VOLUME 8 □ NUMBER 1 IEEE spectrum JANUARY 1971

features

the cover

Scheduled for demolition, this abandoned apartment building is symptomatic of a "depressed" urban system mode, where excess housing and a large low-income population interact to drive out business and reduce the tax base. As the article beginning on page 48 points out, people crowd into some dwelling units while other buildings stand idle and decay.

departments

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

IEEE SPECTRUM is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: 345 East 47 Street, New York, N.Y. 10017. Cable address: ITRIPLEE. Telephone: 212—752-6800. Published at 20th and Northampton Sts., Easton, Pa. 18042. Change of address must be received by the 1st of a month to be effective for the following month's issue. Please send the SPECTRUM mailing label showing your old address, together with your new address to Coding Department, IEEE, 345 E. 47 St., New York, N.Y. 10017. Annual subscription: IEEE members, first subscription $3.00 included in dues. Single copies $1.50. Nonmember subscriptions and additional member subscriptions available in either microfiche or printed form. Prices obtainable on request. Editorial correspondence should be addressed to IEEE SPECTRUM at IEEE Headquarters. Advertising correspondence should be addressed to IEEE Advertising Department, at IEEE Headquarters. Telephone: 212—752-6800.

OTHER IEEE PUBLICATIONS: IEEE also publishes the PROCEEDINGS OF THE IEEE and more than 30 Transactions for IEEE Groups with specialized interests within the electrical and electronics field. Manuscripts for any IEEE publication should be sent to the editor of that publication whose name and address are shown on page 93 of this issue. When in doubt, send the manuscript to E. K. Gannett, Editorial Services, at IEEE Headquarters, for forwarding to the correct party.

IEEE also copublishes, with the Institution of Electrical Engineers of London, England, *Electrical & Electronics Abstracts* and *Computer & Control Abstracts*, as well as their companion titles journals, *Current Papers in Electrical & Electronics Engineering* and *Current Papers on Computers and Control.*

Forum

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

IEEE and the unemployment problem

The article on the employment problem in the November issue of *IEEE Spectrum* revealed what I perceive as inadequacies in the leadership of the IEEE.

I was particularly struck by the statement that an IEEE role of lobbyist or pressure group "would be morally repugnant to most, if not all, of our membership." The sad truth is that the IEEE leadership has never investigated members' attitudes on this and related issues. The officers of the IEEE seem content to act as a clearinghouse for scientific information. There has never been a referendum among members on important policy issues nor is there an opportunity for members to express themselves effectively when they elect officers. Candidates for office are identified solely by professional credentials and not by the programs or visions they might have for the Institute. The officers constitute a society within a society, which apparently cares little about the full spectrum of members' needs and desires.

The IEEE has two major membership groups—the academic and the industrial—with differing and sometimes conflicting needs. Members in the academic group usually belong to other organizations that can represent them with their employers and society at large, such as the American Association of University Professors and faculty senates. Industrial engineers have no such organizations to represent them. It is ridiculous to say that engineers would find organized representation with their employers, the U.S. Congress, or society at large "repugnant" since most other groups within our society gladly accept this representation when it is available.

The problem that faces the industrial engineer is finding an effective technique for organized representation. Unionization is not likely to be successful because the engineering activity is too far removed from the production line, and unionization is effective only for production workers such as teachers, policemen, and construction and assembly line workers.

On the other hand, organized representation to Congress is practical since Congress influences the supply of engineers (through grants to educational institutions) and the demand for engineers. You suggest that such representation is immoral, although our entire political system is predicated on informed groups of citizens organizing and representing their views to their lawmakers. Furthermore, the representation should deal not only with how Congress influences the supply and demand of engineers but also with how it influences the way engineering is done in the U.S. After all, Congress is largely responsible for encouraging the growth of research-oriented universities and the large aerospace complex, as well as a large number of technical government agencies. Informed engineers should influence how these decisions are made.

It appears that the officers of the IEEE not only lack the vision to see how they constructively can influence government decisions but also are unwilling to consult with their membership on what kinds of activities are desirable and appropriate. It was the *New York Times* and not *IEEE Spectrum* that observed, "The issue transcends the personal hardship of highly educated professionals seeking to support their families by working as handymen. . . The deeply troubling, fundamental question is whether a nation can allow itself to be the pawn, rather than the master, of its destiny. Young Americans today are told that Ph.D.'s are a drug on the market. Only a few years ago, to become an engineer or a scientist or indeed a teacher was not only to fulfill one's goal of intellectual satisfaction, but to serve nation and society . . . it is intolerable that brainpower be put in mothballs like unwanted warships. If under such conditions the brain-drain is allowed to proceed, this will clearly tell new generations that America no longer wants the best of their talents, inventiveness and reason. To be incapable of employing the country's intellectual resources for peace and progress as readily as for war and national glory is to clamp an embargo on hope and faith."

I believe that the IEEE should remove its blinders and work with its membership to find new ways to promote the well-being of engineers while promoting the best interests of America.

J. A. Rosenthal
Rochester, N.Y.

I read with amazement (but not amusement) Mr. Granger's article on the employment problem. I am amazed that the IEEE Board of Directors feels it cannot and should not do anything, and further, that it will not do anything but offer token retraining (or is it reprogramming?). Are we so bound by our sacred tax status that we dare not take positive action? Are we so enthralled with the purism of science that positive action in the economic area is beneath our professional dignity? Are we so wrapped up in our own little worlds, so divorced from events around us, that we will not take positive steps to alleviate the employment problem?

To me, the response to the employment problem expressed in that article is unacceptable. We can and must do more. True, we cannot solve the problem completely, but we can help to alleviate it. We can help to remove some of the unfairness that accompanies it. We can work to prevent its reoccurrence. The IEEE is in a position to do things that individuals cannot do. It can take actions on its own and in cooperation with other technical societies and governmental agencies. The time to do something is now.

James W. Berkovec
The Aerospace Corporation
San Bernardino, Calif.

The two articles in November's *IEEE Spectrum*, coming from the President and the General Manager of the Institute, show the gravity of concern felt by engineers in the current employment situation, and provide much needed information for an intelligent discussion.

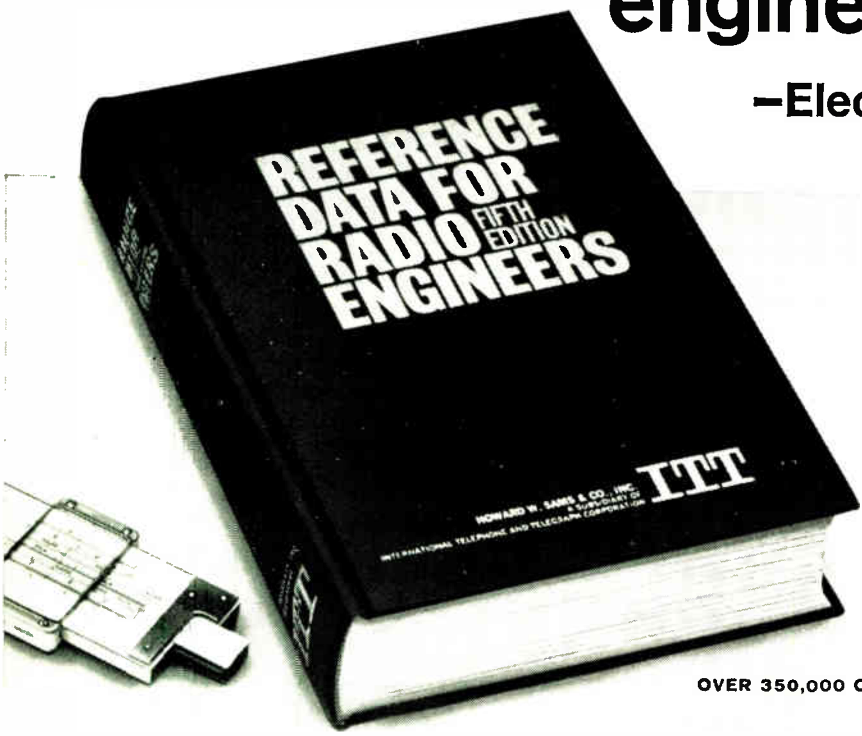
The President points out that the IEEE cannot, legally or ethically, lobby for legislative action to improve employment stability. He states that the Board of Directors has decided that it is not advisable to change the Institute's tax status from "educational and scientific organization" to "business league" (like AMA, ABA, NSPE). Although the decision is arguable (and may be modified at a later date if circumstances justify it), let us accept it and examine what can be done by the IEEE in nonlegislative fields.

