by the rate of charge and discharge and the electrode kinetics of the particular system. The higher the rate of charge or discharge, the lower the voltage efficiency. The current efficiency, on the other hand, can increase or decrease as the rate of charge or discharge is increased, depending on what causes the coulombic loss. If the coulombic loss is due to a side reaction that starts to occur when the voltage reaches a certain value (e.g., gassing in aqueous electrolyte systems) then a lower rate of charge will result in a higher coulombic efficiency. If, however, the coulombic loss is due to reactant solubility in the electrolyte and chemical combination, then the cell self-discharges at essentially a constant rate and hence the higher the average rate of charge and discharge, the higher the coulombic efficiency.

Self-discharge may also be caused by slow reaction with the solvent, by an intermediate that is alternately being reduced and oxidized at the negative and positive electrodes, respectively, and by electronic conductivity in the electrolyte. The last not only affects the coulombic efficiency but also the voltage efficiency and the OCV since it is equivalent to putting a constant external drain on the cell.[9] Thus V, the measured OCV, becomes

$$V = \frac{V_0}{1 + R_i/R_e} \qquad (1)$$

where V_0 = thermodynamic OCV, R_i = ionic resistance of the cell, and R_e = electronic resistance.

Energy losses, whether due to electronic conductivity, reactant solubility, or to energy required to keep the cell hot during standby periods, can be expressed as a self-discharge current. The practical coulombic or current efficiency then becomes

$$\text{Current efficiency} = \frac{I_D t_D}{I_c t_c} = \frac{I_D t_D}{I_D t_D + I_{SD} t_{SD}} \qquad (2)$$

where I_D = average discharge current, t_D = discharge time, I_c, t_c = charge current and time, and I_{SD}, t_{SD} = self-discharge current and time.

If we define the "use factor" for a particular battery application as the fraction of time during which it supplies useful power, then we can relate the current efficiency to the self-discharge current and the use factor. It is clear that the lower the use factor, the lower the self-discharge rate we can tolerate, since self-discharge usually proceeds all the time. We then have use factor $= t_D/t_{SD}$, and combining this with Eq. (2), we get

$$\text{Current efficiency} = \frac{1}{1 + \dfrac{I_{SD}}{I_D}\,\dfrac{1}{\text{use factor}}} \qquad (3)$$

Figure 1 shows a plot of Eq. (3) for various use factors. A use factor of 1.0 can only be achieved in a primary battery or fuel cell, which, after activation, is discharged continuously. Under these conditions quite high self-discharge rates can be sustained without serious effects on the current efficiency. The effective self-discharge may also be the result of heat requirements to keep the system hot, or power requirements for auxiliary controls. In the case of solid electrolyte systems, self-discharge can result from an electronic conductivity in the electrolyte. It is not always desirable to select the electrolyte or the temperature with the lowest electronic conductivity. Rather, the optimum system is a complex function of the total conductivity and the ion transference number.[9]

The lower use factors apply to secondary batteries that must be recharged regularly. A battery in industrial use might be discharged eight hours a day and recharged every night for a use factor of 0.33, whereas a battery for a personal electric vehicle that is driven about an hour a day would have a use factor of about 0.04. In the latter case the effective self-discharge current must clearly be kept low to give reasonable current efficiencies.

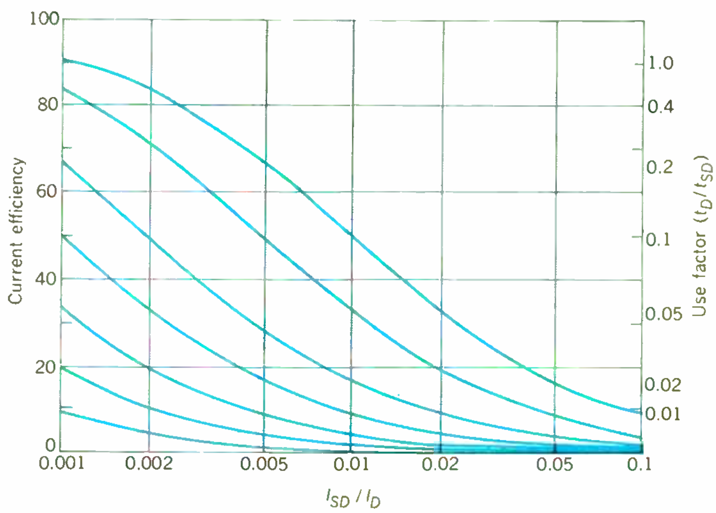
Since the self-discharge current for a given system is generally directly proportional to the electrode area, it appears advantageous to reduce the electrode area to a minimum. However, this increases the charge and discharge current densities with a loss in voltage efficiency.

The voltage efficiency is defined as the ratio of the average discharge voltage to the average charge voltage. The charge and discharge voltage are related to the charge and discharge current densities through the usual electrokinetic equations.[6,8,10] These are summarized here.

$$V_D = \text{OCV} - I\Sigma R - \eta_{c,a} - \eta_{c,c} - \eta_{a,a} - \eta_{a,c} \qquad (4)$$

where V_D = discharge voltage; OCV = open-circuit voltage; I = current; ΣR = sum of electrolyte, electrode,

FIGURE 1. Current efficiency as a function of self-discharge and use factor; Eq. (3).



and lead resistances; $\eta_{c,a}$, $\eta_{c,c}$ = concentration polarization at the anode and at the cathode; and $\eta_{a,a}$, $\eta_{a,c}$ = activation polarization at the anode and at the cathode. For simplicity, all polarizations are here taken as positive quantities.

The concentration polarization at each electrode is due to the concentration gradients necessary to move reactants from the supply zone to the region where electron transfer can occur and to move products from the electron transfer region to the product storage volume. The relationship between concentration polarization η_c and current density i is of the form

$$\eta_c = \frac{2.303RT}{nF} \log \left(\frac{i_L}{i_L - i}\right) \qquad (5)$$

where R = gas constant = 8.314 joules/degree mole; T = temperature, °K; n = number of electrons; and F = Faraday = 96 500 coulombs/mole.

The limiting current density i_L is the maximum current density that can be obtained. It is typically given by

$$i_L = \frac{nFDC}{x} \qquad (6)$$

where C = concentration of the rate-limiting reactant in the supply zone, D = diffusion coefficient of the rate-limiting specie, and x = a characteristic distance equal to or simply related to the distance between the supply zone and the electron transfer region.

It is clear from Eq. (5) that in order to minimize η_c one must maximize i_L. This, in turn, means maximizing D and C and minimizing x. Generally, little can be done about increasing D except by increasing the temperature whereas the concentration C is usually determined by solubility, which may be increased by increasing the temperature or, in the case of a gaseous reactant, by increasing the pressure. It is in the diffusion distance x that one often has a good deal of control over the limiting current density. In a porous gas electrode x is usually related to the particle size or the pore size[11] and to the wetting characteristics of the electrode, whereas in solution x can be reduced by convection or stirring.

The activation polarization (η_a) is the polarization more directly related to the electron transfer processes although surface diffusion of reactant or intermediates may be involved. It is related to the current density in the form

$$i = i_o \left\{\exp \frac{\alpha nF\eta_a}{RT} - \exp \frac{-(1-\alpha)nF\eta_a}{RT}\right\} \qquad (7)$$

where i_o = exchange current density and α = transfer coefficient.

There is no simple way of expressing η_a in terms of the other parameters except in special cases. When α equals the often-observed value of 0.5, we have

$$\eta_a = \frac{RT}{0.5nF} \sinh^{-1} \left(\frac{i}{2i_o}\right) \qquad (8)$$

An approximate relationship can be written for polarization values less than RT/nF when

$$\eta_a = \frac{RT}{nF}\left(\frac{i}{i_o}\right) \qquad (9)$$

whereas for polarization larger than RT/nF we have the Tafel equation

$$\eta_a = \frac{2.303RT}{\alpha nF} \log \left(\frac{i}{i_o}\right) \qquad (10)$$

Figures 2 and 3 show plots of these equations to indicate their ranges of validity. It is clear that to obtain high rates of charge and discharge with low-activation polarization losses one must look for systems with high exchange current densities (i_o). This, in turn, means operating at elevated temperatures using fused salts where exchange current densities of 10^0 to 10^2 are typical for simple one-electron reactions. On the other hand, in aqueous solutions at ambient temperatures exchange current densities of 10^{-12} to 10^{-5} are common for reactions requiring two or more electrons for the overall reaction.

Finally, in this discussion of the various forms of energy losses it is clear that the simple IR losses must be kept to a minimum. Since fused salts generally have resistivities that are an order of magnitude lower than those of aqueous electrolytes and several orders of magnitude lower than organic electrolytes, the fused-salt electrolytes naturally lead to higher power and energy systems. This, of course, also means elevated temperatures

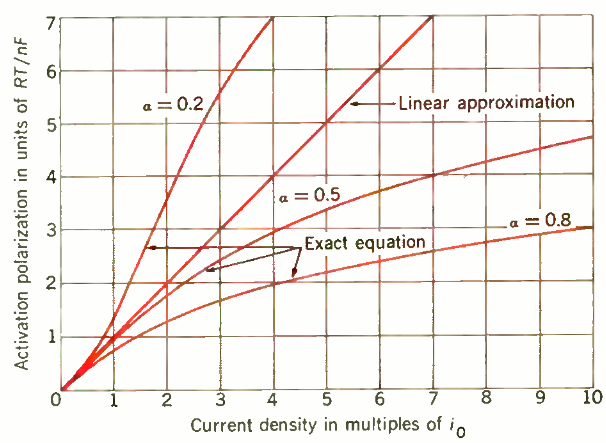
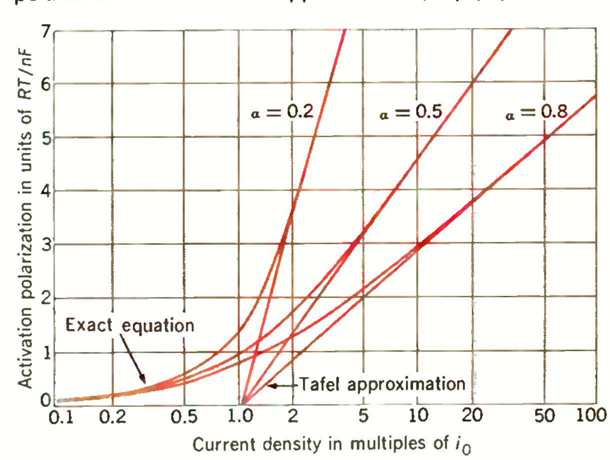


FIGURE 2. Comparison of exact equation for activation polarization and the linear approximation; Eq. (9).

FIGURE 3. Comparison of the exact equation for activation polarization and the Tafel approximation; Eq. (10).



of operation with the resulting benefits of high exchange currents and high diffusion rates, which both lead to low polarization losses.

The Li-Cl_2 battery

The Li-Cl_2 battery has many of the desirable characteristics outlined in the previous sections.[12] The high open-circuit voltage (3.467 volts at 650°C and 1 atm) together with the low equivalent weight, leads to a theoretical energy capacity of about 2.2 kWh/kg. The achievement of but 15 percent of this theoretical energy density in a practical system would suffice for electric car requirements. A further advantage of the Li-Cl_2 system lies in the fact that the reaction product LiCl is also the electrolyte so that no diffusion limitation due to the removal of reaction products can exist. In addition, fused LiCl has the lowest known resistivity of any fused salt at 650°C (ρ = 0.17 ohm-cm). The electrochemical reaction at the lithium electrode is extremely simple since it consists of a single electron transfer from an atom in a pure liquid metal with the formation of a monovalent ion in a pure fused salt. It therefore is not surprising that no polarization can be detected on a lithium electrode at current densities up to 40 A/cm^2.[12] To achieve mechanical stability in this liquid-liquid interface electrode a wick electrode has been developed.[13] Figure 4 shows a sketch of an advanced Li-Cl_2 cell design. The lithium normally floats on the denser LiCl. Since liquid lithium wets most solid metals preferentially to the LiCl, a porous stainless-steel felt structure will serve to wick lithium from a storage region to the electrode region in the way an asbestos wick acts in an oil lamp.

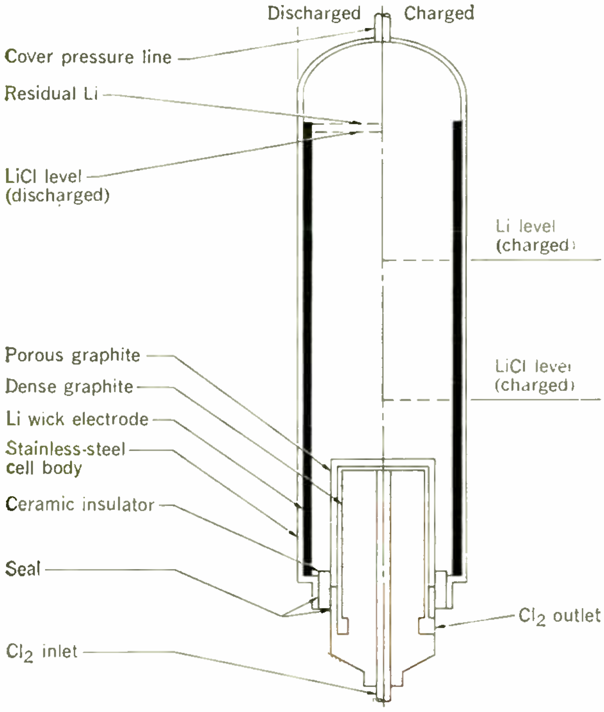


FIGURE 4. Possible configuration of an advanced lithium-chlorine cell.

The driving force for the wicking action is the capillary action of the lithium in the porous metal. This driving force, ΔP, is given by

$$\Delta P = \frac{2\gamma \cos\theta}{r} - gh\Delta\rho \qquad (11)$$

where γ = Li-LiCl interfacial tension, θ = contact angle, r = pore radius, g = acceleration due to gravity, h = wicking depth, and $\Delta\rho$ = difference in density between LiCl and Li. The resulting velocity v of lithium flowing through a cylindrical pore is given by

$$v = \frac{r^2\Delta p}{8\eta L} \qquad (12)$$

where η = viscosity of Li and L = length of pore = τh, where τ = tortuosity of the pores.

Upon combining (11) and (12) and expressing the wicking rate in terms of current per square centimeter of wick cross section, we find

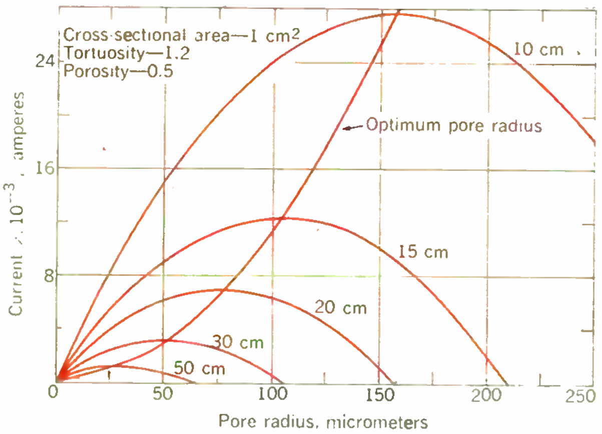
$$I = \frac{nF\rho_{Li}\varphi}{6.94}\,\frac{r^2}{8\eta\tau h}\left(\frac{2\gamma\cos\theta}{r} - gh\Delta\rho\right) \qquad (13)$$

where φ = wick porosity and ρ_{Li} = density of lithium.

A plot of (13) is shown in Fig. 5. For each wicking depth there is an optimum pore size given by

$$r_{opt} = \frac{\gamma\cos\theta}{gh\Delta\rho} \qquad (14)$$

FIGURE 5. Calculated lithium wick electrode performance as a function of pore radius and wicking depth; Eqs. (13) and (14).



For pore sizes smaller than the optimum, flow restrictions limit the total rate whereas for pore sizes larger than the optimum the capillary driving force is reduced. These wicking rates have been confirmed experimentally[13] and are sufficiently high to meet the requirements of most Li-Cl_2 cell designs.

The Cl_2 electrode is much more complex than the lithium electrode. It consists of porous carbon or graphite, which is in contact with LiCl on one side and is fed by Cl_2 gas on the other. Since LiCl does not wet graphite, the electrolyte does not normally penetrate into the pores. The Cl_2 electrode process consists of:

1. Chlorine transport through the porous electrode due to a pressure difference ΔP.
2. Dissolution of Cl_2 in LiCl.
3. Diffusion of Cl_2 to the graphite-LiCl interface.
4. Electron transfer and migration of Cl^-.

The second step is fast compared with the others and the fourth step has an exchange current density of about 0.2 A/cm² associated with it—which means that at practical current densities of 2–4 A/cm² an activation polarization of about 0.2 volt may be expected. Steps 1 and 3 are the main rate-determining steps, depending on the structure of the porous electrode. The diffusion-limited current density associated with step 3 is given by

$$i_L = \frac{8\, nFDKP_2\varphi}{r} \tag{15}$$

where n = number of electrons = 2, F = Faraday, D = diffusion coefficient for Cl_2 in LiCl, K = Henry's law constant, φ = porosity, and r = pore radius.

The partial pressure of Cl_2 at the gas–liquid interface (P_2) is determined by the pressure drop (ΔP) across the porous electrode and the buildup of electroinactive impurities (P_i), which are invariably present in the Cl_2. P_2 is then given by

$$P_2 = P_T - \Delta P - P_i \tag{16}$$

in which

$$\Delta P = \frac{iRTL}{nF\varphi K'} \tag{17}$$

$$K' = \frac{1.013 r^2 (2\, P_T - \Delta P)}{16\mu} + \frac{4\delta r}{3}\sqrt{\frac{2\, RT}{\pi M}} \tag{18}$$

$$P_i = N_0 P_T \exp\left(\frac{iRTL}{nFP_T\varphi D'}\right) \tag{19}$$

In Eqs. (17)–(19), μ = gas viscosity, δ = constant = 0.9, M = molecular weight of Cl_2 = 71, N_0 = mole fraction of impurities in the incoming gas, and D' = diffusion coefficient in the gas.