A very important role for IEEE is to help develop a *code for employers*. Much of this can be done through persuasion and coordination, because many engineers are involved in the management's decision-making process and, as professional people, are interested in the development and welfare of fellow professionals. Since the code would be voluntarily accepted by the industry, it would require no legislation.

Voluntary cooperation in areas of mutual interest, even by competitors, is not new. Examples of uniform codes for hiring and firing are not many, but the universities offer an example where the

"Bigger and better...this engineer's bible..."

—Electronics Magazine



5th EDITION of a world-famous reference

Now contains 50% MORE DATA to include major information developed over the past 12 years. Prepared by an outstanding group of practicing engineers, professors, and industry and government experts, under the direction of the International Telephone and Telegraph Corporation staff.

OVER 350,000 COPIES SOLD IN PRIOR EDITIONS

Reference Data for Radio Engineers

CHAPTER CONTENTS

- Frequency Data
- *International Telecommunication Recommendations
- Units, Constants, and Conversion Factors
- Properties of Materials
- Components or Parts
- Fundamentals of Networks
- Filters, Image-Parameter Design
- Filters, Modern-Network-Theory Design
- Filters, Simple Bandpass Design
- Attenuators
- Bridges and Impedance Measurements
- Magnetic-Core Transformers and Reactors
- Rectifiers and Filters
- Magnetic Amplifiers
- Feedback Control Systems
- Electron Tubes
- Electron-Tube Circuits
- Semiconductors and Transistors
- Transistor Circuits
- *Microminiature Electronics
- Modulation
- Transmission Lines
- Waveguides and Resonators
- Scattering Matrices
- Antennas
- Electromagnetic-Wave Propagation
- Radio Noise and Interference
- Broadcasting and Recording
- Radar Fundamentals
- Wire Transmission
- *Switching Networks and Traffic Concepts
- Digital Computers
- *Navigation Aids
- *Space Communications
- Electroacoustics
- Nuclear Physics
- *Quantum Electronics
- Information Theory
- Probability and Statistics
- *Reliability and Life Testing
- Miscellaneous Data
- Fourier Waveform Analysis
- Maxwell's Equations
- Mathematical Equations
- Mathematical Tables

*New chapters on subjects not covered in 4th Edition

1196 PAGES • 1350 ILLUSTRATIONS
plus 41 page index

The FIFTH EDITION of this widely used reference work has been completely revised and updated. It provides in a single volume comprehensive data on all basic phases of electronics, including tables, formulas, standards, and circuit information—*PLUS*—all-new data on microminiature electronics, space communications, navigation aids, reliability and life testing, international telecommunication recommendations, switching networks and traffic concepts, and quantum electronics. Includes over 1300 charts, nomographs, diagrams, tables, and illustrations. 45 data-packed chapters. 1196 pages of invaluable information for practicing radio and electronics engineers.

10-Day Free Examination

HOWARD W. SAMS & CO., INC., Dept. IE-011
A Subsidiary of International Telephone & Telegraph Corporation
4300 West 62nd Street, Indianapolis, Ind. 46268

☐ Send me REFERENCE DATA FOR RADIO ENGINEERS for 10-day examination, without cost or obligation. If I decide to keep the volume, I will send $20.00 (plus postage) in 10 days; otherwise, I will return the book within 10 days.

☐ $20 payment enclosed. Send my copy postpaid with full 10-day refund privilege. (Include sales tax where applicable.)

Name ______________________

Company ______________________

Address ______________________

City ____________ State ____________ Zip ____________

Circle No. 4 on Reader Service Card

American Association of University Professors (AAUP) has helped develop a code—partly unwritten—governing employment policies.

The employers' code should include the following points:

1. Stabilized employment. This could be accomplished through (a) more realistic prediction of manpower and skill requirements; (b) sharing a manpower pool by defense and aerospace industries where the gain or loss of a contract results in large manpower fluctuations; and (c) classification of professional jobs by each employer into categories with increasing stability and careful filling of these categories with people whose performance is really valuable to the company. These steps would ensure that the organization hires selectively, does not overhire, and identifies its core jobs as well as core personnel.

2. Improved pension plans. These would aim at (a) elimination of waiting periods for eligibility; (b) full and immediate vesting of benefits; and (c) full portability from one employer to another.

3. Improved unemployment benefits. A professional man whose services are terminated (except for incompetence or similar cause) should get a specified fraction (say 25 to 50 percent) of his salary, according to a sliding scale and with prescribed minimum and maximum limits, for a period proportional to the length of his service with the company. Such benefits are already given by several firms in the form of termination salaries, but the practice is neither universal nor uniform. To make it attractive—and fair—to the employer, the benefits could be spread out over a number of months and could be discontinued or diminished if the man gets another job.

4. Retraining grants and loans. The employers, and perhaps professional societies, should make grants and loans available to scientists and engineers who wish to retrain themselves in new fields. The benefits could be similar to those offered by the G.I. Bill or by the existing programs for college loans.

5. Unemployment insurance. This is an area where the IEEE and other professional societies could take the initiative even without the participation of employers. Several societies have already arranged for sickness and accident insurances for their members at group rates. They could now arrange for unemployment insurance, whereby a member could pay an annual premium and could be eligible for receiving a certain amount—say $200 per month—for a specified period in case of unemployment. This income would tide him over the period needed for retraining or searching for a new job.

It is hoped that the IEEE Headquarters and members will include the foregoing proposals in their consideration when discussing and planning the Institute's role in professional matters.

S. H. Durrani
Associate Editor
AES Transactions
c/o COMSAT Labs
Clarksburg, Md.

The IEEE, professional or technical society?

The following comments are a rebuttal to Mr. Granger's opinion (*IEEE Spectrum*, November 1970) of what the IEEE can and cannot do regarding engineering employment and regarding the IEEE's status as a scientific (rather than professional or "business league") society.

I would like specifically to take exception to the premises, based solely, it seems, on the "opinion of the Board of Directors," that:

(a) Jeopardizing the IEEE tax-exempt status in favor of adopting the role of a "pressure group . . . would be morally repugnant to most, if not all, of our membership."

(b) Using "technical expertise or institutional solidarity" to influence national goals in a political manner would impoverish. . . .the professional recognition the public extends to engineers.

(c) "IEEE cannot legally, or ethically, lobby in Congress" for conditions favorable to stabilizing engineering employment.

(d) The IEEE should confine its activities to retraining and should not adopt the role of a "business league" or an "active lobbyist" such as the American Medical Association (AMA), the American Bar Association (ABA), or the National Society of Professional Engineers (NSPE).

Tax-exempt status

First, there is absolutely nothing sacred about our tax-exempt status. Loss of it would require an increase in dues and/or advertising rates. However, if forming a more effective group directly or indirectly increased a member's income by 0.2 percent, it would permit doubling the present dues with no loss to the individual. There is no other reason than cost to resist a change in status.

Parallel to ABA, AMA

Our associates in the medical (AMA) and bar (ABA) associations saw the need for a single society long ago and are now quite willing and quite well able to pay for it. A doctor may pay between $1000 and $2000 a year to his association, but has an income level of more than $50 000. Should we quibble about increasing our dues from $25 to $50 a year?

To answer one traditional disclaimer—it is no longer true that doctors and lawyers are *all* self-employed and, thus, more directly control their own incomes. A significant number *are* employed by government agencies, universities, and privately or publicly owned firms. These professional employees enjoy far better incomes and working conditions than do engineers *solely* because of their professional associations' activities. These associations do have public welfare as their publicized *raison d'être* but they recognize that the public is best served by satisfied, well-paid, stable professionals. Both organizations originally were formed to take care of their members' self-interest. Altruistic concern for public welfare and technical advancement followed.

What should a professional engineering society do?

Engineers should be represented by a single organization that is concerned with: (a) the professional, legal, and ethical standards; (b) the technical development of the individual; and (c) the engineer's position in society. Consolidation of all three functions under a single group permits one function to augment the other in an effective and synergistic manner. Having a single representative for all professional matters heightens member interest and participation, as well as willingness to pay. The inevitable result is a more dynamic group, as witnessed by the ABA and AMA.

Professional standards

Educational, legal, and ethical standards can be maintained by: college accreditation processes, which keep the academic curriculums and bachelor requirements of our science and engineering schools at a high level; legal licensing of those engineers working in fields involving public welfare and safety; establishment of uniform professional requirements for entrance into the professional/technical society(ies); establishment of voluntary or mandatory programs of reexamination, recertification, or reeducation during an individual's career; and formation of ethical and educational review boards.

The present influence of IEEE and other Engineering Joint Council (EJC) societies is too weak to enforce any professional standards. The various professional engineers groups do strongly influence state licensing requirements but unfortunately represent only about 60 000 of the more than one million engineers and scientists in the U.S.

Technical development

The various technical societies are doing a credible job in encouraging the technical development of the individual through seminars, meetings, and journals. This activity could be greatly enhanced by unification under one parent group with some consolidation of paid staffs and facilities.

Political social involvement

There is effectively *no* group representing the engineer in the political and social forum. If we are to achieve financial gains and improve security and/or portable benefits in our employment, we must have effective lobbying and publicity representatives who will win support not only from Congressional and administrative leaders but also from the general public. We must have a professional society that not only ensures that the engineer maintains and increases his value to society but also makes sure that society is constantly aware of and desirous of his contribution.

In short, the engineer must be taken from behind the laboratory curtain of anonymity and be given a public and political image. Politicians, taxpayers, and consumers pay our salaries. If we want a bigger share of the gross national product (than, say, a plumber, auto worker, or airline pilot), we have to organize, assess desires, set realistic long- and short-term goals, and *lobby, lobby, lobby*.

Portable benefits

In addition, the society must provide an acceptable total package of fringe benefits that travels with the member throughout his career. Employer-controlled pension plans no longer serve a useful purpose in minimizing job-hopping. The economy and the employer's success in securing new business is conducting the "hopping tune" now, not employee desires.

Although the pension involves the largest financial factor in a required portable package, medical, life, and disability insurance are also significant. The unemployed or reemployed engineer may lose these protections (or pay dearly for them) until and/or unless he endures a waiting period and qualifies medically for new plans.

The engineer is entitled to these benefits or the cash equivalent. He has at one time or another received them in lieu of a cash raise. If he adopts a private plan of pension and medical coverage in lieu of his employer's plan, he is leaving money on the table. As an individual, he will not get group annuity and insurance rates, and he will not get the employer's premium contribution, which frequently equals or exceeds his share.

When he terminates with a company, voluntarily or involuntarily, he will lose all benefits except those legally vested with him. He'll also leave behind other benefits such as extra vacation time and other seniority-related benefits. Perhaps company policies should tie the latter to salary or level of responsibility instead of, or in addition to, straight seniority.

Private bargaining—salary statistics

The engineer's competitive position in privately bargaining for his own salary demands (and fringe benefits not supplied through a professional society) would be enhanced greatly by comparative statistics. If the professional society collected and publicized data on salaries, vacation and holiday schedules, fringe packages, relocation policies, *et al.*, the individual could exert more rational and realistic pressures to improve his lot. The publicity, and ensuing mobility of the engineering population, would result in a leapfrogging effect on salaries, and benefits would be competitively improved.

Who pays for all this?

The final total compensation level reached would be in response to the inevitable law of supply and demand. It is obvious that there will be extra costs. Some of this can, and will, be absorbed by the public digging deeper to maintain a resource that is truly useful to society, *if* its value is constantly and emphatically asserted.

This presupposes that engineers are assigned to stable programs that are worthy of national goals—no mean task today as we shift our collective interests from military–aerospace developments to problems of education, health, housing, transportation, and ecology.

The public might not have to dig down so deep into its collective pocket, if a more critical look were taken at the utilization of engineers to eliminate wasteful stockpiling and subprofessional assignments for skilled and highly paid people. But the basic funding source for these benefits will ultimately be John Q. Public and his friend, Uncle Sam. We have to make sure they know our story and are willing to pay.

In the past decade, engineers and scientists have been more responsible than any other group for gains in productivity and consequent growth of the Gross National Product (GNP). And yet engineers' salary gains have lagged far behind most production and construction workers and behind professional groups who are more vocal and more persistent. Gains for these groups range from 6 to 10 percent annually and exceed the cost of living. Engineers' salary increases of late do not equal the rising cost of living.

No one company can start the movement upward in engineering compensation. To do so in today's tight competitive market would be to invite financial suicide. If our profession is to secure a larger share of the GNP, it must do so gradually and in a manner that tends to equalize the burden among all industries. If all engineering professionals, both *employer* and *employee*, have the same personal goal and means to achieve it, much of the trauma associated with increased salary demands can be quietly and rationally settled.

What about labor unions?

The professional society as proposed would not be a recognized or certified collective bargaining group. It would represent all engineers. When an individual becomes an employer, either as an owner of a private firm or as a member of management in industry, the more senior professional engineer still has the same financial, social, and technical needs as his juniors or nonsupervisory peers. To preclude managers or supervisors from a professional society would be to lose a valuable resource. The same enthusiasm, energy, and intellect that warranted their promotion could, and probably would, be exercised to benefit all their fellow professionals.

This does not mean an end to collective bargaining for some engineers. Where and when the need exists, the independent union or trade association can coexist with the professional society with possible mutual benefits.

Political involvement—what value to society?

So far I have discussed the value of a professional society in enhancing the public image of the engineer solely in terms of self-interest. For the public to want to pay for this new image, it will demand a more effective argument than the typical Madison Avenue "snow job" and more impressive image than an introverted laboratory mole.

In our increasingly technical society, engineers must recognize the social impact of their innovations, define and quantify this impact, and be outspoken in discussing this with government agencies, the press, and interested organizations. We must actively participate in making the vital decisions that affect our nation politically, economically, and environmentally, both because we are morally responsible for our inventions and because we know how to measure, control, and modify these inventions better than anyone else.

In our technology-laden society, the engineer's technical knowledge probably should represent 20 to 30 percent of the factors involved in decision-making—in fact, it is probably considered as worth only 2 to 3 percent. And so we have automobiles that menace life, lungs, and limb but improve(?) in style every year; power stations designed and located based on narrowly derived options and facts rather than on a full concern for all environmental factors; chemical food additives and crop agents of unknown effect on present and future generations; road routes dictated mainly by politics and doomed to obsolescence and overcrowding before the last concrete is poured; "my lady's" appliance that gets its first real quality check in her kitchen and laundry; hazardous, breakable toys. . .excess garbage. . .cable television. . .on and on in an endless list of problems that clearly demand technical solution, but which will probably be resolved by executive orders, legislative action, or regulatory agencies with little informed advice.

Can we think only with a slide rule?

And what about the problems that cry for solution in areas not clearly technical? It is a fact that science and engineering majors enter universities with a higher I.Q. rating than followers of most other disciplines. Do we suddenly lose our reasoning powers in the laboratories? Have we received a college education or a vocational training? Do we suppress all interest in things social because they cannot be easily measured, molded, and predicted?

We should be in the vanguard of reform groups attacking health, social, educational, and welfare problems along with our fellow professionals—doctors, lawyers, psychologists, and teachers. In most instances, we do not even make the rear-action squad.

Every engineer does not develop the talent for full participation in the political world. Neither does every doctor or lawyer. But every engineer can support a society that helps him participate by representing the members' varied views and alternative solutions to national problems. There are enough articulate spokesmen among our colleagues to take care of publicity. What we need is a rallying point to channel our massive, collective brain trust toward the solution of pressing problems.

What needs to be done?

The IEEE must assess member needs, goals, and willingness to pay. Forum meetings at the Sectional level should provide the groundwork for a well-structured nationwide poll, perhaps with the other members of EJC and the scientific societies. On such critical issues, the IEEE Directors need not and should not guess, opine, or unilaterally dictate what its members want (and are going to get).

IEEE next should determine the possible structures of a single group that can represent the engineer in the professional, technical, and political arenas. Realism dictates maximum conservation of the existing technical and professional societies.

Implementation will require a scheduled entry into those activities over a long period and probably will require the continuity that can only come from a full-time paid staff.

Why IEEE? Why now?

The IEEE, as the largest engineering society in the U.S., is the logical choice to start such a society. It has over 150 000 members comprising about one sixth of the scientists and engineers in the United States. It has the greatest resources to provide preliminary funding, conduct surveys, and make an effective proposal. Actual implementation, in transition at least, could be assigned to EJC or a similar umbrella group.

The IEEE could certainly maintain its distinct (and international) identity, representing a unique, scientific discipline, as one technical division of the professional society. This would be consistent with the structure of the AMA and ABA. The present IEEE administration need not directly involve itself in lobbying. This would be the responsibility of the executives of the professional society; they in turn might elect to fund a separate lobbying corporation as the medical society does.