A plot of i_L versus pore radius shows a maximum indicating an optimum pore size. For pores smaller than the optimum size, flow through the porous graphite is the main limiting process, whereas for pores larger than the optimum size, diffusion of Cl_2 in the electrolyte is the main rate-limiting step. In addition, the accumulation of inactive impurities—such as N_2 and CO_2—in the porous electrode can severely limit the maximum current density.[14] Figure 6 shows a plot of i_L for various thicknesses of a particular porous graphite designated as FC-11 by the manufacturer (Pure Carbon Co.).

In spite of the various limitations of the Cl_2 electrode, power densities of 40 watts/cm² have been demonstrated in laboratory cells using the purest Cl_2 available and a highly porous and well-designed Cl_2 electrode.[15] This power density is well in excess of that required for electric vehicle power plants and is much higher than what has been obtained in any other electrochemical system.

Since both the energy and the power capacity of the Li-Cl_2 battery are high, what limitations must be overcome before the system can be put to practical use? In selecting a couple of such high reactivity to obtain a high OCV, one automatically places high demands on the container materials, in particular, the insulator. In the lightest and simplest cell designs the insulator is in contact with Li, LiCl, and Cl_2, and so must be compatible with all three media at cell operating temperatures (Fig. 4), but very few materials at present can meet this requirement sufficiently to give life expectations of the order of several years.

Safe containment methods for Cl_2 must be developed because of its toxicity; one approach that looks promising is to absorb the Cl_2 on activated charcoal.[16] The requirements placed on the containment materials can be reduced by selecting a somewhat less reactive system but this necessarily reduces the system's energy capacity.

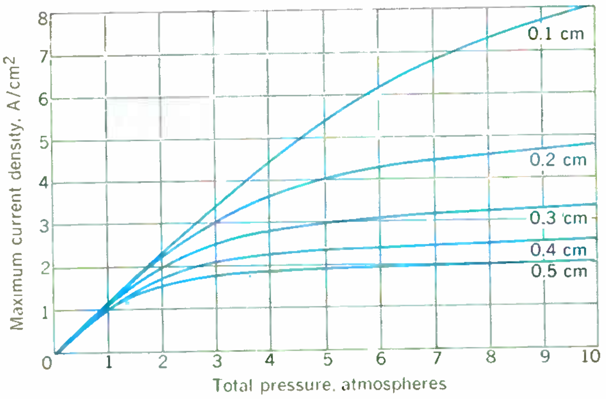


FIGURE 6. Maximum current density for FC-11 porous graphite as a function of electrode thickness and total pressure at a one percent impurity level.

The Na-S battery

Sodium and sulfur are less reactive than Li and Cl_2, respectively. This fact is reflected in the much lower open-circuit voltage for a Na-S cell, which is 2.08 volts at 300°C.[17,18] The lower OCV, together with the higher equivalent weight of sodium and sulfur in the reaction $2Na + 3S \rightleftarrows Na_2S_3$, leads to the much lower theoretical energy density of 760 Wh/kg at 300°C for this system. However, the result of this sacrifice in energy density is a much reduced materials problem since many more materials are compatible with sodium and sulfur than with lithium and Cl_2 at cell operating temperatures. Nevertheless, the Na-S cell is made possible only by the use of Na^+ ion conductive membranes such as glass or β-alumina. Because sodium and sulfur or polysulfides react chemically to form sodium–sulfur compounds containing increasing amounts of sodium, with the final compound being Na_2S, an Na-S cell would be expected to have a high rate of self-discharge unless contact between the metallic sodium and the polysulfides could be prevented.

The use of ionically conductive membranes to separate otherwise incompatible liquids is well known and they have been utilized in many reference electrodes[19] and electrochemical cell studies[20] as well as in fuel cells. The uniqueness of β-alumina for this purpose lies in its layered structure,[18] which allows a high mobility of the sodium ion between the Al_2O_3 layers and results in an ionic resistivity that is orders of magnitude lower than that of other Na^+ conductive solids at the same temperature. Modifications of the β-alumina ($Na_2O \cdot 11Al_2O_3$) have further enhanced the conductivity of polycrystalline bodies made of this material to the point where resistivities of 3.5–5 ohm-cm have been achieved at 300°C.[18]

The overall cell process in Na-S consists of:

1. A simple single electron transfer takes place at the sodium electrode similar to that at the lithium electrode in the Li-Cl_2 cell.

2. The Na^+ ion formed is transferred through the ionic separator to the polysulfide electrode.
3. Since pure sulfur is an insulator ($\rho = 10^{17}$ ohm-cm), it cannot accept the electron from the external circuit. Thus an inert electrode must be found as well as an electrolyte in which the reaction $S_x + ye^- \rightarrow S_x^{Y-}$ can occur.

Similar to the Cl_2 electrode, porous graphite has been used as the inert electrode and sodium polysulfides as both the electrolyte and the reaction product. However, because of the high resistivity of the polysulfides, the sulfur electrode still appears to be the main rate-limiting step in the overall performance of the Na-S cell and a more detailed understanding of its operation should lead to significant improvements in performance.

The simple cell designs that result from the use of a solid ionic separator between the liquid reactants (which can then be stored at cell operating temperatures as opposed to remote storage of one of the reactants in liquid or absorbed[16] form in the Li-Cl_2 cell) has inspired predictions of achieving a rather high fraction of the theoretical energy capacity of the Na-S system. Practical energy capacities of 330 Wh/kg or 43 percent of the theoretical have been forecast.[18]

Chemically bound oxidizers

The use of an ionically conductive separator is, of course, not limited to the Na-S cell but makes possible a large number of cells that would otherwise have high rates of self-discharge, viz., cells using chemically bound oxidizers, such as NiF_2, CuF_2, $CuCl_2$, CuO. These oxidizers are used extensively in organic electrolyte high-energy batteries[6] where the self-discharge is limited by the low solubility of these materials in the organic electrolyte. However, low solubility also severely limits the rates of the electrode reactions, which, with the high resistivities of the organic electrolytes, results in high-energy but low-rate batteries. However, by utilizing an ionic membrane separator, these same cathode materials can be used at elevated temperatures where much faster electrode kinetics may be expected.

Chlorine can also be bound by physical and chemical forces to activated carbon at high temperatures[21] and a cell similar to the Li-Cl_2 cell can be made without the use of gaseous Cl_2. However, a severe weight penalty is incurred since 3 to 12 carbon atoms are required to bind one chlorine atom and hence the theoretical energy density is reduced by 45 to 75 percent. Finally, concentration cells of the type Li/Li^+ fused salt/Li (Te) can also be considered in this category since their operation depends on the formation of compounds such as Li_2Te.[22] This greatly reduces the activity of lithium in the tellurium, which then results in high cell voltages. By using a mixture of LiCl, LiF, and LiI as electrolyte, power densities of up to 5 W/cm² have been demonstrated.

Conclusion

To achieve electric car performance approaching that of today's internal combustion machines it is necessary to develop batteries with high power densities and energy capacities of 220 Wh/kg or better. Since conventional batteries cannot meet this requirement, a number of new battery systems are being investigated. These all involve elevated temperatures and the use of fused salt or solid ionic conductors to achieve high power and energy densities. Aqueous electrolytes generally cannot be used since the open-circuit voltage, and thus the energy density, is limited by the decomposition of water. The zinc–air battery is the only aqueous system that approaches the required energy densities but it has significant power limitations. Organic electrolytes can be used with very active electrodes such as lithium and magnesium but the high resistivity of these electrolytes and the slow electrode kinetics severely limit the power capacity of systems in which they are used.

Several high-temperature battery systems presently under investigation have been reviewed. The electrochemical charge and discharge properties of these systems are such that high-power and high-energy-density batteries can result if the problems of safe containment of these very active materials at elevated temperatures over a period of years can be solved.

REFERENCES

1. Vinal, G. W., *Primary Batteries*. New York: Wiley, 1950.
2. Vinal, G. W., *Storage Batteries*. New York: Wiley, 1955.
3. Collins, D. H., *Batteries*. New York: Macmillan, 1963.
4. Hoffman, G. A., "The electric automobile," *Sci. Am.*, vol. 215, pp. 34–40, 1966.
5. Shipps, P. R., "Secondary metal air system," *Proc. 20th Annual Power Sources Conf.*, pp. 86–88, May 1966.
6. Bockris, J. O'M., and Srinivasan, S., "Electrochemical energy conversion," *J. Electroanalytical Chem.*, vol. II, pp. 350–389, 1966.
7. Henderson, R. E., Agruss, B., and Swinkels, D. A. J., "The electrochemical engine," *Engineering Developments in Energy Conversion*, pp. 293–308, Aug. 1965.
8. Jasinski, R., *High Energy Batteries*. New York: Plenum Press, 1967.
9. Takahashi, T., Ito, K., and Iwahara, H., "The efficiency of solid electrolyte fuel cells," *Electrochim. Acta*, vol. 12, pp. 21–30, 1967.
10. Vetter, K. J., *Electrochemical Kinetics*. New York: Academic Press, 1967.
11. Srinivasan, S., Hurwitz, H. D., and Bockris, J. O'M., "Fundamental equations of electrochemical kinetics at porous gas-diffusion electrodes," *J. Chem. Phys.*, vol. 46, pp. 3108–3122, 1967.
12. Swinkels, D. A. J., "Lithium-chlorine battery," *J. Electrochem. Soc.*, vol. 113, pp. 6–10, Jan. 1966.
13. Swinkels, D. A. J., and Tricklebank, S. B., "Lithium wick electrode," *J. Electrochem. Technol.*, vol. 5, pp. 327–331, Aug. 1967.
14. Swinkels, D. A. J., "Impurity effects in high current density Cl_2 electrodes," *J. Electrochem. Soc.*, vol. 114, pp. 812–816, Aug. 1967.
15. Hietbrink, E. H., Petraits, J. J., and Craig, G. M., "Short duration lithium-chlorine primary battery systems," *Advances in Energy Conversion Engineering*, pp. 933–941, Aug. 1967.
16. Werth, J., and Zeitner, E. J., Jr., "Storage of chlorine on activated charcoal in low pressure, high performance secondary batteries," *Advances in Energy Conversion Engineering*, pp. 927–932, Aug. 1967.
17. Weber, N., and Kummer, J. T., "A sodium-sulfur secondary battery," *Advances in Energy Conversion Engineering*, pp. 913–916, Aug. 1967.
18. Kummer, J. T., and Weber, N., "A sodium-sulfur secondary battery," presented at Automotive Engineering Congress, Jan. 9–13, 1967.
19. Labrie, R. J., and Lamb, V. A., "A porcelain reference electrode conductive to sodium ions for use in molten salt systems," *J. Electrochem. Soc.*, vol. 106, pp. 895–899, 1959.
20. Hamby, D. C., Steller, B. W., and Chase, J. B., "Alkali metal Voltaic cells," *J. Electrochem. Soc.*, vol. III, pp. 998–1002, 1964.
21. Sprague, J. W., Boos, D. L., Rightmire, R. A., Beals, D. L., and Rowland, R. L., "Active carbon plate cathodes for molten salt batteries," *Abstracts of the Fall Meeting of the Electrochemical Society*, Philadelphia, Oct. 9–14, 1966.
22. Foster, M. S., and Lui, C. C., "Free energy of formation of Li_2Te at 798°K by an electromotive force method," *J. Phys. Chem.*, vol. 70, pp. 950–952, 1966.

Electronics and the urban crisis

Electronic technology represents one of our greatest potential hopes for meeting many of the increasingly critical challenges faced by our urban society

Seymour N. Siegel *City of New York*

Although it would be naive to consider electronics as a panacea for urban ills, current and potential innovations in the electronics industry will aid in providing economically feasible answers to many of our cities' most pressing problems. At present, progress is being made in such areas as traffic control, police and fire communications and surveillance, air- and water-pollution control, and transit system routing.

Aside from the war in Vietnam, there is, perhaps, no single topic of greater import to the United States than the condition of our cities. Urban squalor and crime, the contamination of our environment, our failures in education, the inequality of citizenship for the poor—all of these press upon the conscience of the community. Obviously, cities are confronted with problems that they are ill-equipped to solve. The lack of dollars is only one of the reasons.

The role of electronic technology

It has been said that electronics will not, necessarily, meet all of our urban needs, but that the problems of the megalopolis will not be solved without electronics.

In my limited experience in city government, I have found that the need for the protection of the public purse and the numerous safeguards in the expenditure of monies result in our procuring last year's computer two years from now. This doesn't mean that cities all over the nation

Essentially full text of a talk presented at an IEEE dinner held at the Statler Hilton Hotel, Washington, D.C., February 10, 1968.

are not catching up with technology. Although there is a considerable "city hall lag" in adopting new methods and techniques, elaborate and sophisticated devices are being installed all over the nation. I would not go so far as to say that municipal technology has arrived; but, here and there, in small towns and in the large metropolitan areas, the population explosion has virtually forced city fathers to innovate, adapt, or go down to defeat.

The League of Cities has been facing up to the fact that municipalities all over the United States are now compelled to furnish local services to which they have only casually paid heed in the past, and the only way they can do so economically is to exploit technology.

When you talk to mayors and city managers, you become aware of the popular misconception that adequate electronic technology is now available to assist municipalities out of their dilemmas; however, this is not the case. Unlike the federal government, local governments are not in the habit of underwriting research and development. They have neither the resources nor the know-how. There is, perhaps unfortunately, a lack of uniform standards in the quality of local government that would preclude even our larger cities from combining to underwrite research costs. Even NASA had long, expensive experiences in the R & D area. Cities are learning, but there is no easy, royal road in the learning process; as with the individual, there is some strain involved. All too often, manufacturers and vendors fail to appreciate the problems of urban areas. At times the electronics industry solves the wrong problem or, more often, it comes up with the wrong design. Some city leaders are smart enough to attempt to exploit the experience of the military and aerospace technology and adapt it for municipal usage.

Roger Kenneth Field recently said that "successful adaptation occurs only when it is undertaken by unusually bright people who are intimately familiar with both the problems of the cities and the traditional solutions to these problems. Unfortunately, such individuals are rare and are widely scattered." This is the reason why it is said that San Jose, Calif., has the best computer-controlled traffic system; Los Angeles, the best air-pollution monitoring and water reclamation programs; Chicago, the most technologically advanced police department; and New York City, the most up-to-date communication and routing equipment for an admittedly complex rapid transit system.

Every city needs special thinking. Like any good architect, the most successful urban designer is one who combines sensitivity to the problem to be solved with an acute awareness of the environment in which solutions must operate. The best example of this kind of design philosophy is the automobile industry. Any industry capable of designing a machine that can be built by uncaring workers, maintained by ill-trained (and sometimes money-hungry) mechanics, and driven by a smug, oblivious public, and still can perform every day sets an example that might be followed by our urban technological planners. However, if we wanted to, I suppose we could blame everything on the automobile. I think it was Professor Kranzberg who said that the automobile created our suburbs, our parking lots, our gas stations, paved streets, freeways, motels, roadhouses, trailer parks, special courts, mass production, casualty figures, a losing branch of the insurance business, and a host of industries polluting the land, sea, and air. It opened new gateways of "see-it-yourself education, recreation, courtship, and sin."

New York City, which I perhaps know best, has been faced with problems that we thought reached gigantic proportions a long time ago, from the physical problems of urban renewal, transportation, zoning, traffic congestion, all the way through the full gamut of human aspects, such as crime, delinquency and its prevention, education of the disadvantaged minorities, and, finally, the expanded housekeeping chores we never believed possible. Only recently, the Commission on Intergovernmental Relations indicated that all cities were facing dire economic and fiscal difficulties.

If one were to judge by the measurement of decibels at the recent State of the Union Congressional Session as to what constitutes the number one domestic priority in the minds of people and their representatives in Congress, it would, obviously, be the prevention and control of crime in our streets.

Attempts at innovation in New York

We hear a great deal about how important it is to innovate and to make new approaches. In New York City, at least, there are visible signs that we are trying. Even if these are not always successful, we are attempting progressive action to prevent procurement of what all too often was obsolete before it could be obtained and put to use. We have our sights set on new horizons in the area of communications and we try to get involved with all of the new developments. We are still using telephones, but we have installed Centrex systems and have inaugurated mass announcing systems. We are involved in the experimental use of 200 street pay telephones capable of permitting a call to the operator from a street pay booth in an emergency even without a dime; as a matter of fact, the New York Telephone Company expects to have 2000 of these telephones operational within the next few months. Consolidated communications centers are being planned and the city's Police Department Communications Center is well into the construction phase. New York will be among the first cities to institute the three-digit national emergency telephone number 911. This is possible only because we already have a centralized net system that uses seven digits.

It is easy to visualize, within the next few years, that every foot patrolman in the New York City Police Department will be equipped with a two-way radio. The municipality spends altogether too much to recruit and train a policeman. It is not good economic sense if every unit of manpower cannot be located and deployed in accordance with acutely increasing needs. We are awaiting delivery of a thousand walkie-talkie radios for scooter patrolmen and foot men. This will triple the number of these units on the street by this summer.

The institution of the Sprint system, the first truly computer-based police dispatch system in the country, is at the midpoint. This system will shorten police response time dramatically. It is expected to be operational by the summer of 1969. Mayor Lindsay has only recently announced a six-point program, which includes the expansion and the use of one-man patrol cars in low-density areas; the addition of several hundred police trainees to direct traffic and, thereby, release regular policemen for an intensified fight on street crime; the procurement of 300 new scooters, an increase in the city-wide patrol strength in high-density and high-hazard areas; and the modernization and consolidation of various facilities, thereby releasing men from duplicated desk jobs and putting more men on the street. The communications requirements of this rapid expansion and reorganization must be met within the necessary time limits.

We can look forward to the use of a personal oscillator by every fireman, which would enable him to be located in the midst of any conflagration. We have already experimented with consumable television cameras, which can be affixed to the end of fire towers and thus permit the probing of any major fire without risking human beings.

Many of New York City's correctional institutions already are equipped with closed-circuit television installations, which permit the warden in his office to follow the progress of the individual from the moment of reception and processing through the various areas of the institution. Some people consider this a "big brother" watch, but it actually facilitates the administration of these institutions.