If the IEEE does not take the lead now, it may be taken from us forever. The employment situation is bad now and will become worse before it gets better. Situations similar to this have bred the existing professional and trade union associations. It is not too late to change our structure. I would rather see IEEE as a technical division of the American Engineers Association than see it dwindle to an impotent, obsolescent group.

Are these thoughts shared by others?

I should point out this entire subject was thoroughly aired in Washington, D.C., before a distinguished panel including the executive secretary of the President's science advisor, the executive secretary of a national scientists' and engineers' bargaining unit association, the director of the Washington, D.C., American Bar Association office, and a manager from a major electronics firm. The panel was convened and chaired by the President-elect of the IEEE to provide a sounding board for member views.

The reaction of the 160 members at the meeting left little doubt that my views were shared by most of those present. Moreover, the majority of the panel also agreed that the way to improve the engineer's position was through a stronger organization that could function effectively in the political, social, and economic fields.

In addition, there are a multitude of surveys published each month in the trade press that confirm the engineers' demands for professional organization and participation in government (*EDN*, October 1970; *Industrial Research*, October 1970; The *Professional Engineer*, November 1970, to name just a few current surveys). The ASME is using the technique of in-depth colloquiums, lasting several days, to probe these same questions, and is coming up with similar responses.

Conclusion

The engineers in the United States need, and are demanding, an organization that represents them on all fronts. The present state of the economy is adding impetus to this, but even an upswing of the economy will not eliminate the acute hardships and the waste associated with our profession today. The nation's needs demand more technical solutions and more professional involvement in decision-making processes.

If the IEEE does not accept the challenge of this responsibility, its members will certainly turn to the organization that does.

To paraphrase the comments at the recent panel discussion: "We've heard all the objections to change, all the reasons it will not work, but it's got to work—so let's talk about how we're going to make the necessary changes."

R. J. Backe
Silver Spring, Md.

I have received numerous letters from members who are dismayed (or outraged) by what they view as an inadequate (or worse) response by the IEEE Board of Directors to the problems currently confronting so many of our membership. The flow of correspondence on this subject has reached such proportions that I am simply unable to reply individually to each writer. I am therefore resorting, reluctantly, to the device of a form letter. In it, I am attempting to comment meaningfully on those several topics that are raised most often in this correspondence.

Many of my correspondents take issue with my comment, on page 33 of the November issue of *IEEE Spectrum*, that the Board feels that for IEEE to act as a "pressure group" would be "morally repugnant to most, if not all, of our membership." These writers usually imply that the notion that a pressure group is morally repugnant is peculiar to the members of the Board. In fact, of course, the image of a "pressure group" as a group pursuing self-serving goals through political pressure is part of the "common wisdom." Quite aside from the perhaps debatable question of "moral repugnancy," however, the Board must consider, and has considered, the question of whether IEEE, acting as a "pressure group," could effectively alleviate the problems that concern us all. Many members apparently have concluded that we could, and urge that the IEEE be reorganized along the lines of the American Medical Association.

Those who regard medical doctors—and the American Medical Association—as an appropriate model for electronics engineers—and the IEEE—in relationship to society should consider the very great difference between the practice of medicine and the practice of engineering. Medical practice is a personal service in which the patient (employer) entrusts decisions to the skill and judgment of the doctor, decisions that are ultimately and quite literally matters of life or death. By contrast, most electronics engineers work as part of a group enterprise. Their work deals not with individual human life, but with things—devices, mechanisms, systems of inanimate objects. In critical situations, a doctor must make vital decisions on his own, without benefit of time for study and research, without consultation, and without review. When an engineer must make a critical decision, he is expected to take time to do the relevant research, he is encouraged to consult with other specialists, and his critical decisions are invariably subject to review before a final commitment is made.

These differences in roles and methods, and in the ultimate importance of the outcome to the employer (patient), are reflected in the manner in which the employer chooses the individual to whom he entrusts the task, in the form of the relationship between employer and employee, and in the way in which the employee's compensation is determined. The engineering employer chooses a new engineer on the basis of his academic and work record, frequently augmented by specific tests of his skills, and after detailed questioning through which the employer undertakes to determine the prospective employee's attitudes, approach to his work, expectations for salary, and other rewards. A patient chooses his doctor (if circumstances permit him a deliberate choice) on the basis of his reputation and, more important, on his own (the patient's) feeling of confidence in the individual doctor. Having made his choice, the patient accepts a tacit responsibility to follow the doctor's instructions; that is, he yields to his *employee* the authority to direct future events. Once hired, the engineer–employee expects his employer to provide (more or less) explicit direction of his work and to accept the responsibility for assigning him to tasks appropriate to his skills. The engineering employee is never expected to assume personal responsibility for the business success of the enterprise that employs him.

Having chosen a doctor, and having yielded to him a high degree of responsibility for the patient's physical well-being (perhaps even for his very life), the patient is not ordinarily disposed to haggle (negotiate) about the doctor's fees (wages). If his medical care is covered by insurance or Medicare, the patient knows that the amount paid the doctor will, in most instances, have been predetermined by the insurance carrier (or by a government agency). The engineering employer, in comparison, pays his engineering employee a predetermined salary, which reflects the current availability of qualified applicants. The "labor market standard" resorted to serves the needs of both employer and employee. If the proposed salary is too low, by this standard, the employee will refuse it and take a higher paying job. If it is too high, by the same standard, the employer will be incurring higher than necessary costs, and his business will (ultimately) be adversely affected.

It has been argued that it is the existence of the AMA and its aggressive public relations and lobbying programs that permits doctors to enjoy their high professional standing and their correspondingly high income. This argument ignores both logic and history.

As I have described in the foregoing, the very nature of the patient–doctor relationship acts to endow doctors with a high degree of professional standing, and the vital importance of medical care to the patient's survival tends to remove medical fees from the workings of a competitive market. These factors existed before AMA ever came on the national scene, and they would continue to exist if AMA were to disappear. AMA is well known to the public for its long—and unsuccessful—effort to prevent the adoption of Medicare. The AMA position on Medicare reflected a variety of motivations, not the least of which was a desire to head off government intervention in fixing medical fees. Doctors recognized that government-established fees would likely be lower than those they were accustomed to, and that "free" medical care would inevitably increase their work load, particularly by encouraging office visits by many persons who had no specific medical problems, but simply desired reassurance. AMA worked very hard, and spent a great deal of money, in its efforts to head off Medicare. In spite of this, and in the face of the extraordinarily high professional status enjoyed by AMA's roughly 200 000 member doctors, their efforts failed. The adoption of Medicare was a political decision, of course, and, in the end, the Congress recognized the greater political strength of that sector of the public that wanted such a program adopted. (A subsequent development provides a footnote to the history of AMA's efforts, which may be enlightening to IEEE members. After Medicare was adopted, the AMA for a time urged its member doctors to refuse to participate. One group, the District of Columbia Medical Association, did refuse. As a result that group—and the AMA—were brought to trial by the Department of Justice for violation of antitrust laws. The AMA was convicted and paid a very heavy fine.)

IEEE members who seek ways to increase public acceptance of their professional status as engineers, and by so doing hope to achieve more generous (and more dependable) compensation, must look elsewhere than AMA for an appropriate model. I believe IEEE can help its members to achieve a better public understanding of the social value of their work, and IEEE must do so. Success in dealing with the other problems we face—unemployment, forced migration in job seeking, and inadequate compensation for our highly trained services—will come only when the public recognizes an increased need for the products and services we supply. We must recognize that, unlike medical care, most of what electronics engineers do is, in the final analysis, the result of a deliberate choice by the public or its elected representative. Consumer electronics clearly subsists on the personal interest (and pocketbooks) of individual buyers. The Apollo program, which resulted in $20 billion of aerospace and electronics business in less than a decade, filled no need more tangible than national pride and scientific curiosity. Industrial electronics builds on the competitive needs of manufacturers and businessmen, not on politics. Defense electronics waxes and wanes with international tensions, and with the subjective responses these tensions evoke at the polling place. The ability of electronics engineers to contribute to the solution of pollution problems, mass urban transit, our decaying urban areas, and opportunities for all our people is not self-evident. I believe IEEE must provide aggressive leadership in delineating the technical aspects of these problems, and in providing the means by which practicing engineers can develop the specialized skills they need to contribute to them.

In my opinion, the IEEE Board of Directors, and the many volunteer members who serve on its operating committees, are facing squarely and realistically the many basic problems that afflict our membership. They are pressing ahead, with mounting momentum, on programs that will help to deal with these problems in constructive and effective ways. That they have not succumbed to the pressure of certain members who demand that IEEE become a "pressure group" is, in my opinion, a measure of their realistic appreciation of the total problem and their determination to respond effectively. Those who picture the IEEE Board as isolated atop ivory (or porcelain) towers cannot know them as I do.

J. V. N. Granger
President IEEE 1970

An open profession?

I wish to comment on Dr. Willenbrock's article in "Spectral lines" in the August 1970 issue.

Dr. Willenbrock contends that the engineering profession is not "open" to certain members of our society, but the logic with which he supports his contention is faulty. He commits a very basic error by asserting that the absence of certain groupings of people in our profession is prima facie evidence that the profession is not open to them. This is, of course, not true. Since it is quite obvious that anyone with the appropriate knowledge can become a member of the engineering profession, I do not really see how it can be said that the profession is not "open."

It is expected that a profession will treat all who seek membership with equality. Therefore, I do not think that the IEEE should set up any programs for the purpose of giving special advantages to any segment of our society.

Jesse B. Ring
McGraw-Edison
Canonsburg, Pa.

After reading the August "Spectral lines" I felt it important to report that some of the Student Branch members here at Berkeley started in 1968 the Committee for Engineering Student Development (CESD), whose goal was to encourage individuals from underrepresented segments of society to enter the engineering profession. After a rough start, a relatively successful program has evolved and has been in operation for more than a year. It supplements the Berkeley faculty-sponsored Special Opportunity Scholarship (SOS) Program.

SOS recruits high school students of disadvantaged backgrounds who have shown potential in some way and provides them with a program that improves their academic preparation for college. CESD provides to those participating students an engineering interest group. This class meets for four afternoons a week in the summer and on Saturdays during the school year.

Many of the students have written assembly language programs and have run them on a PDP-7 computer; also, students are writing programs in Fortran and are running them on an IBM 1620. Thus, the committee feels that it accomplished its main purpose by developing the students' interest in engineering and their confidence in their ability to enter the engineering profession.

To date, the only financial support for CESD's program has come from student engineering organizations, and their funds have been depleted. The San Francisco Section has agreed (subject to approval by higher IEEE authorities) to contribute some money, but even so it will cover only 15 percent of a year's budget. CESD is now applying for financial support from Western Electronics Education Fund (WEEF). WEEF was established by Western Electronics Manufacturer Association and the IEEE, and its purpose is to encourage young people to enter the electrical engineering and allied professions. It is hoped that CESD will receive from the IEEE Section and WEEF the substantial on-going monetary support the program is going to need in order for it to continue. Because of the transitory situation of students, this program undoubtedly would benefit from the personal participation of the established members of the profession and so CESD also hopes that it will receive their support as well.

Terry M. Kvam
CESD, U.C.L.A.
Berkeley, Calif.

In response to Jesse B. Ring's letter, there are a few points that I would like to bring out.

It is meaningless to call any profession "open" if large segments of society do not enter it because of blocks that are social, educational, economic, or psychological in nature.

A number of professional societies in the United States, such as the American Bar Association and the American Institute of Architecture, have noted that there are extremely few members from minority groups in these professions. Also many universities have taken a hard look at their student bodies and faculties and have found that they, too, have few members from minority groups. A number of industries have noted this same phenomenon.

Although all of these organizations can assert that anyone who is qualified can join them, it is evident that there are, in fact, individuals from certain segments of society who do not.

In a number of instances, significant changes in the percentage of minority group members have been made if the organization involved has developed a positive program of recruitment. Frequently educational, economic, social, and mental blocks have had to be removed. When they are, the number of members of minority groups within the organization have started to approach the percent of these members in society at large.

It is my position that a profession or organization that ignores the problem is not as "open" as one that undertakes an active program to correct the situation.

A clear example of the type of work that must be done to inform members of these minority groups that they have an opportunity to join our profession is the work of the Berkeley Student Branch of IEEE. This is exactly the type of program that I feel is necessary to help right some of the imbalances that presently exist in our society.

F. Karl Willenbrock
Senior Past President, IEEE

Logic functions

This is to bring to your attention a few simple errors in Part I of L. S. Garrett's article, "Integrated-Circuit Digital Logic Families," which appeared in the October issue of *IEEE Spectrum* (pp. 46–58).

The logic expression for the output of the NAND gate of Fig. 12, as shown in Fig. 14, should be A B C D, not ABCD; i.e., the two expressions given in Fig. 14 should be equivalent by De Morgan's theories. Also, a similar error appears in Fig. 24, but this error is on the NOR expression, which should read $\overline{A} + \overline{B} + \overline{C} + \overline{D}$. In Fig. 29(A), the schematic diagram for DTL shows D_1 as a Zener diode, but it should be shown as a regular diode.

H. P. Anderson
Bell Telephone Laboratories
Denver, Colo.

There are several common conventions for representing logic functions. The MIL-STD-806B convention is most commonly used throughout the IC industry and was followed in the article. The small circle at the input or output of a distinctive shape (AND, OR) does not necessarily indicate inversion, but only the voltage that activates that function and the output voltage level that appears when it is activated. A circle on an input indicates that a low level at that terminal activates the function and a circle on an output indicates that the function has been activated when the output is low. For example, a positive logic NAND function can be stated as: only when all inputs are high will the output be low. The difference between positive logic and negative logic depends upon which levels, (H) or (L), are defined as true or logical "1." If all (H) levels are defined as 1 or true, we are using the positive logic convention. If all (L) levels are defined as 1 or true, we have the negative logic convention. These two conventions produce dual functions, not complement functions, as can be ascertained from the level table of a given function. A NAND gate is the dual of a NOR gate. The two are not equivalent by De Morgan's theories. This also holds true for other specifications such as ASA Y32.14. Figures 14 and 24 in the article are correct according to the commonly accepted conventions.

Figure 24(A), which shows a schematic diagram of the basic modified DTL gate, is in error. D_1 should be a regular diode and not a "Zener." The error crept in during the final *IEEE Spectrum* layout. Technically it is a moot point anyhow since in IC processing a base–emitter (base–collector short-circuited) diode is normally used. This transistor connection does have a "Zener" breakdown in the reverse direction and a normal diode characteristic in the forward direction.

Lane S. Garrett
Motorola Semiconductor Products Inc.
Phoenix, Ariz.

Least-squares estimation

We wish to make a brief comment concerning an item in Prof. H. W. Sorenson's article, "Least-Squares Estimation: from Guass to Kalman," which appeared in July.

The author implies on page 65 that the $(n + 1)$-order matrix inversion required for the solution of the vector-matrix equation associated with the Wiener filter becomes computationally impractical when n is large. If the matrix is inverted by conventional means, required computer storage is proportional to n^2 and the associated computer time is proportional to n^3. In 1947 Levinson made use of the special structure of this matrix and published an algorithm that leads to a filter solution with storage requirements proportional to n and computer time proportional to n^2. (Levinson, N., *J. Math. Phys.*, vol. 25, no. 1, pp. 261–278, 1947; also published as Appendix B in Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: Wiley 1949.) The resulting gain in computing efficiency has rendered the Wiener filter an invaluable tool in our own field, seismic prospecting.

Those who wish to pursue these matters further are referred to the copious literature that has appeared over the last decade in such journals as *Geophysics* and *IEEE Transactions on Geoscience Electronics*.

S. Treitel
R. J. Wang
Pan American Petroleum Corp.
Tulsa, Okla.

Upon rereading the statement in my article to which Messrs. Treitel and Wang refer in their letter, I must agree that the statement can be regarded as being stronger than is warranted. I did not intend to imply that the Wiener filter for a finite data set has seen no application. This is obviously not the case. However, the statement that I make in the article can be interpreted as implying this and as a result may be misleading. It would have been more appropriate to include a comment such as made in the letter. Thus, I think that the point is well taken and thank the writers of this letter for bringing it to my attention.

Harold W. Sorenson
University of California, San Diego
La Jolla, Calif.

Confusion of noise terms

Much confusion is caused by the somewhat vague definitions of standard terms. Perhaps this is worst in the field of noise. Take, for example, the recent article on "Noise in Amplifiers" by Letzter and Webster (*IEEE Spectrum*, pp. 67–75, Aug. 1970). Here the authors differentiate between noise figure and noise factor although the IRE Standards Committee gave the two as equivalent (*Proc. IRE*, vol. 48, p. 61, 1959). Then the authors define noise figure in terms of . . . "Input power SNR (amplifier disconnected) . . ." What does this mean? Surely input power with amplifier disconnected is zero. I appreciate their difficulty caused by the frequent use of "available input noise power" and "available output noise power" in the noise figure definitions, in which the word "available" does not always seem to mean what it should, i.e., power transfer under matched conditions.