The city is in the midst of equipping its police helicopters with television cameras for purposes of traffic control, and probably for the control of air pollution as well. The use of these cameras in crime detection, such as the maintenance of surveillance on various types of motorcars from way above ground, is not far off.

We have fallen flat on our collective faces in connection with installation of electronic traffic-light control systems. We put too much faith in one of the large aerospace manufacturers who failed, after three and a half years, to deliver any equipment that really worked dependably.

© General Drafting Co., Inc.

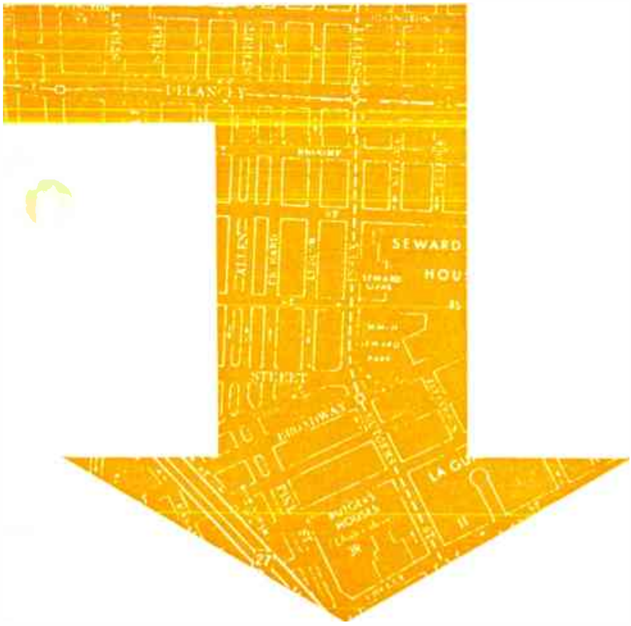
Traffic Commissioner Henry Barnes has not given up hope and I have every confidence in his ability to achieve his dream of taming the megalopolis, but we are starting all over again. The installation of two-way radios in buses as well as on subway trains will, certainly, be vital in the control of crime and will also permit more efficient dispatching and loading procedures.

New York City has just combined all of its environmental services into a pioneering single superagency. For the first time, we are seeking to deal with our physical surroundings in a related manner. The problem of air pollution cannot be solved if such solutions have an adverse effect on water pollution. We are coordinating our efforts to deal with solid waste disposal, water supply, noise abatement, and air and water pollution. All of these activities need and require a special electronic technology and pose special communications problems as well.

It was not so long ago that the great City of New York depended on a man in a rowboat in upper New York Bay to pull a pail of water every hour to provide the necessary sampling to control the city's sewage disposal activities. We are now planning the installation of various in-stream sampling devices, electronic sensoring equipment, and electronic reporting devices measuring drops in water pressure and speeding up the location of water-main breaks. Telemetering will solve other problems for us.

Likewise, in air-pollution control, after finally deciding what we want to look for, 21 monitoring points will permit the electronic sampling of our air. Hopefully, we'll obtain regional adoption of our standards and we'll solve the problem of what levels of pollutants are really dangerous. Monitoring air pollution is one thing; controlling it, of course, is another. Since this function must take place at the source, motor vehicles and industrial and home smokestacks must be controlled in some technological fashion.

Noise abatement poses additional problems, especially since New York has two of the nation's most congested airports as well as rapidly increasing helicopter traffic. In all of these areas, we are seeking and installing the necessary technology to help improve the quality of life in



the big city. I am not prepared, however, to report any glowing success stories at this time.

Like many other large cities, New York is faced with an acute shortage of radio frequencies in the Land Mobile Services—a shortage that affects all municipal operations. However, after two years the present City Council has begun to implement the Lindsay plan for reorganizing some 50 independent departments of city government into a more manageable group of ten agencies. Since this is not merely a physical regrouping but represents a highly professional breakdown by function, the problem of mere reorganization of communications systems is a formidable one. In the context of what we are discussing, this may be a minuscule footnote, but I find myself in the middle of a dilemma. As a broadcaster, I have devoted a large portion of my life to the development of UHF television. On the other hand, I am all too aware of the acute critical need for spectrum space for our public-safety functions. Unfortunately, broadcasters have already hardened their position with regard to suggestions for reallocations. Some hard-nosed decisions will have to be made and, at the same time, some concessions will have to be granted.

Needed: a sense of urgency

Enough of this casual cataloguing of urban uncertainties. It is probably superfluous for me to underline the sense of urgency that is called for. Neither a crash program nor "business as usual" will suffice. It was just a little more than a decade ago that a man-made object in space rallied the academic, scientific, industrial, and government resources of this country to produce a tangible answer. The urban Sputnik has become visible and has been flashing before our eyes for about four summers. With regard to the technological aspects of city needs, neither the innovators, product designers, nor technicians have, as yet, reacted with the urgency that the situation demands. Perhaps, many of us have individual doubts about the identification and extent of the problem. No one suggests that gadgetry and new tools will, of themselves, provide a solution. But technology can help our cities do other jobs more efficiently or will release resources and manpower. Anything that will permit our municipalities to concentrate on the new human problems cannot help but be of assistance in the restoration of social health.

It is easy to conclude that the time is long overdue for all of us in the universities, in industry, in state and federal government—in fact, all of us as individuals, wherever we live—to consider ourselves directly involved in the issues and conflicts that affect local governments.

Foundations and industry are channeling funds into our universities, and the institutions of higher learning are devoting more and more effort to the various facets and proposals of the cities and to the field of urbanology. Centers for urban study are flourishing at Harvard, M.I.T., the University of Chicago, New York University, the City University of New York, and other colleges and universities. For the last two decades, numerous institutions have been involved with communications research centers; perhaps a marriage is called for. The research center might be able to contribute more effectively by concentrating on the two disciplines of urbanology and communications. Certainly, universities could and should devote some of their research and development effort into the direct application of electronics technology to city functions.

Private industry, itself, has an obligation to put aside high profits in what could well be described as an emergency situation. The field of urban renewal housing programs abounds with examples in which private enterprise has forgone normal immediate financial returns but is, nevertheless, surviving and prospering. The total cost of research and development need not be applied *in toto* to a single application in a single city just because this was the first order received.

Insofar as federal and state governmental jurisdictions are concerned, you know it takes a bureaucrat to know a bureaucrat. There is nothing more frustrating for the local-level animal than an exercise of ministerial discretion at a higher level when the action is adverse. It is popular, and probably necessary, to look to Washington for dollars, but it is equally necessary to cut out red tape, delay, and procrastination.

Not so long ago, Commissioner Nicholas Johnson suggested the creation of an Office of Urban Communications in the Federal Communications Commission. There are many people who would like to see something like this if the purpose is to provide necessary priorities and to be helpful to cities. The educational establishment has been pampered over a considerable number of years and the results have been salubrious. The municipalities should receive the same kind of tender loving care and, perhaps, public safety ought to be given a higher priority over purely commercial requirements.

As a broadcaster, I have found that listeners and viewers often write in and ask what they, as citizens, can do. We usually suggest that they bring to bear their own expertise. Beneficial suggestions are always welcome when they represent an individual's own brand of experience and training. You'll find it an exhilarating experience to concentrate on your city's problems, even while you shave. Some of our best engineers are generalists. If any single one of you comes up with the workable answer to any single problem, you will, to some extent, free today's urban man from the crippling fetters of insecurity and rejoice with fellow citizens in the success and viability of "our town." We have few options left.

Laser applications

Improvements in efficiency and reliability have expanded the practical usefulness of laser devices. The extremely diverse applications include, for example, cutting and welding, retinal attachments, measurement of current in EHV power lines, and obstacle avoidance for the blind

R. D. Haun, Jr. *Westinghouse Research Laboratories*

When lasers were new many people seemed to expect that this exciting development would rapidly revolutionize civilization. More than seven years have now passed since the advent of the first laser and the world has changed only slightly. As a result, it has been said that the laser is a solution in search of a problem. It is the purpose of this article to show that, in spite of comments to the contrary, there are some problems for which lasers are a solution. It is appropriate to do this now because the state of the art has reached the point at which useful energy and power outputs can be achieved with sufficiently high efficiency and reliability to make some applications practical. The emphasis here is on industrial and medical applications, as opposed to the many present and potential military applications. Despite the progress made during the past few years, much more research and development will be required before the full potential of lasers is realized.

Prior to the introduction of the laser the only available sources of light were essentially isotropic radiators. One of the consequences was that the temperature that could be produced by using lenses or mirrors to focus the light emitted from the source was limited by the temperature of the light source itself. Consequently, attempts to use light for such applications as welding and cutting had only limited success because the requirement for high source temperature resulted in short source life and high enough temperatures could not be attained to cut or weld some materials.

The laser circumvents this limitation by generating light throughout a large volume in a unique way that allows the light to be focused on a small spot to produce temperatures exceeding the temperature of the laser itself by many orders of magnitude. For example, the outputs of some lasers have been measured to be greater than 10^9 watts with a pulse duration of about 10 ns.[1] Using a lens to focus this on a spot 10 μm in diameter produces a power density of 10^{15} W/cm^2. In comparison, at the sun's surface the radiated power is only about 10^5 W/cm^2.

Although at least one application (the use of lasers to try to start thermonuclear reactions) requires power densities this high, most laser applications are much more modest in their requirements. Laser output power levels of a kilowatt are easily achieved for times longer than a millisecond.

Applications that take advantage of the high degree of collimation of laser output light include optical alignment, hole punching, spot welding, continuous cutting, and retina reattachment, as well as such military applications as optical radar and laser range finders. Only the more promising nonmilitary applications will be treated in this article.

Another limitation of light sources prior to the advent of the laser was the difficulty of obtaining high power in a narrow range of wavelengths. Although monochromators could be used to filter incoherent light down to a very narrow spectrum, this filtering was achieved at the expense of intensity. However, the output of the laser is not only directive but extremely monochromatic as well. In fact, some laser output spectra are so narrow that the range of wavelengths cannot be measured with a monochromator. Instead, techniques analogous to RF heterodyne detection must be used. With these techniques it has been determined that the bandwidth of some lasers is less than one part in 10^{10} of the frequency of the laser radiation.[2]

Applications that employ the monochromaticity of laser outputs include holography, optical information processing, interferometric measurement and alignment, optical inspection, laser gyroscopes, and Doppler measurements of fluid-flow velocity.

How does the laser work?

Interaction of radiation with matter. Since there are many excellent descriptions of laser physics, only a summary will be given here. The reader is urged to consult one of the cited references to obtain a more detailed picture.[3-5]

Those working with lasers usually define "light" as electromagnetic radiation having a wavelength between 300 μm and 0.01 μm. Electromagnetic radiation can interact with matter in three ways: (1) absorption, (2) spontaneous emission, and (3) stimulated emission. Radiation is absorbed or emitted in discrete quantities called "quanta," which can be absorbed or emitted only if their energy is equal to the difference in energy between two allowed energy levels in the atoms, ions, or molecules making up the material with which they interact. If this difference in energy is E, then the frequency ν of the electromagnetic radiation must be $\nu = E/h$, where h = Planck's constant $= 6.62 \times 10^{-34}$ W·s². Quanta are absorbed when the energy of the atom (or ion or molecule) increases and they are emitted when the energy of the atom decreases. If the emission occurs in the absence of any quanta it is called "spontaneous" emission. Spontaneous emission is roughly isotropic; that is, quanta are emitted in all directions. If the emission is induced by the presence of quanta at the frequency $\nu = E/h$, it is called "stimulated" emission. Stimulated emission tends to be emitted in the same direction as the stimulating quanta.

Figure 1 is a schematic representation of three interactions between matter and radiation. It shows a hypothetical material containing four atoms, each represented by a solid circle. Each transition that results in absorption or emission of a quantum of energy $h\nu$ is indicated by a dotted arrow. Two energy states are shown, with the higher energy state represented by the upper horizontal line.

The process of stimulated emission is the exact inverse of the process of absorption. To a first approximation, the probability for absorption of each quantum by each atom in the lower energy state is equal to the probability of stimulated emission from each atom in the upper state when induced by each quantum. Consequently, whether stimulated emission or absorption will predominate depends on whether more of the atoms are in the upper or the lower state, respectively. If more of the atoms in a material are in the upper state than in the lower one, stimulated emission will predominate and on the average each quantum of energy $h\nu$ will be amplified by inducing the stimulated emission of other quanta as it moves through the medium.

Most lasers are used as oscillators rather than amplifiers. In a laser oscillator the quanta that are spontaneously emitted by some of the atoms decaying from excited states are amplified by inducing stimulated emission from other atoms in the laser.

To make the output of a laser oscillator directive, the shape of the amplifying medium is chosen so that the amplification will be greater in one direction than it is in another. This can be done simply by making the laser longer in one dimension than it is in the other two.

FIGURE 1. Three different types of interaction of radiation with matter. A—Absorption. B—Spontaneous emission. C—Stimulated emission.

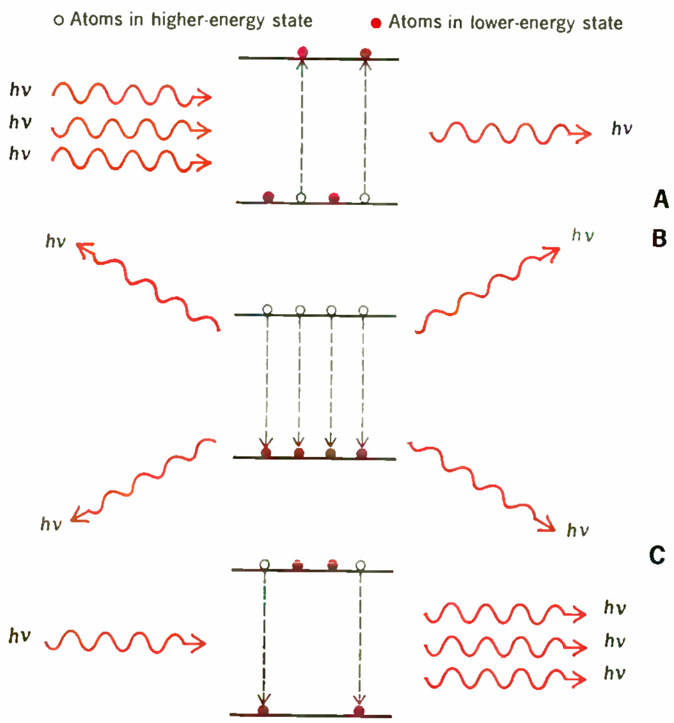
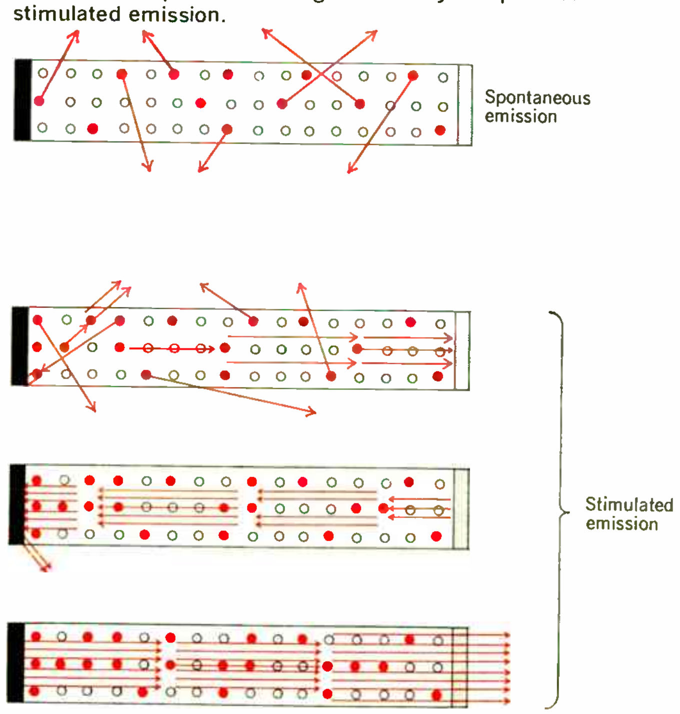


FIGURE 2. Amplification of light waves by the process of stimulated emission.



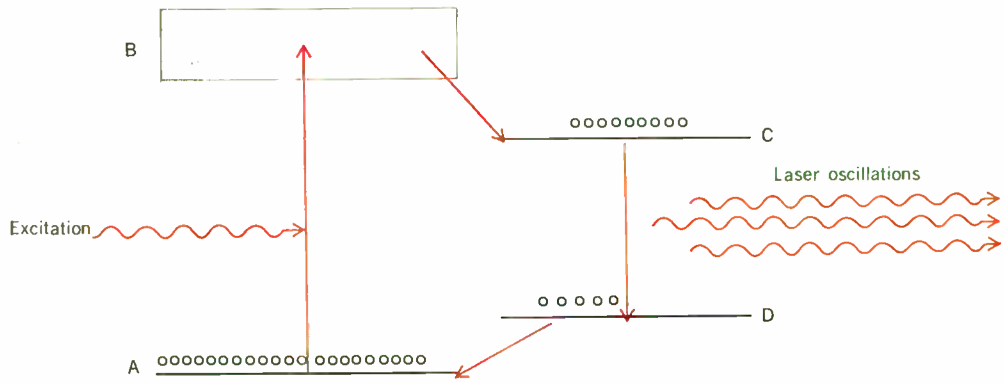
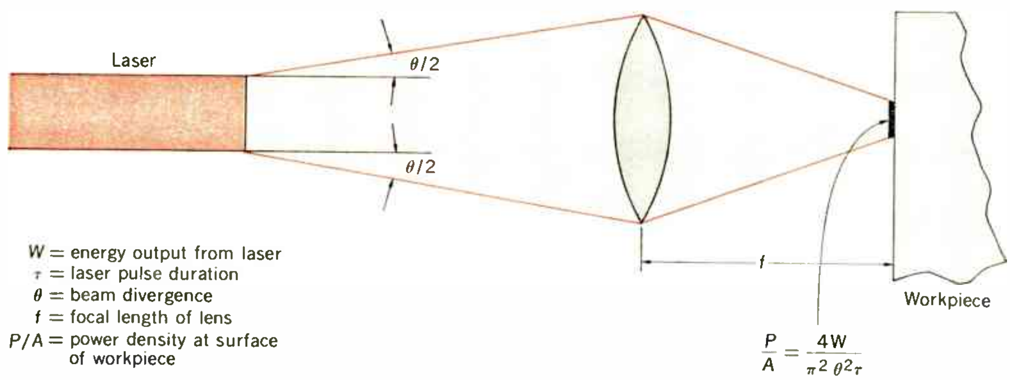


FIGURE 3. Laser oscillations in a fluorescent system.