Perhaps someone should be instructed to sort this one out.

M. H. N. Potok
Royal Military College of Science
Swindon, England

Unaware of the IRE Standards Committee position, we differentiated in our article between noise figure and noise factor insofar as we took the former term to be a logarithmic representation of the latter, expressed in decibels, e.g., noise figure (in dB) = 10 log (noise factor). This differentiation of terms has been used by others (*Handbook of Military Infrared Technology*, sponsored by Office of Naval Research, Washington, D.C., 1965, p. 601, Eqs. 14-97 and 14-98).

As to the noise figure definition—"Input power SNR (amplifier disconnected) . . ."—I would substitute "mean-square volts" for "power" so Eq. 7(b) would be: "Input mean-square volts SNR (amplifier disconnected) . . ."

Seymour Letzter
Princeton Applied Research Corp.
Princeton, N.J.

Spectral lines

Birth of a concept. The Board of Directors has approved the formation of three IEEE Societies, effective January 1, 1971. Responding to the initiatives of the existing Groups in these areas, the Board approved the IEEE Power Engineering Society, the IEEE Computer Society, and the IEEE Control Systems Society. These actions are the culmination of more than three years of study and analysis by the Technical Activities Board, originally under the leadership of James H. Mulligan (now IEEE President) and, currently, Harold Chestnut, Vice President for Technical Activities. Together with the related reorganization of the Board of Directors, approved last fall, which established Divisional Directors responsible for each of the six groupings of IEEE technical activities, this move is the most significant step taken in the organization of IEEE's technical activities since the "Professional Group" concept was introduced in IRE by W. R. G. Baker more than 20 years ago.

The objectives of this new development are defined in a policy statement adopted by the Board:

"The change should stimulate the member of an IEEE Society to greater participation in its activities. Increased technical activity, in turn, will benefit the IEEE member who does not belong to one of these Societies. The change should attract new members in competition with other societies, particularly from disciplines other than electrical engineering. It should improve the image, prestige, and effectiveness in dealing with non-IEEE professional societies, with government agencies, and with the public at large. It should provide a good basis for merger of Groups in closely related fields and for bringing non-IEEE societies into the IEEE."

In this policy statement, the Board established the criterions to be considered in forming a Society from one or more constituent Groups. These involve size and growth rates, financial viability, and the scope and significance of the associated publication and meeting activities. Although some overlap with the scope of other IEEE Groups is in all likelihood inevitable, it is fundamental that an IEEE Society should be the primary organizational element in a major technical area within the scope of the IEEE.

The proliferation of technical and application specialties within the broad area of electrical and electronics engineering began in World War II, and has continued, virtually unabated, ever since. The IRE Professional Group concept was an effort to provide workable organizational mechanisms for assuring a positive response to these developing interests. By and large, the Group system has been successful in this regard. Two decades of experience with it have shown, as might have been expected, that the various Groups progress at different rates and develop different patterns of service and, as a corollary, experience different patterns of income. One reason for this, but clearly not the only reason, is the obvious fact that specific technologies and applications specialties produce differing economic impacts and may, or may not, lead to new curricular specializations in engineering education. Some of these specializations prove to be of major importance, involving large numbers of engineers and enjoying widely based economic success in the marketplace. Others, concerned with important technical specialties involving a more limited number of practitioners, find an important niche in the broad IEEE program without impacting directly on the marketplace. Still others never gain significant numbers of adherents, although the individuals involved may produce important work and frequently exhibit that tenacious dedication to their specialty that turns up elsewhere in, for example, railroad buffs and Sanskrit scholars.

Recognizing that the resources of the IEEE—staff and volunteer manpower as well as financial—are too limited to hope to satisfy all of the demands put on them, the TAB and the Board of Directors have been concerned for some years with the problems of differentiating among the multiple constituents of IEEE's technical activities. The creation of IEEE Societies is an important first step. As this concept matures, it is reasonable to foresee the emergence of a (relatively small) number of IEEE Societies, each in an area of major technical activity of broad interest and each possessing a membership of sufficient size and vitality to insure continued successful service to the profession as a whole as well as to specialists in that field, operating with a substantial degree of managerial and financial autonomy. A larger number of the present Groups, and in all likelihood a number of Groups still to emerge, will (hopefully) find the traditional Group concept, focussing on rather sharply defined technical specialties and drawing on the managerial and financial resources of the whole IEEE for efficiency and to insure continued viability, of continuing utility. The most difficult unresolved problem relates to that small number of present Groups who have never achieved "critical mass," either in size or in the quality and significance of the services they perform. Some will undoubtedly find merger with other Groups with compatible interests a sensible course of action. Some might function more effectively as technical committees, concentrating on *ad hoc* projects rather than attempting to sustain a schedule of regular publications and meetings for which new and significant contributions, in the required quantity, may not be available. The total problem will not be solved, however, until the IEEE evolves a successful method for achieving the demise of Groups whose formation, in retrospect or in the face of subsequent developments, proved to be a mistake. *J. V. N. Granger, IEEE President, 1970*

The application of electron/ion beam technology to microelectronics

Although beam techniques are not yet widely used or fully developed, their unique features qualify them as a potential major element in the fabrication of future microelectronic devices

George R. Brewer Hughes Research Laboratories

The use of electron and ion beams in the field of microelectronics has grown rapidly over the past ten years. For example, fine-focused electron beams, which initially were employed in diagnostic tests on integrated circuits and semiconductor devices, now are being utilized in the laboratory for creating devices and circuits with resolutions considerably higher than the values achievable by photolithography, and ion beams are being employed in the production of integrated circuits. It is seen that the beam approach offers a number of advantages compared with contemporary techniques. This article describes some of the devices that can be fabricated with electron beams, thus indicating the state of the art, discusses the advantageous features of ion implantation doping, and points out the limitations on beam processes imposed by fundamental electron/ion optical effects.

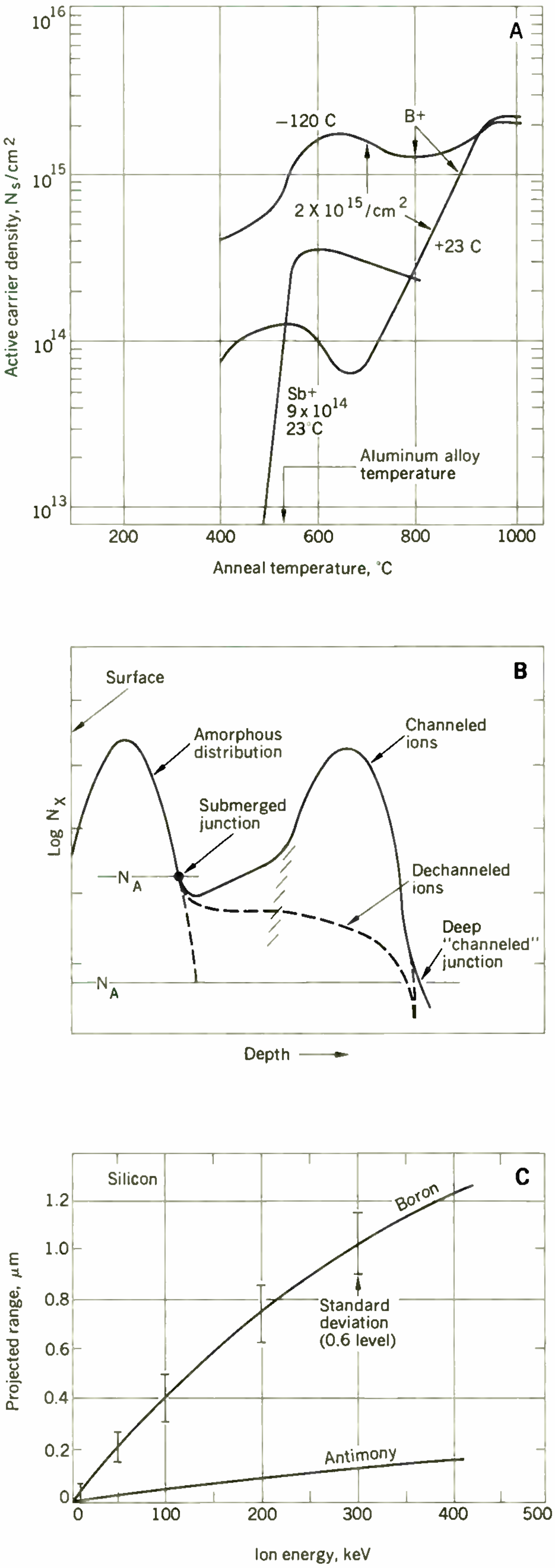
The trend in the development of microelectronic devices and circuits for many years has been toward higher performance (higher frequency, higher power, lower noise), higher process yield, and greater reliability. Truly phenomenal advances have been achieved, largely as a result of improvements in the technology employed to fabricate the devices and circuits. Several recent comprehensive reviews[1-4] of the technology of integrated circuits have brought out this growth in performance, reliability, and yield resulting from the evolutionary improvements in fabrication techniques. Moreover, this progress has been made using techniques that are basically extensions of the technology developed originally for fabricating silicon transistors plus that of thin films; the pattern definition, for instance, is achieved through photolithography. However, these fabrication methods are now being pushed to their limits, some of which are rather fundamental. For example, the resolution obtainable with photolithography is limited by lens aberrations and by diffraction in the projection aperture or by the mask. It thus appears that the next major advance in the performance of microelectronic devices and circuits will be made with the aid of qualitatively different fabrication technologies.