FIGURE 4. Focusing-laser output.



Since there are practical limitations on the length of a laser, use is frequently made of partially reflecting mirrors on the ends of the laser to increase the amplification of quanta emitted along the length of the structure; see Fig. 2. If the reflectivity of the reflector at one end is 100 percent, and of the one on the other end is R, then on the average each quantum will have to hit the R reflector approximately $1/(1 - R)$ times before it will escape from the laser. Consequently, the effective length of a laser of actual length L is roughly $2L/(1 - R)$.

Production of amplifying states. Under normal conditions the number of atoms in each higher energy state is lower than the number in each of the lower energy states. Consequently, materials normally absorb radiation rather than amplify it.

There are several different general types of lasers. These differ from one another in the method used to establish or maintain an amplifying state in the laser material. In general, in each of these schemes the atoms in the laser medium are excited by some process that selectively excites a higher-lying energy state more rapidly than it does a lower state. If the rate of excitation is higher than the rate of spontaneous or stimulated emission to the lower state, it may be possible to achieve a condition wherein the population of the higher state will exceed that of the lower state, and laser oscillations will occur.

Optically pumped lasers

Optical pumping. Many laser materials can be excited by illuminating them with sufficiently intense light from an incoherent light source. Examples are ruby (Al_2O_3 containing small amounts of triply ionized chromium ions), glass containing Nd^{3+} ions, $Y_3Al_5O_{12}$:Nd (nicknamed "YAG"), $Y_{1.5}Er_{1.5}Al_5O_{12}$:Ho, Tm, Yb (nicknamed "alphabet YAG"), and $Ca_5(PO_4)_3$:Nd (nicknamed "FAP"). Although the first laser (ruby) required a very-high-power pulsed flash lamp to achieve laser oscillations, more recent materials (such as YAG and FAP) can oscillate continuously using a tungsten filament as the pump.

Long-pulse optically pumped lasers. All of the materials mentioned in the preceding paragraph can be pumped by means of pulsed flash lamps. As a specific example, the Nd-doped glass laser will be described in some detail.

If the rare-earth element neodymium is added to optical glass, each trivalent neodymium ion has the energy level scheme shown in Fig. 3. The two levels C and D have an energy difference that corresponds to quanta having a wavelength of 1.06 μm. Atoms in states B and D decay rapidly to states C and A, respectively, by giving up their energy in the form of heat into the glass. The energy difference between A and B corresponds to quanta having several broad bands of wavelengths near 0.58, 0.74, and 0.81 μm. A high-energy-pulsed flash lamp emits sufficiently intense radiation at these wavelengths that transitions from A to B can be excited much more rapidly than the ions spontaneously decay from C to D. Since the ions excited to state B by the light from the flash lamp decay to state C rather than back to state A, then although the glass is illuminated by the flash lamp a population excess can be built up in state C relative to state D, and laser oscillations can be achieved at a 106-μm wavelength.

In a flash-lamp-excited neodymium-glass laser, the laser material might typically be in the form of a 30-cm-long, 1.3-cm-diameter glass rod of high optical quality containing about 4 percent neodymium. The flash lamp is excited by dumping an electric charge of up to 10 000 W·s from a capacitor bank through it in a few milliseconds. The dissipation of this power in the gas-filled tube causes it to become incandescent. A large fraction of light emitted by the lamp impinges on the laser rod, which absorbs those portions of lamp radiation that overlap the absorption bands of the neodymium ion. The capacitor bank is recharged from a dc power supply during the time between pulses. The laser output pulse is nearly as long as the pulse of electric current that heats the lamp.

As the input energy (per pulse) to the lamp is increased from zero, the population of state C in Fig. 3 increases. As long as the rate of pumping by the flash lamp is too low to give population inversion, the isotropic spontaneous emission (fluorescence) at 1.06 μm increases. However, when the rate of pumping becomes large enough to maintain a population inversion, the material becomes amplifying and laser oscillations begin. Above the threshold energy input that just produces oscillations, the laser output increases roughly linearly with the input energy.

Pulsed-laser hole punching. If the output of a laser is focused by a lens of focal length f the diameter d of the spot produced will be given roughly by $d = f\theta$, where θ is the angular spread of the output of the laser. The power density P/A at the focal spot is therefore given by

$$\frac{P}{A} = \frac{4W}{\pi f^2 \theta^2 t}$$

where W is the output energy of the laser and t is the time duration of the output pulse.

For the neodymium-glass laser described in the preceding section the angular spread of the output is about 5 mrad, where the output is 60 joules and the pulse duration is 2.5 ms. Using a 7.0-cm lens having a 7.0-cm focal length to concentrate this output pulse gives a power density of 2.5×10^7 W/cm^2 at the lens focus. Using the setup shown in Fig. 4, holes 0.4 mm in diameter were punched through 5.3-mm-thick cast iron using a single pulse from the laser to produce each hole.

Although pulsed lasers built to date for industrial application more typically provide 10 rather than 50 joules, experience with the smaller industrial units and with laboratory units with outputs up to 200 joules per pulse indicates that it should be reasonable to expect at least 100 000 shots between maintenance at a pulse repetition rate of about two pulses per minute at the 60-joule output level. The item that needs to be replaced after this number of shots is the flash lamp, which typically costs less than $100. Lasers operating at 10 joules per pulse are currently capable of pulse repetition rates in excess of one pulse per 5 seconds. Holes with smaller depth can be produced using lasers with lower energy output per pulse. Using a microscope instead of a simple lens makes it possible to produce very small holes, ranging down to 10 μm in diameter.

The use of the laser to make diamond dies has been extensively investigated by the Western Electric Company.[6] For miniaturized circuitry it is desirable to have wire as small as 42 AWG (about 0.1 mm in diameter). The most practical method for drawing such small wire diameters is to use diamond dies. Conventional methods for making these dies utilize mechanical grinding techniques, which typically require 24 hours to pierce a 20-point die stone. However, if multiple shots from a 10-joule-output pulsed ruby laser are used, the time required to pierce the same size stone is less than 10 minutes. The laser can also be used to resize holes in dies that have become worn. Even though the initial cost of the laser die-drilling equipment is more than that of the mechanical grinding equipment, the reductions in time and maintenance required result in a major cost reduction in the manufacture of diamond dies.

To date very little systematic study has been made to determine how the properties of the holes produced vary with the physical and chemical properties of the material irradiated with the laser. The holes produced in some materials, such as cast iron, have quite large depth-to-diameter ratios, whereas those produced in some other materials have much smaller ratios. The roundness and uniformity of the holes also vary with the material. It is believed that deep holes may be the result of the recoil momentum produced by the vaporization of the material at the surface of the material illuminated by the laser radiation. This may produce some sort of shock wave, which propagates through the material in a very directional manner to punch out the hole. Some explanation rather than simple heating of the removed material is required for three reasons: (1) To vaporize completely the amount of material removed by the laser in punching the desired hole would require much more energy than was contained in the single pulse. (2) The time required for heat to be transmitted through the entire thickness of the irradiated material is much longer than the short pulse duration. (3) The heat would tend to diffuse uniformly in all directions from the irradiated spot rather than being confined to the uniform-diameter cylinder which is punched out.

The advantages of using the pulsed laser for hole punching are as follows:

1. Because there is no physical contact between the punching tool and the material, such problems as drill breakage and wear are eliminated. Use of the laser process also may ease the problems of jigging and fixturing and make it easier to transport the article to be drilled into and out of the work station.

2. The time duration of the laser drilling process can be so short that the article to be drilled can be moving while the hole is drilled.

3. Thermal damage of adjacent regions can be minimized, because the time duration of the hole-punching process is too short for thermal conduction to occur.

4. The optical system that is used to locate the position of the hole to be produced and to magnify its image can also be used to focus the laser radiation.

5. Very-small-diameter holes can be drilled because the laser radiation can be focused onto a very small spot.

Pulsed-laser spot welding. If the power of the laser is reduced, if a lens with a long focal length is used, or if the material to be irradiated is placed some distance away from the focal spot of the lens, the power density at the focal spot may be high enough to weld the material but too low to produce vaporization. Under these conditions the laser can be used to produce spot welds.

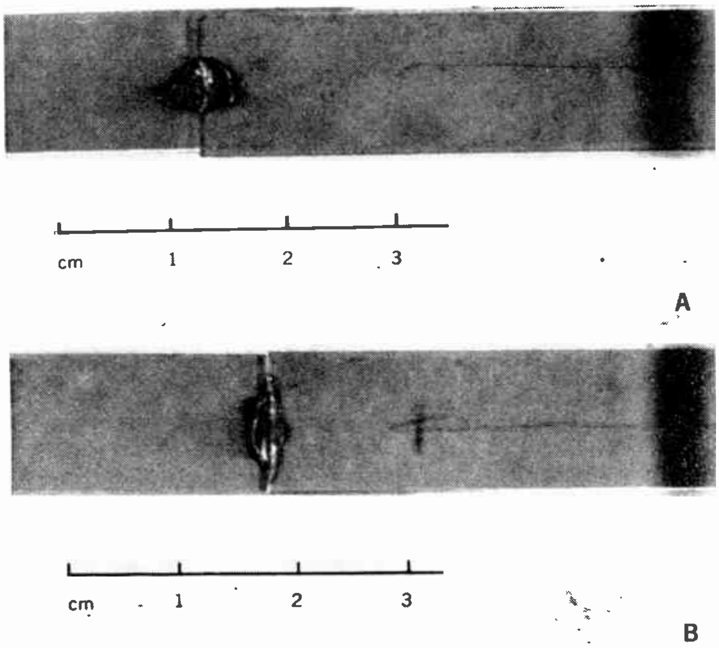


FIGURE 5. Lap welds made by use of a 200-joule laser. A—Spherical lens. B—Spherical-cylindrical combination.

FIGURE 6. Gold wire, 0.076 mm in diameter, laser-welded to 200-Å-thick gold.

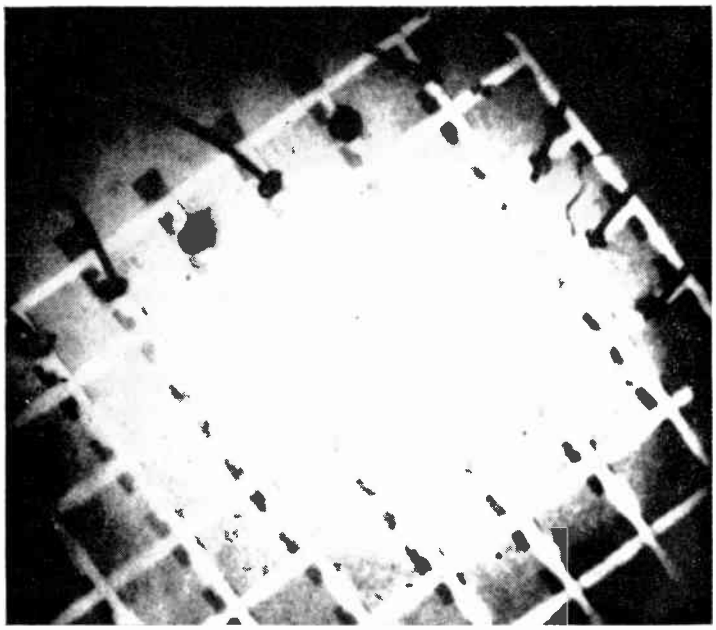


Figure 5 shows lap welds between two pieces of 0.76-mm-thick 304 stainless steel using a 200-joule 5-ms-duration pulse from the neodymium glass laser described earlier. The top portion of the figure shows a weld produced using spherical lens with a focal length of 25 cm to focus the laser output onto the junction of the two pieces of metal, which was located 23 cm from the lens. The weld shown in the lower portion of the figure was produced under identical circumstances except that a cylindrical lens with an 18-cm focal length was used in addition to the spherical lens. The cylindrical lens defocuses the beam in one direction so that it is spread out along the junction to be welded. The mechanical strength of these welds has been experimentally determined to be comparable with that obtained with other more conventional welding techniques. It was not necessary to use fill metal in making these welds.

In one application, repeated shots from a 40-joule 5-ms-duration neodymium–glass laser were focused using a cylindrical lens to produce a vacuum-tight seal around a small rectangular relay can. Approximately 12 shots were required to seal the 3.8-cm-long perimeter seam of the can.

In another application, several faulty electron tubes were repaired by welding 0.76-mm-diameter Kovar wires to 0.25-cm-thick steel tabs through the glass vacuum envelope.

In the manufacture of miniaturized electronic components, microscopes are routinely used to position components to be welded together. By making minor modifications it is possible to use the microscope to focus the output of a laser to do the welding. Figure 6 shows 0.076-mm-diameter gold wires laser-welded to gold film circuits, 200 Å thick, deposited on a silicon substrate with a 1.5-joule laser output.

In order to produce good welds it is necessary to use long pulse durations. For most metals 5 ms is sufficient to allow the heat being generated by surface absorption of the laser radiation to be conducted deep enough to give a useful volume of molten material. In comparing laser welding with electron-beam welding there is sometimes a misconception that the electron beam penetrates much further into the irradiated material than does the laser radiation. Actually, absorption depths for both electron beams and laser radiation are usually negligible as compared with the thickness of welds produced. Perhaps the misconception has arisen because electron beams are known to produce welds that can be as deep as 15 cm, but lasers have not yet done so. However, the deep penetration of the weld must result from some mechanism other than direct penetration of the electrons through the metal. Perhaps the mechanism for deep electron-beam welds is the same as the mechanism that produces the deep holes when the pulsed 1.06-μm laser is used for hole punching in metals and when the 10.6-μm laser is used to cut plastics. More systematic studies of pulsed laser welding with long pulse durations need to be performed to determine whether conditions exist that would yield deep penetration for laser welds.

The pulsed lasers used for spot welding are essentially the same as those used for hole punching, described earlier, except that for welding pulse-forming networks are added to the energy-storage capacitor-bank portion of the power supply to increase the pulse duration and flatten the power output pulse.

The use of pulsed lasers for spot welding has several advantages:

1. The fact that there is no physical contact between the welding tool and the work piece eliminates the need for periodic maintenance, such as dressing of electrodes. Therefore, as in the case of laser hole punching, it may simplify transporting the article to be welded into and out of the work station, as well as jigging and fixturing.

2. Welding can be done in air, vacuum, or a controlled atmosphere. The laser beam can enter the enclosure through a transparent window.

3. The same optical system can be used for focusing the

laser radiation as for magnifying and locating the position of the spot to be welded.

4. Very small welds can be made.

5. Welds can be produced in regions that are not readily accessible to conventional welding equipment.

Medical applications of long-pulse lasers. The laser provides a valuable supplement to other techniques used for repairing detached retinas and for other surgical operations in human eyes.[7] In the repair of a retina that has fallen away from the eyeball, it is desirable to cause scar tissue to form at several points along the torn retina so that the retina will re-establish close enough contact with the choroid to receive nutrients from it. By the use of an intense light beam directed into the eye through the lens, a small region on the retina can be burned; scar tissue will then form to repair some types of tears and detachments. For several years incoherent short-arc lamps have been used in conjuction with mechanical shutters for this application. However, the radiance available with the short arc is not high enough to allow the irradiation time to be short in comparison with times of spasmodic motion of the eye and the burned area is consequently larger than is necessary. Pulsed glass and ruby lasers have high enough radiance that the irradiation time can be much shorter, and thus the scar area is minimized. No anesthetic is required for this treatment and negligible pain is sensed during or after irradiation.

The use of lasers in cancer research is being studied widely.[8–10] In some experiments irradiation of small spots on large, malignant tumors in hamsters and mice has been followed by apparent disappearance of the tumor, but much more work needs to be done to determine the role of lasers in aiding or hindering tumor growth. The phenomenon that occurs during laser irradiation is probably primarily localized intense heating due to absorption of the laser radiation; this heating may result in the formation of chemical agents that impede the growth of the tumor.

Q-switched lasers. Lasers can be operated in a mode, called "*Q* switching," in which energy can be stored in the laser rod prior to being released in a very short high-power pulse. Typical pulse durations are 10^{-8} second or less, and peak powers from 10^6 to 10^{10} watts can be achieved. The energy storage is accomplished by inserting a variable absorber between one end of the laser rod and one of the laser end reflectors. During the first part of the pump pulse this absorber is kept in a highly absorbing state so that little amplification occurs along the length of the laser and a larger population excess builds up in the upper level than if the laser were oscillating. Near the end of the pump pulse the variable absorption is reduced to a small value so that the laser output light can be strongly amplified as it bounces back and forth between the mirrors. *Q*-switched lasers are useful both because of the very high attainable powers and because of the very short pulse lengths.

Clear-air turbulence detection. When a radar employing a laser is directed into the atmosphere on a perfectly clear day a return signal is sometimes detected.[11] These laser radar echoes are caused by discontinuities, inhomogeneities, and turbulence in the atmosphere. Experiments attempting to use airborne laser radar systems to detect clear-air turbulence have been inconclusive. These experiments have used laser radars fixed on the airplane, and thus it is difficult to assure that the airplane will fly through the small region of turbulence that may be detected by the laser. The use of scanned laser sources might lead to more conclusive results in these experiments and might be useful as a means for avoiding these small regions of danger.