Several of the more critical limitations of contemporary microelectronics technology can be overcome by the use of electron and ion beams. This article presents a view of the application of these beams to the fabrication of microelectronic devices and circuits. Although the unique or advantageous features of such beams for such processes will be shown, it is made clear that they do not provide a magic solution to all present-day fabrication problems. This is true because electron and ion beams suffer from rather fundamental limits of their own, the implications of which will be described. To be effective, the beams must be used so as to take advantage of their unique features and with full knowledge of these limiting factors. The current state of the art in beams is used as a basis for projecting their potential capability for fabricating advanced devices and circuits in the years ahead. (Note that in this article the much-used term "microelectronics" is applied to include both digital-

I. Attractive features of beams for microelectronics

Resolution Using Electron Beams	Improved Processes Using Ion Beams
Spot size $\leq$ 500 Å versus 5000 Å = λ_{light}	Ion implantation doping: Low temperature → simpler processes Low temperature, S/C materials
Electron beam → resist → <0.2-μm line	Precise dose control
therefore, higher density, faster LSI	Little lateral spread Uniform junction depth
Beam deflectable, thus programmable electrically up to ~10 000 spot diameters	Variable doping profile Doping through oxide Ion-beam sputtering: High-vacuum (clean) deposition and removal

FIGURE 1. Principal features of ion implantation doping. A—High electrical activity possible at low temperature. B—Wide variety of doping profiles. C—Depth versus energy for implantation into an amorphous material. For a dose of 10^{10}/cm^2, reproducibility of the implanted dose is ~ 15 percent; for 10^{17}/cm, ~ 2 percent.



and analog-type integrated circuits as well as various forms of microwave devices.)

Advantages of beams

Some of the principal advantageous features of using beams of electrons or ions in a microelectronic fabrication process are summarized in Table I. In general, electron beams can create patterns with higher resolution than is possible with photolithography because beams can be focused to a spot diameter that is two or three orders of magnitude less than the wavelength of light. Light rays passing through a photographic mask are diffracted to a line broader than the slit in the mask. Similarly, electron scattering will broaden the pattern drawn by an electron beam, but the resulting minimum line width is much less than that of the light beam. The electron beam can be deflected by electrical control and programmed to write a complex pattern. Thus, the pattern can be created directly from control signals, such as derived from a computer.

Ion beams offer a qualitatively different means for introducing dopant atoms into a crystal than that involved in diffusion,[5] thus providing several practical advantages. The doping process can be carried out at a relatively low temperature, compared with equivalent diffusion temperatures, which raises the possibility of implanting materials for which diffusion is difficult. The basically electrical nature of the ions permits the precise measurement of the number of atoms introduced (the dose). The ions penetrate the crystal in the direction of their approaching trajectory; as a result there is significantly less lateral scattering than from diffusion. Since the depth of penetration depends principally on the ion energy, the depth of the junction surface can be made very uniform over large areas. The ions can be introduced after a passivating oxide is grown, resulting in very good junction characteristics. Also, the profile of doping density versus depth can be varied over wide limits. Thus, ion implantation offers a number of new approaches to processes that allow the fabrication of improved devices or of new classes of semiconductor devices. Ion-beam sputtering as a means of material removal or deposition also provides a means of achieving clean surfaces or impurity-free layers as well as high-resolution machining.

Ion implantation

Let us examine in a little more detail some of the principal features offered by ion implantation, as illustrated in Fig. 1. This doping process is fundamentally different from that of diffusion in several respects. Implantation is not an equilibrium process—the motion of the ion into the crystal results from its initial kinetic energy, not from a concentration gradient, as in diffusion. Therefore, the usual solid solubility limits of dopant atoms sometimes can be exceeded. Further, the ions can be injected into a relatively cold crystal. Whereas subsequent annealing is usually necessary to achieve a high percentage of electrically active dopant atoms, the annealing temperature can be well below comparable diffusion temperatures. This means, for example, that semiconductor processing steps can be varied substantially from those followed when diffusion doping is employed; typically, doping can follow aluminum metalization if desired. It also means that low-melting-temperature materials such as InAs, as well as very-high-diffusion-tem-

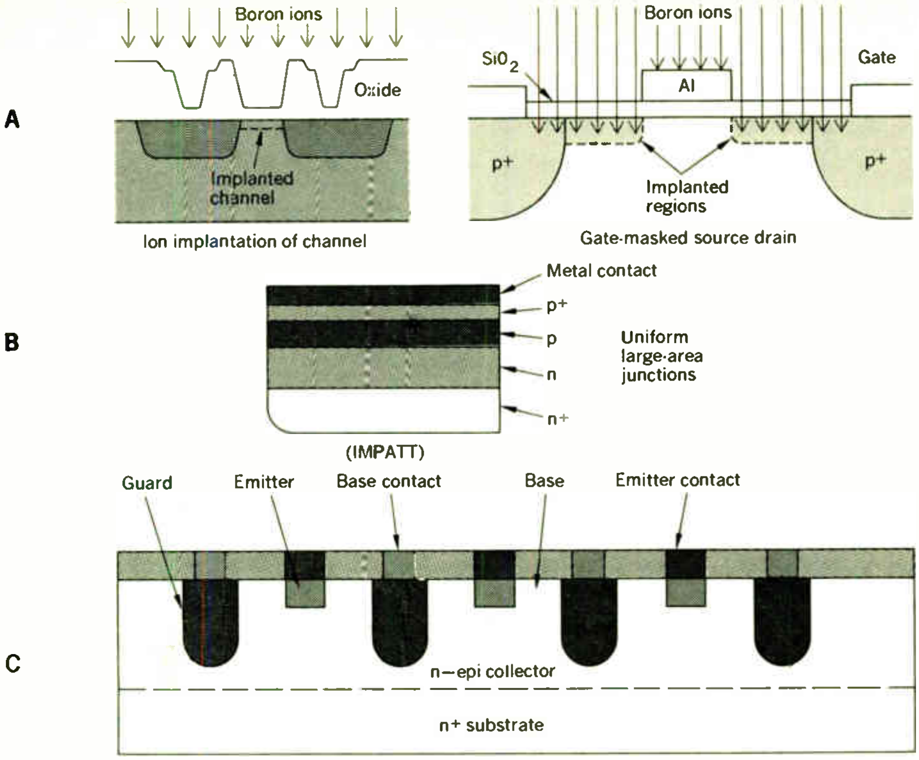


FIGURE 2. Examples of the uses of ion implantation in the fabrication of several classes of semiconductor devices. A—Threshold adjustment and source-drain extension of MOS devices. B—Creation of uniform large-junction area. C—Fabrication of a bipolar transistor by implantation and diffusion. Special low-dose devices are used in microwave amplifiers and transistors.

perature materials, such as SiC, can be doped in short times at comparatively modest temperatures. Typical data on the density of active carriers—the effectiveness of converting ions implanted into a silicon crystal into active carriers—are shown as a function of annealing temperature in Fig. 1(A). It is seen that under proper conditions (for example, a low crystal temperature with boron implantation) most of the dopant atoms can be made electrically active at temperatures below those for aluminum alloy formation in silicon or for diffusion of the same dopant.