Plasma trigger. To achieve self-sustaining thermonuclear reactions for power generation it is necessary to achieve plasma temperatures greater than 5×10^{8}°K and density greater than 10^{14} cm^{-3} while also confining the plasma.[12] Basov and Krokhin[13] have suggested that it may be possible to achieve these conditions by focusing the output of a *Q*-switched laser having a 100-joule 10-ns pulse deviation on a small pellet of material. Experiments are currently under way at several laboratories using lower power and energy levels to explore the mechanism of laser plasma heating.[14]

Ambartsumyan *et al.*[15] have recently demonstrated ruby laser output pulses with a peak power of about 5×10^9 watts (10 joules of energy in a pulse of less than a 2-ns duration). The French Compagnie Général d'Electricité[16] manufactures and sells a unit that provides 1.5×10^9 watts (45 joules of energy in a pulse of less than a 30-ns duration).

Continuous optically pumped lasers. Although gas-discharge lasers (discussed in the following section), particularly carbon dioxide, currently provide the highest efficiency for conversion of electrical input to continuous laser output, optically pumped solid-state lasers are capable of competitive efficiencies and are much more compact than those using gas.

Overall efficiency in excess of 5 percent has been achieved in "alphabet YAG" using a tungsten filament as the pump.[17] This material is YAG ($Y_3Al_5O_{12}$) multiply doped with trivalent rare-earth ions Ho, Er, Tm, and Yb. The output is at 2.1 μm. More than 15 watts of output have been obtained for 300 watts into the lamp. This laser has two major drawbacks: It requires cooling to liquid nitrogen temperatures, and the output wavelength is in an inconvenient region for some types of optical detectors. An additional problem is that difficulty in growing large rods of multiply doped YAG may limit power levels that can be attained with this material. The high efficiency of "alphabet YAG" is obtained because the Er, Tm, and Yb ions absorb energy at many different wavelengths in the near infrared and transfer this energy efficiently to the Ho ions. This results in much more efficient utilization of the broad-spectrum tungsten emission than would be obtained if only one rare-earth element were used.

Another way to achieve high efficiency in continuous optically pumped lasers is to shape the emission spectrum of the pump source to provide a high degree of overlap with the absorption spectrum of the laser material. For example, if sodium is added to a mercury-vapor discharge the emission is concentrated into a few narrow lines near 5890 Å, which can be made to match several absorption lines of Nd in YAG. Recent experiments have shown that it is feasible to make YAG:Nd lasers with continuous output powers of 20–30 watts at 1.06 μm using a 1-kW spectral additive lamp as the pump.[18]

Output powers as high as 27 watts have already been obtained at 1.06 μm by pumping a 3.0-cm-long 0.5-cm-diameter YAG:Nd rod with 1700 watts into a tungsten filament.[18]

Mode locking. Perhaps one of the most interesting laboratory applications for optically pumped lasers is

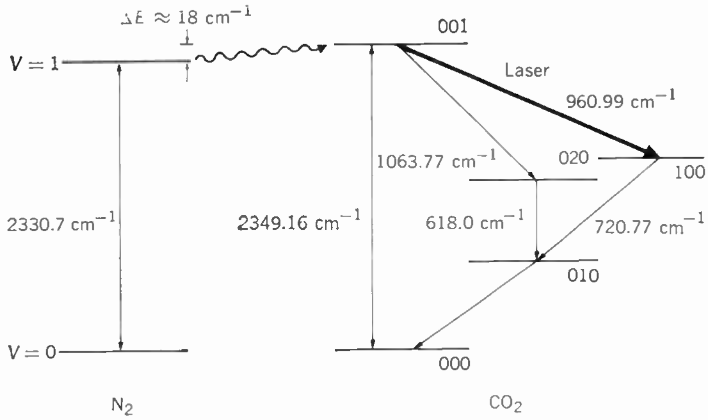


FIGURE 7. Energy-level diagram for CO_2 laser.

FIGURE 8. High-power CO_2 gas laser.

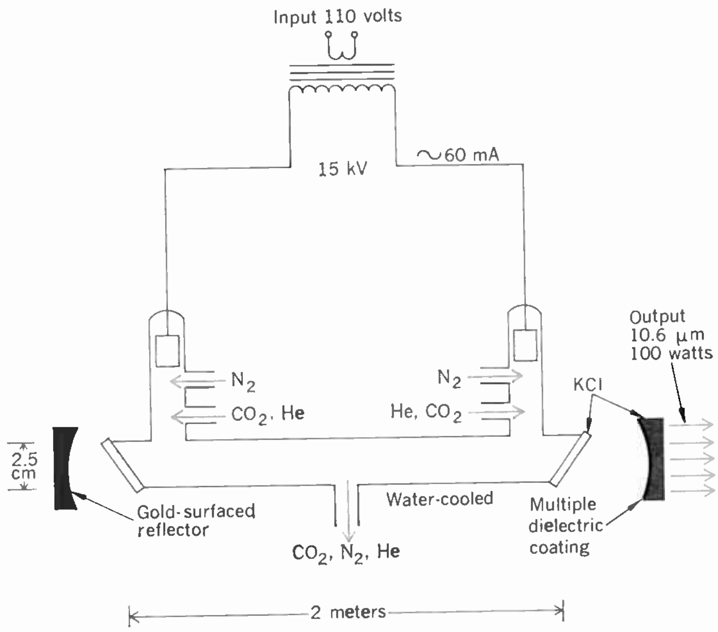
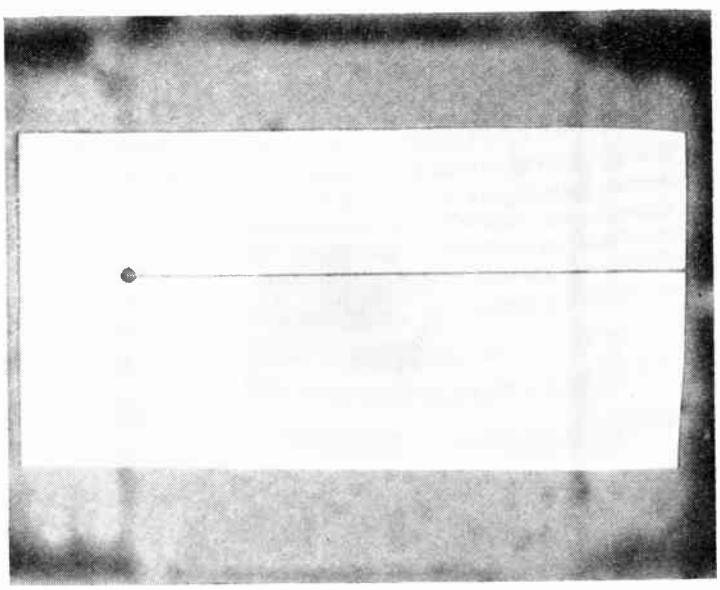


FIGURE 9. 0.25-mm-thick paper cut with CO_2 laser.



mode locking. To achieve mode locking, a modulator is inserted inside the laser resonator, and the resonant frequency of the resonator is thereby modulated at a frequency corresponding to the difference between two of the Fabry–Perot resonator modes. When this is done, several of the resonator modes oscillate at once in such a way that the Fourier spectrum of the resultant corresponds to a series of pulses repeated at a frequency equal to the spacing between Fabry–Perot modes (typically a few hundred kilohertz). The remarkable characteristic of these pulses is that the duration of each one may be as short as 4×10^{-12} second.[19] Since the time duration of each pulse is inversely proportional to the frequency spread of the emitted energy, and since the laser emission line is the source of the energy, the shortest pulses are attained by using materials that have broad laser-emission transitions. Inasmuch as the optically pumped solid-state materials have broader emission lines than do gas lasers, they are more attractive for this application. Specific practical applications for this extremely short-pulse capability have not yet been developed.

High-power continuous carbon dioxide gas lasers. In a mixture of carbon dioxide and nitrogen gas each of the molecules has the energy scheme shown in Fig. 7. By placing a mixture of carbon dioxide, nitrogen, and helium gases in a structure such as that shown in Fig. 8 one can achieve laser oscillations at a 10.6-μm wavelength by simply running direct or alternating current through the discharge. The population inversion between the 001 level and the 100 level is maintained by the following processes:

1. Nitrogen molecules are excited by collisions with electrons in the gas discharge from the $v = 0$ level to the $v = 1$ level.

2. When excited nitrogen molecules collide with CO_2 molecules in the ground state 000 they transfer their excitation energy to the CO_2 molecules, leaving the CO_2 molecule in the 001 state and the N_2 molecule in the $v = 0$ state after the collision.

These two processes occur rapidly enough to maintain higher population in the upper CO_2 level than in the next lower one. If the exciting current is continuous, then continuous laser oscillations are achieved at 10.6 μm (corresponding to 960.99 waves per centimeter).

This laser is very efficient and the power output is high. Continuous power output in excess of 1 kW has been reported with an efficiency for conversion of electric input power to laser output power greater than 10 percent.[20,21] Operation for hundreds of hours without maintenance is attained at power output levels of 150 watts and below. If resonator techniques can be improved to withstand operation at higher powers there seems to be no fundamental limit on the power that could be attained with a carbon dioxide laser.

The maximum output power of a carbon dioxide laser is directly proportional to its length and has very little dependence on the diameter.[21] Diameters vary from 2.5 to 10 cm. For 150 watts output the laser tube is about 1.8 meters long. For a kilowatt output the laser must be 15 meters long. Techniques are being developed to fold the carbon dioxide laser tube into more convenient geometries.

Cutting and slitting using continuous lasers. The output of the carbon dioxide laser can be focused by using curved mirrors having a thin gold reflective coating or by using lenses of an infrared transmitting material such as germanium. The earth's atmosphere has extremely low absorption at 10.6 μm.

The carbon dioxide laser appears to be more useful for cutting and welding organic materials and inorganic insulators than for metals. This is because most metals have very high reflectivity at 10.6 μm so that the use of focused CO_2 laser output to heat the metal is very inefficient. It may be possible to cut metals by using the laser radiation as a sort of catalyst to accelerate chemical etching processes.

Deep cuts have been successfully produced in such materials as wood and lucite without charring of the edges. The carbon dioxide laser has also been used in cutting ceramics.

Further research needs to be carried out to determine whether the depth of the cut produced by the laser increases in proportion to the output power.

Thin-sheet materials such as plastic and paper can also be cut by the carbon dioxide laser. Figure 9 shows a piece of 0.25-mm-thick paper, which was cut with a 70-watt laser by moving it at a rate of about 15 cm/s past the focus of a mirror having a focal length of 18 cm. Even though 60-Hz current rather than filtered direct current was used for the excitation, the edges of the slit are smooth and little charring is present. Because the heat input from the laser is extremely localized and rapid, very little conduction to the adjacent material takes place and the formation of a bead is minimized when plastic sheet is cut.

It seems reasonable to expect that for thin samples such as these the rate of slitting would increase linearly with laser output power. Thus a 1-kW laser might be expected to cut 0.25-mm-thick paper at a rate of 216 cm/s. The use of filtered dc for the driving current would be expected to give a smooth edge to the cut.

Much work remains to be done to determine the kinds of materials and cutting operations that can be effectively handled with the CO_2 laser.

The advantages of using the continuous carbon dioxide laser for cutting are:

1. As in the case of pulsed-laser hole punching there is no contact between the cutting tool and the workpiece, so that tool wear is eliminated and jigging, fixturing, and material transport are simplified.
2. By the use of mechanical or electronically controlled techniques to move the mirrors, the output of the laser can be deflected rapidly. Thus it is possible, for example, to cut sheet plastic to length without having to stop its motion for the cutting operation.
3. Rapid cutting rates can be obtained.
4. Bead formation at the edges of cuts in thermoplastic materials is minimized.
5. The depth of the cut produced can be varied rapidly by changing the output of the laser.

Welding using continuous lasers. To date, very little has been done to investigate the possibility of using the output of the carbon dioxide laser to weld thermoplastic materials. Very limited work on lucite has shown that welds can be produced; however, the strength of the welds and possible advantages over other techniques have not yet been determined.

Low-power continuous helium–neon gas lasers. Neon gas can be excited by collisions with helium in very much the same way as carbon dioxide is excited by collisions with nitrogen molecules. The energy levels for helium and neon are shown in Fig. 10. The helium atoms in a tube filled with helium and neon gas are excited by collisions with electrons and transfer their excitation to the neon atoms fast enough to result in a population excess in the 3*s* and 2*s* levels relative to the 3*p* and 2*p* levels. By choosing the wavelength dependence of the reflectivity of the mirrors, continuous oscillations can be obtained at any one of several wavelengths.[22] The most useful ones are 0.63 and 1.15 μm. A helium–neon laser can be excited as shown in Fig. 11. Light with a wavelength of 0.63 μm is detectable by the eye. Commercially available 1-mW 0.63-μm helium neon dc-excited lasers have a size of about 30 by 20 by 13 cm and weigh less

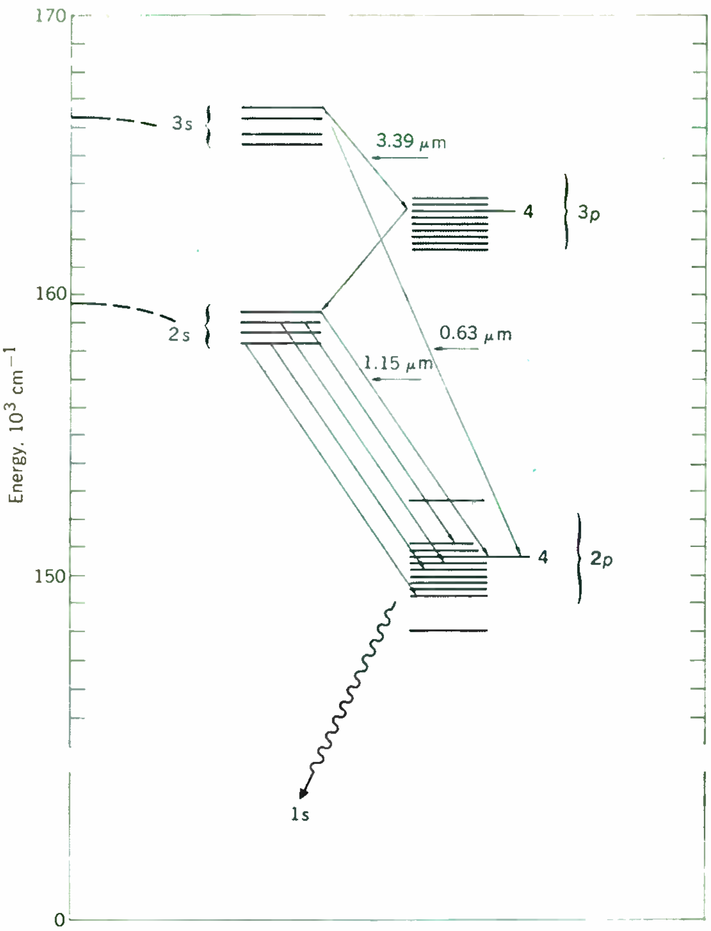
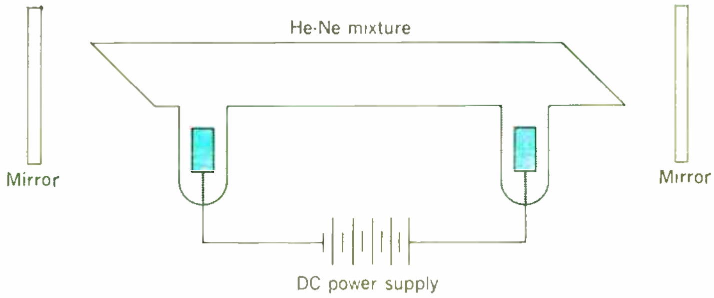


FIGURE 10. Helium–neon energy levels.

FIGURE 11. Helium–neon dc-excited gas lasers.



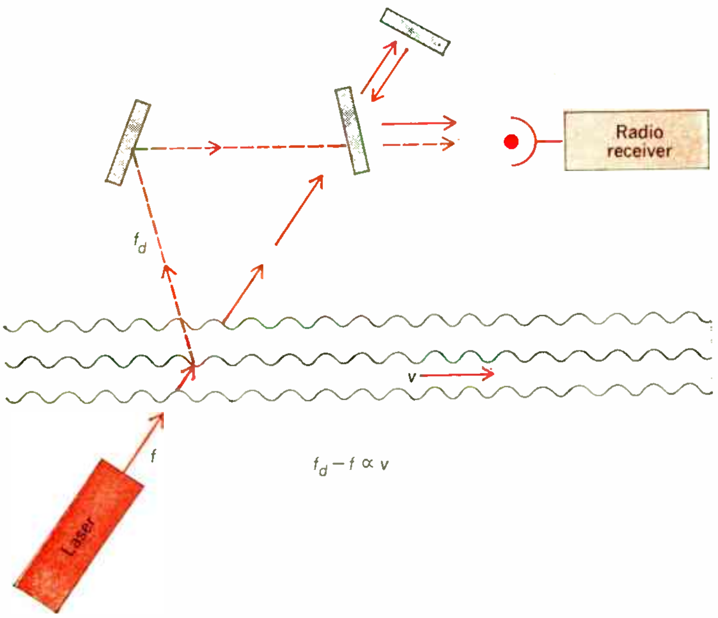


FIGURE 12. Laser measurement of fluid-flow velocity.

than 16 kg. Operation for several months without maintenance is not unusual.

The helium–neon laser is useful not only because of its monochromaticity but also because of the high directivity of its output. The beam divergence of a 1-mW visible helium–neon laser is typically 10 mrad. Other slightly larger and more sophisticated helium–neon lasers have beam divergences as small as 1 mrad. Using a collimating telescope in conjuction with a 1-mrad laser can yield a beam spread as small as 2×10^{-5} rad. In other words, a beam 5 cm in diameter will spread only by 0.5 cm in 300 meters. One application for this high directivity involves the use of the laser beam as a reference line for such operations as tunnel drilling.

Because of its high monochromaticity the helium–neon laser can be employed to produce interference between light reflected by surfaces that are separated by relatively large distances. This capability suggests such applications as interferometric control of machine tools.

Doppler-shift applications. One of the more novel applications of the helium–neon laser utilizes the Doppler frequency shift, which results from the scattering of light by moving objects.[23] For measurements of fluid-flow velocity, as shown in Fig. 12, the output of a laser at frequency f is directed through the transparent fluid. Laser radiation that is scattered by particles or inhomogeneities in the fluid will have its frequency shifted to a new value f_d, which differs from that of the unscattered radiation by an amount proportional to the velocity v of the scatterer. The scattered and unscattered radiation are caused to impinge simultaneously on a photoelectric emitter. Since the number of electrons N_e emitted per unit time is proportional to the intensity (that is, power per unit area) of the total optical radiation, N_e is also proportional to the square of the electric field of the optical electromagnetic field. Consequently, a photoelectric surface can be used as a nonlinear element to mix two optical signals having different frequencies. The electron current generated this way will contain a frequency component corresponding to the difference between the two frequencies of the optical signals. The difference frequency electron current can be amplified and its frequency measured by conventional electronic techniques. In the laser velocimeter the measured difference frequency is just the Doppler shift, which in turn is directly proportional to the velocity of the fluid. For a velocity of 5 cm/s the frequency shift is about 20 kHz. Shifts of this order of magnitude are easily detected. The laser velocimeter has also been used to measure the velocity of solid sheet material, such as paper and aluminum.[24]

Current measurement. Engineers at Tokyo Electric Power and Tokyo University in Japan have recently described the use of a gas laser to measure the current in 500-kV transmission lines.[25,26] The current is determined by measuring the amount of rotation of the plane of polarization of a beam of light as shown in Fig. 13. The light beam in their apparatus is the output of a helium–neon laser located at ground potential. Since apparatus used to measure the rotation of the polarization is also at ground potential, the only parts of the system that are close to the high-voltage wire are two mirrors and a piece of flint glass. This device has been tested with 10 000-ampere 1.5-MV impulse current, and the response time can be as short as a few microseconds. The device can measure currents between 30 and 30 000 amperes. The Japanese say that this unit "is compact and costs less" than the conventional types of current transformer used to monitor EHV power transmission lines.

Holography. The monochromaticity and high intensity of laser radiation make possible a new form of three-dimensional imaging, called holography.[27] A hologram is made by illuminating an object and a reference mirror simultaneously with the light from a laser. A photographic film is used to record the interference pattern between light reflected from the object and the mirror. After the film is developed, one can view the hologram by looking toward a monochromatic light source (not necessarily a laser) through it. The image one views is a monochromatic reconstruction of the original object, which appears to be located at the same place with respect to the hologram as was the original object. In addition to their obvious potential as advertising displays, holograms may also be useful for three-dimensional computer outputs, training devices that use three-dimensional replicas of actual objects, masks for exposing photoresists in the manufacture of microcircuits, and measurements of small surface strains in machine parts. A helium–neon gas laser with an output at 6328 Å is commonly used for these applications. For applications requiring shorter exposure times, however, the argon gas laser is used instead since it is capable of several watts of output at a wavelength (about 5000 Å) that is more effective in exposing film than is the helium–neon output.

Inspection applications. The monochromaticity and high intensity of laser radiation also make practical optical inspection techniques that were only conceptually possible with conventional light sources. One example is the use of a helium–neon laser to inspect the pitch of a cylindrical spring while it is being wound.[28] Illuminating a spring with the laser radiation produces a diffraction pattern consisting of a number of parallel lines, the spacing between which is inversely proportional to the pitch of the helix. By cutting holes in a mask at the positions of the

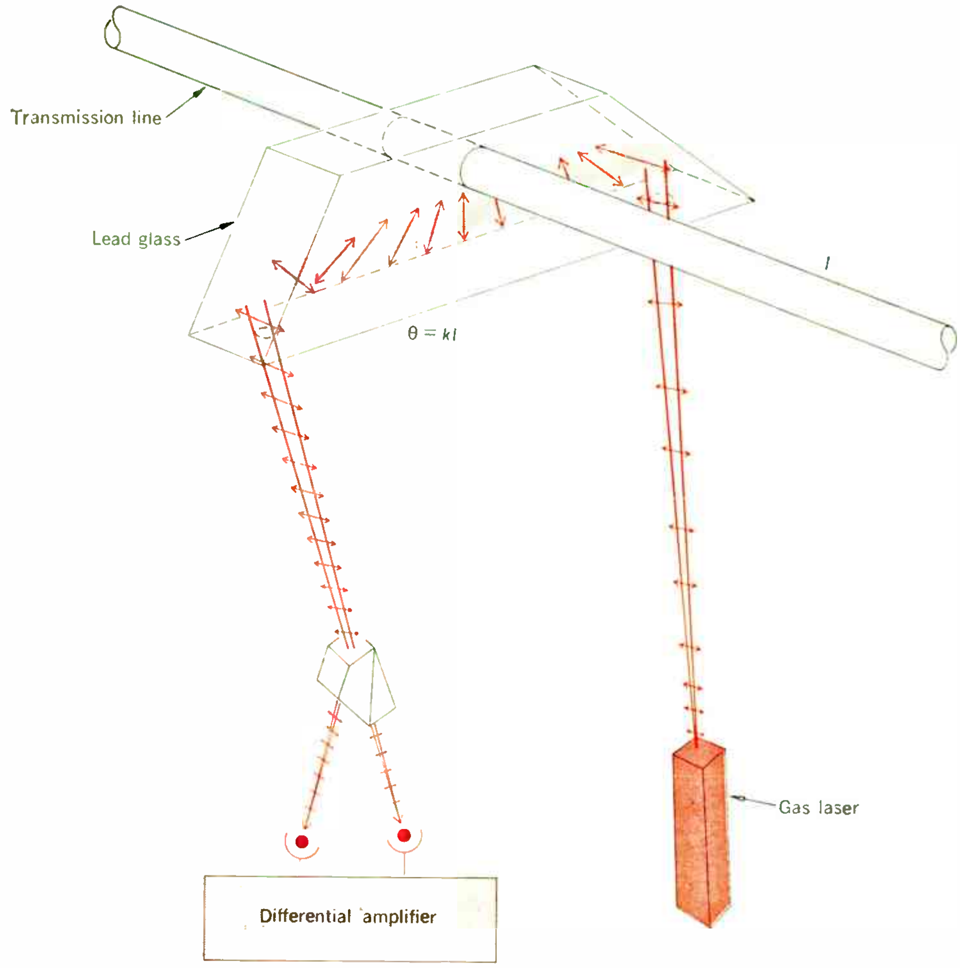


FIGURE 13. Laser current monitor for measuring current in high-voltage transmission lines. The angle of rotation θ is proportional to the current.

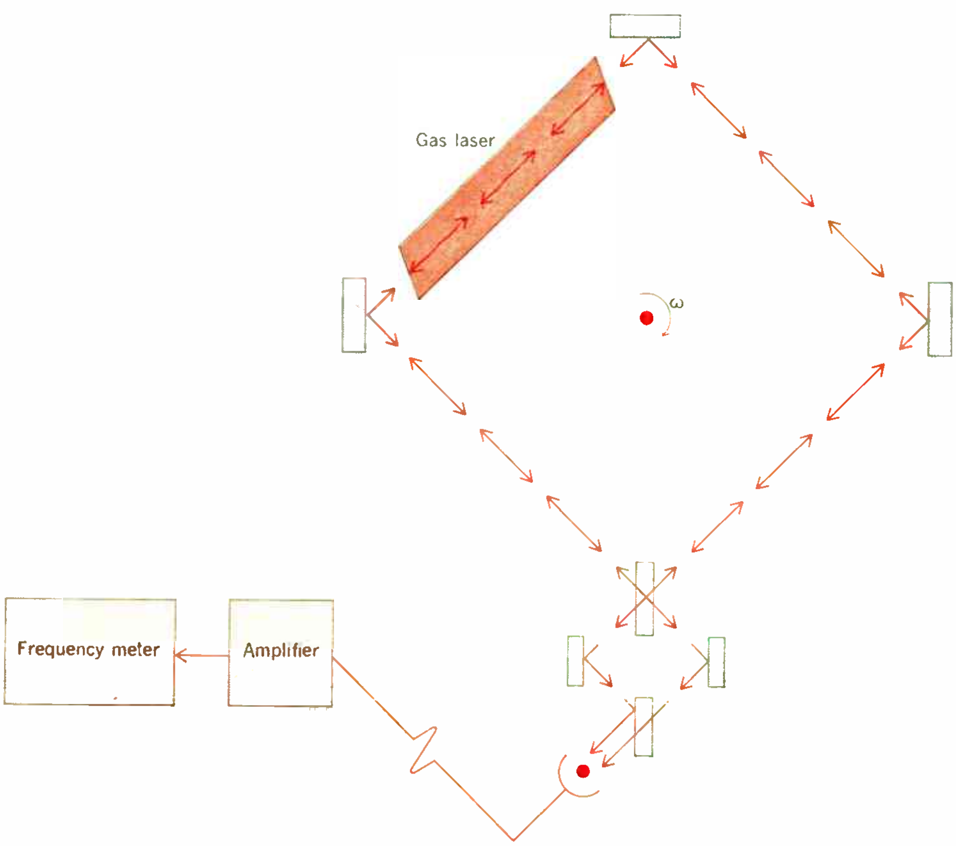


FIGURE 14. Typical prototype gyroscope employing the laser principle as the basis for its operation.

diffraction maxima corresponding to the desired pitch, one can automatically monitor the pitch; by using more holes to indicate when the pitch has become too tight or too loose, one can servo the winding machine to maintain the desired value.

Laser gyroscopes. If a laser is used in a closed optical resonator the rate of rotation of the resonator platform can be measured without the need for reference to any external object.[29] Prototype gyroscopes have already been demonstrated using this principle (Fig. 14).

Communications on laser beams. Because optical radiation has a frequency of more then 10^{14} Hz the available information bandwidth is very large in comparison with, for instance, microwave systems, which have carrier frequencies of about 10^{10} Hz. This frequency range provides the attractive possibility of transmitting many more messages over a single carrier network, but its practical realization will depend on the development of many new techniques for modulating, guiding, amplifying, and detecting light.[30]

Air-pollution measurement. The use of lasers to detect small amounts of contaminants in the atmosphere has been proposed.[31]

Obstacle avoidance for the blind. Instruments employing small pulsed injection lasers (using a principle of excitation not discussed in the preceding portion of this article) are being developed to warn blind people of obstacles a few steps in front of them.[32] These devices measure distance by triangulation of the reflected beam relative to the transmitted beam. The laser has the advantage of high directivity in a small device so that very small changes in height (such as curbs) can be detected.

Other types of lasers

Since there are many different types of lasers, choice for inclusion in this article was made on the basis of relevance to current nonmilitary applications and on the likelihood of availability of equipment for use in these applications. There are more than 50 different laser materials and many of them are capable of operating at any one of several output wavelengths.[33] In addition to the methods of excitation described in this article, laser action has been produced by the current flowing through a semiconductor junction and by exciting a semiconductor with an electron beam. Laser action has also been achieved during the chemical combination of two gases.

Conclusion

This article is intended to generate curiosity rather than satisfy it. It is the author's hope that the article will provide a general understanding of what the laser is and what the properties of its output are. It is also hoped that the brief descriptions of these applications will suggest other potential uses for lasers. Those who wish to learn more about how lasers work, and their characteristics, are urged to consult one of the many excellent review articles or books.[3–5,33] The feasibility of using lasers for some specific application is most easily determined by consulting those engaged in laser research, and questions of this type are generally welcomed by them.

REFERENCES

1. Young, C. G., *et al.*, "Optical avalanche laser," *J. Appl. Phys.*, vol. 37, p. 3419, Nov. 1966.

2. Jaseja, J. S., Javan, A., and Townes, C. H., "Frequency stability of He-Ne masers and measurements of length," *Phys. Rev. Letters*, vol. 10, p. 165, 1963.

3. Lengyel, B. A., *Introduction to Laser Physics*. New York: Wiley, 1966.

4. Birnbaum, G., *Optical Masers*. New York: Academic Press, 1964.

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

6. Epperson, J. P., Dyer, R. W., and Grzywa, J. C., "The laser now a production tool," *Western Elec. Engr.*, vol. 10, p. 7, Apr. 1966.

7. Campbell, C. J., *et al.*, "Laser photocoagulation of the retina (a critical evaluation)," *Klin. Monatsbl. Augenheilkunde*, vol. 149, p. 636, Dec. 1966.

8. Goldman, L., "Applications of the laser beam in cancer biology (a review)," *Internat'l J. Cancer*, vol. 1, p. 309, July 15, 1966.

9. McGuff, P. E., *et al.*, "Laser radiation of malignancies," *Ann. N.Y. Acad. Sci.*, vol. 122, p. 747, May 28, 1965.

10. Minton, J. P., "Experimental results from exposure of Cloudman S-91 melanome in the CDBA/2F hybrid mouse to neodymium or ruby laser radiation," *Ann. N.Y. Acad. Sci.*, vol. 122, p. 758, May 28, 1965.

11. Franken, P. A., Jenny, J. A., and Rank, D. M., "Airborne investigations of clear-air turbulence," *IEEE J. Quantum Electronics (Digest)*, vol. QE-2, p. lix, Apr. 1966.

12. Bishop, A. S., *Project Sherwood*. Reading, Mass.: Addison-Wesley, 1958; see particularly the figure on p. 169.

13. Basov, N. G., and Krokhin, O. N., "The conditions of plasma heating by the optical generator radiation," *Soviet Phys. JETP*, vol. 21, p. 1061, 1965.

14. Englehardt, A. G., "Generation of dense high-temperature deuterium-tritium plamas by means of coherent optical radiation," presented at a meeting of the American Physical Society, Philadelphia, Pa., Mar. 1964.

15. Ambartsumyan, R. V., *et al.*, "Short-pulsed *Q*-switched laser with variable pulse length," *IEEE J. Quantum Electronics*, vol. QE-2, pp. 436–441, Sept. 1966.

16. Compagnie Général d'Electricité, Data sheet on Unit VNDM 1600.

17. Johnson, L. F., Gensic, J., and Van Vitert, L. C., "Efficient high-power coherent emission from Ho^{3+} ions in yttrium aluminum garnet, assisted by energy transfer," *Appl. Phys. Letters*, vol. 8, p. 200, 1966.

18. Liberman, I., and Church, C. H., Private communications.

19. Armstrong, J. A., "Measurement of picosecond laser pulse widths," *Appl. Phys. Letters*, vol. 10, p. 16, Jan. 1, 1967.

20. Ehrilich, J. J., Hutcheson, G. J., Hales, W. L., and Roberts, T. G., "High-power N_2-He-CO_2 laser development," presented at IEEE Conf. on Laser Engineering and Applications, Washington, D.C., June 8, 1967.

21. Whitehouse, D. R., "High-power CO_2 laser," *1966 NEREM Record*, p. 190.

22. White, A. D., and Ridgen, J. D., "Continuous gas maser operation in the visible," *Proc. IEEE (Correspondence)*, p. 1697, July 1962.

23. Foroman, J. W., *et al.*, "Fluid flow measurements with a laser Doppler velocimeter," *IEEE J. Quantum Electronics*, vol. QE-2, pp. 260–266, Aug. 1966.

24. Flower, R. A., "Industrial speed and length measurement using a laser velocimeter," *IEEE J. Quantum Electronics (Digest)*, vol. QE-3, p. 263, June 1967.

25. Saito, S., *et al.*, "The laser current transformer for EHV power transmission lines," *IEEE J. Quantum Electronics*, vol. QE-2, pp. 255–259, Aug. 1966.

26. Saito, S., *et al.*, "Developments of the laser current transformer for extra-high-voltage power transmission lines," *IEEE J. Quantum Electronics*, vol. QE-3, pp. 589–597, Nov. 1967.

27. Leith, E. N., and Upatnieks, J., "Wavefront reconstruction with diffused illumination and three-dimensional objects," *J. Opt. Soc. Am.*, vol. 54, p. 1295, Nov. 1964.

28. Brandt, G. B., and Asars, J., Private communications.

29. Killpatrick, Joseph, "The laser gyro," *IEEE Spectrum*, vol. 4, pp. 44–55, Oct. 1967.

30. Miller, S. E., and Tillotson, L. C., "Optical transmission research," *Proc. IEEE*, vol. 54, pp. 1300–1311, Oct. 1966.

31. Jacobs, G. B., and Snowman, L. R., "Laser techniques for air pollution measurement," *IEEE J. Quantum Electronics (Digest)*, vol. QE-3, p. 267, June 1967.

32. Bolgiano, D. R., and Meeks, E. D., Jr., "A laser cane for the blind," *IEEE J. Quantum Electronics (Digest)*, p. 268, June 1967.

33. See *Proc. IEEE*, vol. 54, Oct. 1966 for a number of survey papers on lasers.

The day Albert Jones conceived a new I R test technique

Technically, it was the day after he and Frank Porter, AGE major domo, agreed there was a problem. No combination of available components could field-test the aircraft IR sensor practically. "Well," said Al, "looks like we start from scratch." Which understated the problem, since they'd only 6 months to develop a working prototype. Al left Frank with a "Don't worry". But he'd already started worrying.

At home, Al popped the family off to bed after an evening's TV . . . and pondered. Everything seemed to hinge on the weight. Even latheways cast in magnesium or aluminum clocked in at 1500 lbs.; available sensor-holding units hit the scales at 300+. What was needed would have to weigh far less. So, he gave it considerable thought.

With considerable success. Resulting in the basic parameters for a portable, air-transportable, autocollimated, lab quality optical bench. Capable of ultra-narrowband spectral analysis; sensitivity checkout; testing resolution and total field of view, with micrometer sensor positioning and a built-in substitute cryo-source. Total weight: 300 lbs!

Funny thing about Al. He doesn't consider what he did especially unusual. It's typical of the all-of-a-sudden problems that crop up in developing multi-function AGE systems whose frequencies run from DC to light. And, he's gotten used to pioneering ideas like automated test equipment using a time-shared central computer. He likes it. All of it.

If there are any more of you out there like Al, why not investigate what our AGE laboratory can offer you. Even if you'd prefer to pit your intellect against a provocative study or development program in another advanced area, we'd like to hear from you: we'll tell you about ongoing programs in our radio communications, navigation aids, data equipment, countermeasures and tracking systems labs.

Direct your resume, in confidence, to Mr. J. P. O'Reilly, Dept. 171.

GENERAL DYNAMICS

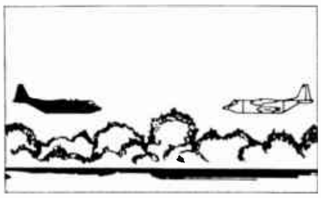
Electronics Division

1400 N. Goodman Street, Rochester, New York 14601

We are an equal opportunity employer in deed, not only an equal opportunity employer in words.

Scanning the issues

Collision Avoidance Systems. For more than a decade, airlines have been aiming at developing a system that will enable aircraft to avoid what seems about to happen in this picture.



In fact, although the scheduled airlines established the goal of providing a collision avoidance system (CAS) in 1955, it was not until a disastrous midair collision that occurred over the Grand Canyon in 1956 that real concern was manifested in a flood of ideas submitted to the airlines industry organization, the Air Transport Association (ATA). An earlier request by the ATA for proposals from engineers, inventors, and manufacturers for collision avoidance systems had brought no response. The log of these events from 1955 up to the present appears in "Airborne Collision Avoidance System Development—Introduction," by F. C. White, which forms the introduction to a complete set of nine papers on aspects of a CAS system. These interesting papers, which reveal the intense investigations that have been going on and which affect a vital interest of all those who fly, appear in the current issue of the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.

The investigations have reached the point where a CAS system using stable time and frequency technology has been defined jointly by the airlines and selected manufacturers, and it is now expected that several manufacturers will provide airborne systems for flight evaluation during the first quarter of 1969.

A few highlights from White's log: In July 1956, ATA brought together airline experts, engineers, and inventors to compare airline requirements with the then-current technology. By September 1956, Collins Radio submitted a proposal for a system thought to be the kernel for a CAS system, but subsequent development showed that its airborne Doppler radar would create erroneous collision predictions. Analytical work continued, and continues to this day. In 1960, a collision between two fighter aircraft at McDonnell Aircraft sparked an all-out effort by that company to develop a CAS for use in its isolated flight test area. In July 1962, Collins outlined at the Airborne CAS Symposium what was to be a major milestone, namely, a method for testing the features of proposed CAS techniques. For the first time, it offered a quick and minimum cost method for choosing the most promising CAS techniques from among a wide variety of theoretical concepts. In 1963, Collins got a contract to study CAS techniques by simulation, a study that by 1965 indicated that a time–frequency system was the most promising that would work well for the case of an airliner en route though the system showed some limitations in terminal areas where aircraft are most likely to be maneuvering in flight. In 1966, the ATA circulated widely throughout industry and government a detailed listing of functional requirements for a CAS. By June 1967, the ATA released a requirements paper containing a "Technical Description of the CAS System." What was important about this description was that it was the result of intensive cooperative investigations by manufacturers, experts in either time and frequency or airline avionics systems. It represented, in fact, according to White, a unique pooling of knowledge in the competitive world inasmuch as manufacturers disclosed results of company-sponsored research.

A Technical Working Group formed by ATA had to select the most suitable technology for the CAS system on the basis of its understanding of four major and basic facets of CAS design. These four design points, which are discussed in quite some detail in the set of nine technical papers, include: (1) the method of time synchronization; (2) the format of the CAS message; (3) the signaling technique; and (4) the protected volumes and related warning times.

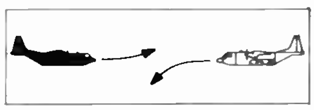
In "History of Time–Frequency Technology," M. R. Bates and his colleagues report that time and frequency technology is reasonably well advanced and has already undergone several generations of development for use by airborne "stationkeepers." They discuss the differences between stationkeepers and the air-to-air collision avoidance problem.

The time synchronization technique adopted for the CAS employs start-up from ground stations using a single master time, with air-to-air relay of the master time to extend outward the influence of the ground master time. The functioning and accuracy of the synchronization technique is discussed in "Collision Avoidance System Synchronization" by R. L. Jaycox, and the maintenance of a single master time by the ground stations is discussed by C. O. Thornburg in "Master Timing of CAS Ground Stations."

White points out in his introduction that the CAS system will be unique in the communication world. It will organize all the aircraft CAS participants into a sequential communications environment in which each aircraft transmits in turn while all others listen. With air space at a premium and traffic densities growing, it is most important that the CAS signals be such that the likelihood of not receiving the data necessary to identify a collision threat be extremely low. In the CAS message, these data are the range, range rate, and altitude. Altitude rate and velocity data may be nearly as important as the "big three," and these, too, have been provided in the message format. R. E. Perkinson discusses the "CAS Message Format."

Aspects of the CAS signaling technique are discussed in three papers, by K. E. Toerper, C. E. Steen, and E. A. Steinberg. Elements of the CAS system are discussed by W. G. Shear. The maze of time allowances inherent to the functioning of the CAS system is treated by J. M. Holt and R. M. Anderson.

In all, the nine papers filling nearly a hundred pages brings to the interested reader a book on the subject of CAS, which, White concludes, will in time be used by all aircraft that share air space regardless of when or where they fly.



(*IEEE Trans. on Aerospace and Electronic Systems*, March 1968.)

Quartz Crystal Resonators. What used to take six months now takes less than a week. This remark by the editor of the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT refers to an improvement in quartz oscillators reported in a paper by R. J. Byrne and J. L. Hokanson.

Up till now, quartz oscillators have been relegated to second place as frequency standards by cesium oscillators. However, these authors have studied the influence of contamination level on the aging behavior of quartz crystal resonators, and they have introduced thermocompression bonding and high-vacuum techniques in the manufacture of quartz crystals in such a way as to minimize the level of contamination. The result has been a significant reduction in the rate of frequency drift of the crystal, and a reduction in the time required for the crystal to stabilize its drift rate.

The need for very stable frequency sources in many diverse applications has led to more stringent requirements on the stability of quartz-crystal-controlled oscillators. The frequency of these precision oscillators is controlled by a piezoelectrically excited quartz plate maintained at constant temperature and amplitude of vibration, and usually operated on the fifth overtone of the thickness shear vibration mode. The best oscillators to date approach a long-term frequency drift rate of the order of 1 to 3 parts per 10^{10}/month after an initial stabilization period of from one to six months of continuous operation. If continuous operation is not maintained, due to power or component failure, the reestablishment of a stable frequency again requires from one to six months. It is this stabilization and restabilization period that is markedly reduced to less than a week for the units made by the new technique. (R. J. Byrne and J. L. Hokanson, "Effect of High-Temperature Processing on the Aging Behavior of Precision 5-MHz Quartz Crystal Units," *IEEE Trans. on Instrumentation and Measurement*, March 1968.)

Radiation and Semiconductors. The effects of radiation on semiconductor devices has been a subject of keen interest for many years now. In a recent paper, it is pointed out that nuclear reactor radiation is frequently employed in studies of the radiation tolerance of semiconductor devices. In addition to neutrons, ionizing radiation is also present in a nuclear reactor. Fast neutrons give rise to permanent *bulk* effects—lifetime decrease and carrier removal. Ionizing radiation gives rise to permanent *surface* effects—surface recombination velocity increase and positive charge buildup within the oxide of planar devices. In radiation studies that employ nuclear reactor radiation, it is necessary to know the extent to which surface effects are responsible for the observed changes in device characteristics.

Through studies of surface effects on planar devices, reported in the current issue of IEEE TRANSACTIONS ON ELECTRON DEVICES, it has become possible to separate unambiguously surface radiation effects from bulk radiation effects and to compare their relative importance. This kind of comparison is made by the authors of this paper. They made radiation studies on several types of surface and bulk semiconductor devices and separated the effects. In brief, their work shows that at high dosage levels the primary effect of nuclear radiation is on the bulk and not on the surface properties of the semiconductor devices. (D. J. Fitzgerald and E. H. Snow, "Comparison of Surface and Bulk Effects of Nuclear Reactor Radiation on Planar Devices," *IEEE Trans. on Electron Devices*, March 1968.)

Information and Evolution. An index of the increasing engineering interest in aspects of evolution appears in a recent paper that investigates how information might be transmitted from the environment to an evolving species. The study is based on a model previously established for the random-mutation portion of the synthetic theory of evolution.

The author of the study notes that living organisms are equipped with many complex devices and instinctive behavior patterns that allow them to survive and to take advantage of some aspects of their environment in order to live and reproduce. Such adaptation is generally understood to result from hereditary changes over many generations. Any theory of evolution must postulate a mechanism whereby such changes occur.

The genetic information imbued in an organism the author calls the *genetic message*. It is theorized that this message evolves, to provide better adaptation of the organism to the environment, through the transmission of information from the environment into the genetic storage. The probability that a given adaptive character is produced decreases with increasing information measure of the adaptive character. The average information transmitted in any evolutionary step (the product of the information and its probability of transmission) turns out to be very small for a reasonable number of trials and gene size. This is consistent with the conjecture that evolution proceeds in small steps, and the size of such steps can be discussed quantitatively in terms of adaptive information. Using methods described, it should be possible to estimate the probability of achieving a particular character within a large number of adaptive steps. (L. M. Spetner, "Information Transmission in Evolution," *IEEE Trans. on Information Theory*, January 1968.)

What's a Man–Machine? The members of the Human Factors in Electronics Group have been uncomfortable with their name for a long time. Now they have done something about it. They have changed their name to Man–Machine Systems Group. To mark this significant occasion, the editor of the TRANSACTIONS, Thomas B. Sheridan, muses on the question of what people see in names. It is worth reading. Not every field of engineering endeavor has had to contend with such an awkward appelation as "Human Factors," but the tide of interdisciplinary investigations sooner or later brings almost every specialty to some form of low-grade identity crisis. So Sheridan's thoughts have some relevance for us all.

His definition of the relation of the Man–Machine Systems Group to other groups within and without IEEE concerned with physiological and psychological factors is particularly worth repeating here. It would seem, he says, that there are really three such IEEE groups, each operating at a different level. The Engineering in Medicine and Biology Group seems to be dealing with part of a man (an internal organ, a cell, etc.) interacting with electromechanical machines, a part-of-a-man–machine system. At a multiman–multimachine level, embracing economics, politics, etc., there is the Systems Science and Cybernetics Group. We, the Man–Machine Systems Group, Sheridan goes on, are right in the middle at the one-man–one-machine level. We ought to be cooperating more with these other two groups and be more conscious of our potential leverage and mediating capacity as middleman between levels as well as between social and life science and engineering disciplines.

Sheridan does not fail to ask whether or not the name "Man–Machine Systems" is wholly satisfactory or is a fad. To counter such doubts, one might observe that at least on grounds of euphony the Group is more happily established. And, besides, what matters is that the work goes on. If a new name helps, then by all means change the old one. (T. B. Sheridan, "What's a Man–Machine?" *IEEE Trans. on Man–Machine Systems*, March 1968.)

Advance abstracts

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

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

Proceedings of the IEEE

Vol. 56, no. 5, May 1968
M—$1.00; L—$1.50; NM—$2.00

Voltage Variable Capacitor Tuning: A Review, *M. H. Norwood, E. Shatz*—The history, device theory, characteristics, applications, and future trends of voltage-variable capacitor tuning are discussed. All equations are stated in terms of two general exponents of power law functions, namely, the impurity distribution proportional to x^m and the differential capacitance proportional to $(V + V_0)^{-n}$. The role of these exponents is shown in the device theory, the temperature drift, linearity in a VCO, and the cross modulation in an FM modulator. Important results from many papers and reports are cited to reinforce the ideas presented. The desirability of the exponent $n = 2$ is reiterated throughout.

A Model for Electromagnetic Propagation in the Lithosphere, *F. K. Schwering, S. W. Peterson, S. B. Levin*—The propagation properties of a crustal waveguide are derived using a model in which dielectric constant and conductivity increase with depth according to a simple algebraic relation. The model is more realistic and consistent with currently available geological information and with reasonable extrapolations of crustal properties than are previously treated models. The mathematical analysis, moreover, is straightforward and rigorous, leading to an explicit analytical solution in the form of a mode expansion, the individual terms of which can be written in closed form. Numerical evaluation of the model leads to a total attenuation rate (for the dominant mode) of 0.256 dB/km for a frequency of 1 kHz. A geologically critical attribute of the model is a positive temperature gradient with depth as the principal controlling influence on the electric profile. A temperature gradient lower than the one used in constructing the model would significantly reduce the attenuation rate.

Analysis and Improvement of Photostore Error Rates, *R. T. Chien, D. T. Tang, E. S. Barrekette, A. M. Katcher*—Statistical experiments have been performed on the photostore to test its reliability as a mass storage system. Errors are analyzed with respect to rate, distribution, and susceptibility to correction by several codes. The experimental results reported are based on disks recorded by optical techniques. Experimental results on disks recorded with electron beams will be reported separately.

Digital Computer Calculation of the Electric Potential and Field of a Rod Gap, *M. S. Abou-Seada, E. Nasser*—The electric field and potential distribution in the gap between a cylindrical rod having a hemispherical tip and an infinite plane perpendicular to the cylinder axis was determined using a charge simulation technique. This method assumes a charge at the center of the hemisphere and a finite number of semi-infinite axial charges in the cylindrical portion of the rod electrode. Boundary conditions in the cylindrical and spherical portions permitted the formulation of simultaneous equations whose digital solution yielded the assumed lumped charges. The digital computer program provided values of the potential and both field components anywhere in the gap with an accuracy of about 2 percent.

Stability of Linear Multivariable Feedback Systems, *C.-T. Chen*—The stability of linear time-invariant multivariable feedback systems is studied from their open-loop transfer function matrices. It is shown that the stability of a multivariable feedback system depends on the determinant of the loop-difference matrix (det $(\mathbf{I} + \mathbf{G}_1\mathbf{G}_2)$) and the characteristic polynomials of its open-loop transfer function matrices. The controllability and observability properties of the feedback system are considered. Hence, the stability conditions insure the stability at the output terminals as well as at the state variables of the system. These conditions can be easily checked by using the Nyquist plot, the root locus technique, or the Routh–Hurwitz criterion.

Sea Echo at Laser Wavelengths, *A. V. Jelalian*—Experimental data on sea echo at a 1.06-micrometer laser wavelength at normal incidence are discussed. The experimental data over a variety of sea states indicate that the return signal is definitely range square dependent, and that the effective target cross section ($\sigma°$) for smooth and rough surface water conditions is 0.15 ± 0.08 and 0.24 ± 0.12, respectively. Additional considerations concerning laser spot size at the ocean and its effect on signal returns are discussed.

State of the Art in Pattern Recognition, *G. Nagy*—Statistical, adaptive, and heuristic techniques used in laboratory investigations of pattern recognition problems are reviewed. The discussion includes correlation methods, discriminant analysis, maximum likelihood decisions, minimax techniques, perceptron-like algorithms, feature extraction, preprocessing, clustering, and nonsupervised learning. Two-dimensional distributions are used to illustrate the properties of the various procedures. Several experimental projects, representative of prospective applications, are also described.

Proceedings Letters

Because letters are published in PROCEEDINGS as soon as possible after receipt, necessitating a late closing date, we are unable to list here the letters in the May issue. This will appear in the next issue of SPECTRUM. Listed below are the letters from vol. 56, no. 4, April 1968.

Electromagnetics and Plasmas

Comments on "Potential Equations for Anisotropic Inhomogeneous Media," *A. Ishimaru, D. A. Ross*
Comment "On the Thermodynamic Paradox in Ferrite-Loaded Waveguides," *J. Brown, F. E. Gardiol*
Comment on "A Theorem in the Field of Steady Current Flow," *R. K. Arora*
Measurement of Antenna Current Distributions in a Hot Plasma, *H. Judson, K.-M. Chen*
Notes on Propagation in a Time-Space Harmonic Dispersive Medium, *J. Bruno, C. Shulman*
Radiation from a Charge in a Uniform Circular Motion, *S. R. Seshadri*
Multipacting Breakdown in Coaxial Transmission Lines, *R. Woo*
Microwave Measurements of Transmission and Reflection Properties of Symmetrical Two-Terminal-Pair Structures, *M. Fitaire*

Circuit and System Theory

A Stable, Accurate Method of Numerical Integration for Nonlinear Systems, *D. A. Calahan*
Addendum to "On the Linear-Slope Delay Approximation," *J. Valand*
Linear Two-Port Characterization Independent of Measuring Set Impedance Imperfections, *J. G. Evans*
Comment on "Counting Complex Roots in Polynomials with Real Coefficients," *A. Lepschy, A. J. Calise*
Synchronization of Pulsed Oscillators, *F. M. Magalhaes, W. O. Schlosser*
Compensating Negative Converters, *N. Balbanian, C.-K. Kuo, K. L. Su*

Electronic Circuits and Design

Practical Design Information for Broadband Transmission Line Transformers, *O. Pitzalis, Jr., T. P. M. Couse*
FM Detection Using a Product Detector, *A. Bilotti*
Correction to "Comment on 'The Monostable Tunnel Diode Trigger Circuit,'" *M. A. Schapper*

Electronic Devices

Gain in Solid-State Traveling-Wave Amplifiers, *M. Ettenberg, J. Nadan*
Mesh Emitter Transistor, *M. Fukuta, H. Kisaki, S. Maekawa*
Analytical Representations of Solid-State Devices, *J. S. Linder*
Effect of the Intervalley Scattering Time on LSA Oscillations, *T. Ohmi, K. Murayama, H. Kanbe*

Classified Advertising

Positions Open

The following positions of interest to IEEE members have been reported as open. Apply in writing, addressing reply to address given or to Box Number, c/o IEEE Spectrum, Advertising Department, 72 West 45th St., New York, N.Y. 10036.

Classified Advertising Rates for this column: $6.00 per line. No advertising agency commission is granted. Copy must be received by the 2nd of the month preceding date of issue.

Faculty positions open in Electronics Technology Dept. at William Rainey Harper College in northwest suburb of Chicago. BSEE and MS (Education or Technical) preferred. Must assume responsibility for teaching of technical and math courses in the electronics area. Salary: $7,000–$17,200. Reply to: Dr. H. R. Pankratz, Harper College, 510 W. Elk Grove Blvd., Elk Grove Village, Ill. 60007.

Research Engineers and Instructors to work for Ph.D. Research projects include: Radar for space, automatic image processing, remote sensing, analog and hybrid simulation, biotelemetry, medical electronics, solid state device studies, electromagnetic theory and antennas. Send resume to: Chairman, Department of Electrical Engineering, University of Kansas, Lawrence, Kansas 66044. An equal opportunity employer for U.S. citizens.

Electrical Engineer. Opening for graduate electrical engineer with experience in public utility transmission and distribution system operations and construction. Emphasis on planning and economics. Excellent opportunity for advancement into management of growing utility company. Desirable suburban office location close to New York City. Please forward complete resume, including salary history, to Box 6103.

Power Plant Project Managers—Established East Coast engineering-construction firm has long term staff positions open for graduate electrical or mechanical engineers with 12 years minimum experience in design of steam-electric power stations including project inter-office and client coordination. Preference given applicants having design experience on units greater than 400 MW capacity. Salary commensurate with experience and relocation expenses paid. P.E. license preferred. Reply giving full details including personal data, education, experience, references, and salary requirements to, Mr. C. D. Millar, The Kuljian Corporation, 1200 North Broad Street, Philadelphia, Pennsylvania 19121. An equal opportunity employer.

Teaching and Research Position in EE. Bioelectronics Engineering/Bioinstrumentation. Joint appointment in Department of Electrical Engineering and Institute for Environmental Research. Teaching at graduate level and independent research. Salary commensurate with experience. Reply to: Ralph G. Nevins, Dean, College of Engineering, Kansas State University, Manhattan, Kansas.

Electric Utility Manager for an investor-owned company to direct all functions of an operating division in northern New England. Offers broad scope of responsibilities and the opportunity for growth through performance to a person now in middle management. Requires a person with supervisory experience who is familiar with the technical, operating and commercial functions of an electric utility. Forward complete resume and salary history to Box 6109.

R&D—Vehicular Electric Propulsion. Electrical engineers for expanding long-range research and development of electric propulsion for vehicles. Requires experience in either advanced electrical machine design (induction, synchronous, and dc), or design of inverters, cycloconverters, and solid state control circuitry. Please submit resume in confidence to: Dr. Paul Agarwal, Head, Electric Propulsion Department, General Motors Research Laboratories, Warren, Mich. 48090.

Graduate electrical engineers to work for M.Sc., Ph.D. in power systems, electronic device and circuit studies, and in control systems. Generous support in the form of teaching assistantships and research assistantships. Please send resume to: Head, Dept. of Electrical Engineering, The University of Calgary, Calgary, Alberta, Canada.

Faculty teacher-researcher. To contribute to the development of a Ph.D. program in a university interested in both students and research. Ph.D. or equivalent publications. Send resume to Dr. W. H. Pierce, Head, E. E. Dept., University of Louisville, Louisville, Ky. 40208.

Power and Control Design Engineer. College graduate. Registration desirable. Project responsibility. Three or more years of experience in power plant or substation design on bus arrangement, wiring, drawing preparation, specification, or cost estimates. Bonus and other fringe benefits. Location—Denver. Firm pays employment fee, interview expenses and moving costs. Send resume in confidence. John L. Howell, Director of Personnel, R. W. Beck and Associates, 200 Tower Building, Seattle, Washington 98101. An Equal Opportunity Employer.

Electrical Engineer needed for responsible teaching position in expanding Two-year, Associate Degree Program in electrical engineering technology on new 12-million dollar campus. Competitive salary, overload pay, retirement, paid insurance, and tuition waiver are available. Minimum of Bachelor's Degree plus industrial experience required. Contact Walter Edling, P.E., Lorain County Community College, 1005 N. Abbe Road, Elyria, Ohio 44035.

Project Development Engineer. An electrical engineering graduate required to assume responsibility of development and promotion of new ideas and products to electrical utilities. Acceptance of product to the utilities' standards is the ultimate objective. Electrical utility company experience of some duration in transmission and distribution is necessary. Must be willing to travel and call on utility engineering departments to accomplish the above objective. Salary commensurate with experience. Location is in Western New York State with a medium sized company producing material for the electrical utility industry. An equal opportunity employer. Box 6110.

Full-time faculty positions in Mechanical and Electrical areas in rapidly expanding evening degree programs. M.S. degree required. Teaching and industrial experience desirable. Duties involve teaching, curriculum development and supervision of part-time faculty. Send resume to: R. A. Norton, Assistant Dean, Evening College, Rochester Institute of Technology, P.O. Box 3416, Rochester, New York 14614.

Pakistan Air Force desires to enroll suitable Pakistanis presently working or studying in U.S.A., as Professors and Associate Professors in Avionics. Interested Personnel are invited to contact Air Attache, Embassy of Pakistan, 2201 R Street N.W., Washington D.C. 20008 for terms and conditions.

An Engineering School of a prominent midwest state university requires a graduate electrical engineer or someone with equivalent design experience for a professional staff position. Responsibilities include electronic instrumentation design for research experiments and supervision of the School's electronic shop and technicians. Opportunities exist for pursuing educational objectives. Send resume to Box 6116. Salary commensurate with experience.

Research Professor. Outstanding opportunity for active researcher at large Eastern Engineering College. Teaching load restricted to three hours, competitive salary and excellent fringe benefits. Applicant should have publications and proven contract research ability. Teaching experience helpful, but not necessary. Box 6117.

Positions Wanted

Classified Advertising Rates for "Positions Wanted" column: $6.00 per line per insertion. 50% discount for members of IEEE (please supply membership number with copy). Copy must be received by the 2nd of the month preceding date of issue. Address replies to Box Number given, c/o IEEE Spectrum, Advertising Department, 72 West 45th St., New York, N.Y. 10036.

Getting full value from your computer system? Eight years experience providing and managing hardware and software technical support for fixed military command post information system users. Instructor, contract technical monitor, site manager, on-site rep. Analyzed requirements, determined needs and provided response. Available near Boston in October. MSPE BA Math 1947, BSEE Phi Kappa Phi 1958, MS Syst Engr (Reliability) Tau Beta Pi 1963. Box 8119.

Electronic Engineer. PE, MSEE, MA Math. 12 yrs. exp. seeks position in systems or equipment R&D in Met. N.J. Box 8126.

Prof. of E.E.: Currently holds British chair. Previous U.S. experience as full prof. E.E. and industrial vice pres. Major public'ns: control. Minor: power, management science. Visiting U.S./Canada July–Sept. exploratory talks interested engineering schools. Box 8127.

Technical correspondence

Climate control claims disputed

The August 1967 issue of IEEE SPECTRUM contained the article, "Economics of Indoor Climate Control," by H. L. Laube. On the basis of extensive research, I regret to report that the article is misleading, inaccurate, and deceptive; here is a sampling.

In his opening paragraph, the author states: "Thus we will limit our consideration to privately owned buildings, since it is necessary for the private sector of our economy to be cost-conscious." Yet, he devoted an important segment of his article to the all-electric Nassif Building in Washington, D.C., "which is occupied by the Defense Department." Can such a building be representative of "privately owned buildings"? The only answer to this question is no. He does not tell us much about the occupancy other than the short phrase, "occupied by the Defense Department." Writing for another publication in 1966,[1] the author discussed the Nassif Building somewhat more fully: "During the first year of operation, the building was fully occupied. Leased to the General Services Administration, Department of Defense, it has been 100 percent 'rented' since it was completed in May 1964. *It has been used virtually 24 hours a day with heavy overtime*" (emphasis added). Such occupancy would not be in the true spirit of his stated objective of discussing costs in "privately owned buildings."

It is also interesting to note that he takes this one office building having a gross floor area of 25 300 square meters (272 000 square feet), and compares the "total energy cost for this building, including cost of equipment maintenance" on a per square meter per year basis, with a "group of Washington office buildings." No identification is given of the buildings, whether they are large or small, new or old, or heated and cooled centrally or "incrementally." He blandly finds that one "Defense Department" building, operating 24 hours a day, has a lower operating cost than an unidentified group of other Washington office buildings. Let's tell the full story.

The group of Washington office buildings to which he referred is that included in a recent annual experience exchange report.[2] The study is statistical in nature and one which is meaningful because of its continuity over a great many years.

The use of the annual experience exchange report for the author's purpose was deceptive. The 1965 report included statistical information on 598 buildings located in 94 cities (Table I). The table shows that more than half of the buildings considered in the study were over 40 years of age; 75 percent were over 25 years of age. Of the 598 buildings, 18 were identified as being in Washington, D.C. This is the group to which the author referred, and only eight of these buildings were air conditioned. There is no basis for comparing these costs, associated with buildings the majority of which are in excess of 25 years of age, with the first-year costs of the Nassif Building, not three years old. Such a "comparison" simply does not do the job.

I. 1965 annual experience exchange report[2]

Buildings under 10 years of age	93
Buildings 10 to 25 years of age	53
Buildings 25 to 40 years of age	134
Buildings over 40 years of age	318
Total	598

Consider the statement on energy costs: "Actually, a careful analysis of the energy cost to heat a 20-story 8350-m² office building in a 5000-degree-day location indicates something of a standoff when using gas at 57¢ per gigajoule and electricity at $4.17 per gigajoule." Mr. Laube gives no further information in his article on what he means by the standoff, but in a paper[3] that he had published elsewhere he gives the figures for this hypothetical case (Table II).

II. 20-story, 8350-m² office building energy (heating) costs[3] (5000-degree-day location)

	Central Induction System, Gas-Fired	Incremental System, All-Electric
Gas, 41 000 therms at 6¢ (or 4330 gigajoules at 57¢)	$2460	
Electricity, 239 000 kWh at 1.5¢ (or 860 GJ at $4.17)	$3585	
Electricity, 605 470 kWh at 1.5¢ (or 2170 GJ at $4.17)		$9082
	$6045	$9082

Not only does he not disclose these important figures in the current article, but his statement that they "indicate something of a standoff" misrepresents the facts. One number is 50 percent greater than the other!

The article deals only in generalities; statistical averages of a vast number of unidentified buildings are the basis of comparison, and, even at that, important information is not given. The author states that the results of a study indicate that a 24-story building in Philadelphia and a 26-story building in Pittsburgh have shown favorable operating costs for cooling, again on the basis of comparison with the annual experience exchange report. The uniform system of accounts is used for the yardstick buildings. In the comparison case, he gives no hint of accounting methods other than that it contained "total cooling cost, including maintenance and filter cleaning."

The costs that are itemized in the annual experience exchange reports are based on uniform accounting methods. The uniform system of accounts includes wages, social security, compensation insurance, supplies, fuel, and unclassified items—each of these being separately itemized under three categories: electrical, heating, and air conditioning and ventilating. The "accounting" used by the author included only the annual cost of electricity, air conditioning maintenance, and electrical maintenance.

Comparisons of this type are usually based on the uniform system of accounts; a definite procedure is available, yet the article ignores accepted practice. The casual reader would not get the help he should expect from a responsible publication such as we consider IEEE SPECTRUM to be.

Many other points can be brought to the attention of the reader:

1. The Nassif Building contains 500 incremental units plus a 200-ton (air conditioning) package unit system for the core of the building and cafeteria. Maintenance of so many units is quite a chore and is one that will mount with years.

2. Eighty-five percent of the buildings were substantially smaller than the unidentified Philadelphia and Pittsburgh buildings. In fact, almost 40 percent were less than one half the height.

3. The author reduces the function of air conditioning to terms of thermal, acoustic, and olfactory environment; this is not in accord with the accepted definition,[4] "The process of treating air so as to control simultaneously its temperature, humidity, cleanliness, and distribution to meet the requirements of the conditioned space."

These and other points could be brought out in detail, but, for brevity, only a small sample is here presented.

Lawrence J. Hollander
American Gas Association
New York, N.Y.

REFERENCES

1. Laube, H. L., "The practicality of an all-electric building," *Skyscraper Management*, vol. 51, p. 16, Apr. 1966.
2. "Forty-Sixth Annual Experience Exchange Report, Office Building Operations Calendar Year 1965," National Association of Building Owners and Managers, Chicago, Ill.
3. Laube, H. L., "Selling kilowatts for comfort in multiroom buildings," a summation of papers presented by Mr. Laube before the Edison Electric Institute, Chicago, Ill., April 6, 1964, *et al.*
4. "ASHRAE Handbook of Fundamentals," American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., New York, N.Y.

Mr. Hollander is wrong in assuming that the Nassif Building is not privately owned. Its owners are Suffolk Properties, Inc. They lease the building to GSA and he is correct when he says that occupancy is not comparable to that experienced in a multiple-occupancy office building. Perhaps he missed the point: Despite the fact that the building is "used virtually 24 hours a day with heavy overtime," the actual operating costs for all energy and all maintenance and energy-using equipment amounted to $4.56 as against a comparable cost of $5.25 per square meter per year for the 14 air-conditioned buildings in the Washington area that reported their figures to the National Association of Building Owners and Managers for the year under consideration. These figures appear in the 1965 office building experience exchange report of that association.

In the writer's opinion, the operating

expense of the Nassif Building would have been considerably less had it operated on the same schedule as is typical of multiple-occupancy office buildings in Washington. In that case, the saving in operating cost of 13.2 percent of $14 911 per year would have been even greater for the all-electric installation.

In estimating energy usage and operating costs for multiroom buildings, long established practice is to consider only significant factors, such as exterior wall area and construction, exterior glass area, quantity of ventilating air, and interior heat loads from electricity, people, and other possible sources, and, of course, climate. The age of the building and its number of floors are, as such, not considered. Therefore, Mr. Hollander's discursive preoccupation with these two subjects is irrelevant.

Mr. Hollander has brought out a discrepancy between what the author reported in the article under discussion and what he reported in another publication. That other publication, which was dated, was prepared on the basis of information available in 1963. At that time, much fewer actual operating data were available on the cost of electric heat than have been accumulated in the subsequent four years. At that time, also, to forestall criticism, gas consumption was purposely estimated low and consumption of electricity high. As additional information became available there was no point in continuing to favor gas over electricity in this manner.

Table II of Mr. Hollander's presentation is incomplete. The complete table shows annual energy costs for cooling also, as well as the costs of other operating expenses. As a result, the all-electric system showed an annual saving in operating cost alone of $11 691, not to mention the saving in fixed charges resulting from a lower first cost.

Although, as Mr. Hollander indicates, a difference in accounting methods may result in an unfair comparison of operating costs, such is not the case in this instance. The audience before which the paper was presented was not one to be belabored with excessively detailed accounting breakdowns. The important figure, in the eyes of building owners and managers, is the cost for a service such as air conditioning per square meter per year. These are the figures that the author used.

Mr. Hollander expresses concern over what he anticipates will prove to be excessive maintenance costs of the equipment in the Nassif Building. The

owners do not share Mr. Hollander's concern, inasmuch as they placed a ten-year maintenance contract at a fixed price per year (no escalation clause) with a local air-conditioning maintenance firm of long experience, at a cost of $21.43 per ton of refrigeration per year. Although this price is high, as against a price at which other maintenance contracts (including escalation clauses to offset inflation) have been placed by other buildings with similar systems, it compares favorably with the $28 per ton of refrigeration recently quoted in New York City for maintaining a heat-actuated absorption refrigeration system. And the latter figure included only the refrigeration system, not the terminal element nor the boiler.

In conclusion, it should be observed that the air-conditioning industry is increasingly polarizing around two energy sources—gas and electricity. Increasingly, therefore, the air-conditioning industry needs an unbiased referee of integrity for digging out and publicizing the credible information that is so sorely needed today. The logical body for this job appears to be the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). Efforts are now being made to move in that direction and such efforts are worthy of support from all concerned.

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

The systems approach

Kudos to the COSINE committee and its report in the March issue of IEEE SPECTRUM about "Computer Science in Electrical Engineering." With regard to the third recommendation for new course development, i.e., "the offering of a course in finite-state systems at the junior or senior level," the importance of such a course transcends the essential finite-state nature of computational devices. As the committee noted, "finite-state systems constitute a very basic class of systems particularly well-suited for the introduction of such basic concepts as state, equivalence, identification, decomposition, etc." As a result, the utility of a course in finite-state systems is great enough to merit the course being required rather than elective, as the committee suggested.

In recognition of this belief, the Systems Engineering Department at the University of Arizona has required a course in finite-state systems of all its undergraduates for the past 4½ years. It has been taught from Arthur Gill's excellent book with great success. Currently, the core program for systems-engineering undergraduates requires the course in the *sophomore* year.

R. B. Hunt
Albuquerque, N.Mex.

Basic transportation!

In Mr. Hirsch's letter in March's issue, he comments on how an individual could get to the bus. One method he notes is to walk, possibly up to one kilometer. If this one kilometer is converted to the standard units used in the United States, it is slightly over 0.6 mile.

I feel sure that if all commuters lived within 0.6 mile of reasonable bus service, there would be much less congestion on the highways today. I submit that this is a rather nominal walk. Many of us walk nearly that far in the morning trying to find the alarm clock.

James C. Arnold, Jr.
Arlington, Va.

Cybernetics concern

IEEE SPECTRUM articles are, in general quite good and capture succinctly the important aspects of the field. However, cybernetics, which is well known for its overstated claims, vague generalizations, and meaningless models, cannot be surveyed in an uncritical manner.

The article, "Human Enhancement: Beyond the Machine Age," (Feb. 1968), although offered to "stimulate a dialogue among all kinds of systems designers," is too bland to stimulate much else than letters such as this. I would like to see a presentation that lies somewhere between the innocuous one in SPECTRUM and the slashing attack of "Computers and Common Sense" by M. Taub (New York: McGraw-Hill, 1961).

Perhaps something on the style of "Great Ideas in Information Theory, Language, and Cybernetics" by J. Singh (New York: Dover, 1964) would be appropriate. In particular, I was thinking of his eight-page epilogue that discusses "cybernetics" in a critical, but understanding manner.

I realize the problem in trying to obtain an author who is knowledgeable in the field without being part of the "establishment" (having a vested interest in presenting a positive survey), and hope that future papers in the cybernetics area can be more meaningful.

Oscar Firschein
Lockheed Missiles & Space Company
Palo Alto, Calif.