Two other advantages of ion implantation are the control of junction depth and the wide range of density profiles attainable.[6] The penetration depth and final distribution of the ions in the crystal depend on the ion energy, the crystal and the ion species, and the angular alignment of the ion beam with the crystal axis; these characteristics do not depend strongly on the crystal temperature.[7] We can distinguish two primary classes of ion-crystal interaction that yield quite different penetration depth and ion-density distributions. Ions penetrating a target that is either amorphous or crystalline but with the ion path misaligned from any crystal axis will suffer collisions that reduce their inward motion in such a way that the resulting density profile with depth is roughly Gaussian. This case is illustrated in Fig. 1(B), where it is shown as the "amorphous" distribution near the surface.

If the ion dose is low and the trajectories are directed precisely in an open crystallographic direction, the ions can penetrate deeply into the crystal (this is called channeling), and will stop rather abruptly at the end of their range. The resulting profile is shown in Fig. 1(B) by the curve marked "channeled ions." Some fraction of the ions traveling along a channel leave the channel prematurely and are shown as "dechanneled ions" in this illustration. However, the channeled distribution cannot be maintained at high dose levels since energetic ions striking the surface of a crystal will displace the crystal atoms, resulting in a near-amorphous condition close to the surface. Although the ion dose required to create this amorphous layer varies somewhat with the ion and crystal species, it is usually about 10^{14} ions/cm^2 or more. Thus, with high dose levels, the implanted distribution will always be near-Gaussian.

Many practical applications of ion implantation have been successfully based on the approximation to the Gaussian amorphous distribution. Figure 1 shows the depth of the amorphous doping peak—and an indication of the width or standard deviation of the distribution—in silicon as a function of beam energy for a light ion (boron-11) and a typical heavy ion (antimony-122). It is seen that the typically desired junction depths ($\sim$ 0.1 to 1 μm) can be attained with ion energy between 30 and 300 keV for most ions. Thus, a wide range of useful doping processes can be carried out using relatively simple ion accelerators and power supplies. By properly superposing implants at different energies and doses, or even altering the ion-crystal alignment, the device designer has greater flexibility in obtaining his desired impurity distribution than he does with conventional diffusion.

One of the very practical advantages offered by ion implantation as a doping process is extremely precise control of the number of ions implanted in a given area. This advantage is, of course, the result of the basic electrical nature of the ion beam and the fact that highly precise electric current measurements can be made, resulting in devices of greater uniformity.

The density of implanted ions necessary to produce typical desired semiconductor characteristics is up to

$\sim 10^{15}$ ions/cm² of surface area. The doping time is thus $\tau \approx 10^{15} \times 1.6 \times 10^{-19}/J_t$, where J_t is the ion current density at the target. A typical value of J_t is $\sim 10^{-5}$ A/cm², so, in this example, the implant time is $\sim$20 seconds. The number of ions implanted is seen to correspond roughly to one monolayer on the crystal surface. The ion implantation systems and sources therefore must be capable of delivering to the target a current density of the order of 10 μA/cm² of the desired ion species. The ions desired are generally those that will become substitutional donors or acceptors in the crystal. The types of ions required and the modest energies used thus are the main factors differentiating ion implantation system technology from much of the prior ion-beam-generation work.

Figure 2 illustrates some of the ways in which ion implantation doping has been used to improve semiconductor devices. In MOS devices, ions implanted through a thin oxide have been utilized to introduce charges in the channel region, thereby allowing adjustment of the threshold voltages of the device. In particular, devices with very low values (1 to 2 volts) of threshold voltage are manufactured with little increase in process complexity.[8] The use of ion implantation, again through a thin oxide, to extend the diffused source and drain to exactly the edge of the gate [see Fig. 2(A)] yields a device with minimum parasitic capacitance.[9] In the first MOS example, the precise dose control possible with implantation allows close control of threshold voltage. In the second example, the property of the implanted ions to penetrate straight into the crystal without significant lateral spread allows the gate to be used as an *in situ* mask to achieve self-alignment with the edges of the channel.

Figure 2(B) illustrates those classes of semiconductor devices that require a large-area p-n junction with very uniform junction surface. They include IMPATT microwave diodes, photodetectors, and electron-beam semiconductor devices. These devices are generally operated near a breakdown level of electric field. Because of the uniform depth of penetration of the ion-implanted dopants (depth relatively independent of crystalline imperfections), sharp protrusions from the junction do not develop and the tendency to form microplasmas at the junction is greatly reduced. An ion beam that is raster-scanned over the target provides very uniform (< 1 percent variation) dose over large areas (>1 cm²). Ion implantation has made possible IMPATT microwave oscillators[10] and photodetectors with improved performance and/or greater reproducibility.

Figure 2(C) shows one concept of a bipolar transistor made by ion implantation or a combination of ion implantation and diffusion. The flexibility of ion implantation doping allows the creation of deep and shallow regions, high and low concentrations of n or p dopants, so, in principle, all of the required impurity regions can be created by implantation. A narrow emitter width required for high-frequency operation can be obtained by implanting through a mask. We shall show later that lines narrower than 0.5 μm can be created by an electron beam; in this way a contact mask can be made to delineate the doped region very precisely. This mask could be a 0.75-μm layer of aluminum or a 2-μm layer of polymethylmethacralate (PMM) resist, either of which will stop the 10-keV boron ions.

An example of a microwave device requiring an extremely low dose ($\approx 10^{10}$ to 10^{11} ions/cm²) with the dopant atoms close to the surface ($\approx$1000 Å deep) over a large area ($\approx$3 cm²) is the semiconductor part of a surface acoustic wave amplifier.[11] Devices of this type have been manufactured with dose control of about 15 percent at 10^{11} ions/cm² and the amplifiers have performed well. High-resistivity (greater than 10 kΩ/square) resistors have been achieved throughout ion implantation[12] as a result of the technique's shallow-depth, low-dose, and close-dose-control capabilities. Such values are difficult to attain by diffusion.

The application of ion implantation to an integrated circuit is illustrated in Fig. 3. This circuit is a 2048-bit read-only memory in which the memory pattern is created by selective ion implantation to activate the desired MOS device.[13] Since the implantation, using resist mask, is the final step in the processing, the uncoded circuits can be made and stored prior to coding. Further development of the technology should enable the memory pattern to be written with a small programmed ion beam; that is, without a mask. The circuit exhibits a 100-ns cycle time and is contained on a 0.25-cm by 0.25-cm chip.

Patterns by tiny electron beams

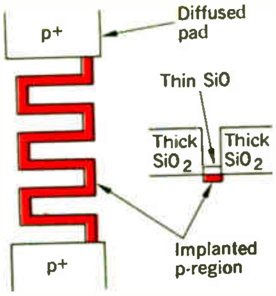
Some of the attractive features of electron and ion beams for use in microelectronics fabrication are summarized in Table I. Electron beams have been focused to a spot size as small as 5 Å in diameter[14]; 100-Å beams are used routinely. Since these dimensions are a small fraction of the wavelength of light, it is clear that beams can "see" and make patterns with much higher resolution than is possible using light. These beams of charged particles can be programmed in intensity and positioned over wide ranges by electric signals. Thus we should be able to create by an electron beam the very tiny patterns or masks needed to define the elements of a high-performance microelectronic device or circuit.

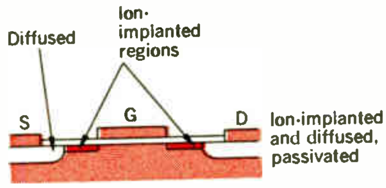
There are two general ways in which beams can be used to irradiate a surface and create a pattern: parallel exposure of all pattern elements at the same time and sequential exposure of one pattern element at a time. The simplest sequential method is to scan the surface with a raster pattern as in a television picture; the beam is turned on and off along each line where it is desired to expose the surface. Analog input signals, such as from a flying-spot scanner looking at an enlarged mask of the desired pattern, control the beam current. The bandwidth required is determined by the time needed to turn the beam on and off in one beam diameter (to attain the resolution possible with a spot diameter d_s). This bandwidth is given by the equation:

$$B_w = \frac{L^2 F}{2 d_s^2}$$

where L is the length of the scan line and the height of the raster, F is the frame rate, per second, and d_s is the electron spot diameter. In order to scan a large area, such as a square centimeter, with a fine beam perhaps 2 μm in diameter, a very large bandwidth—around 375 MHz—is required with television scan rates (F = 30/s). Thus, to achieve good resolution, a slow-speed flying-spot scanner must be used as the modulation technique.

Another method of sequentially generating a pattern by making use of digital input data is illustrated in Fig 4(A). In this case the fine-focused beam is programm





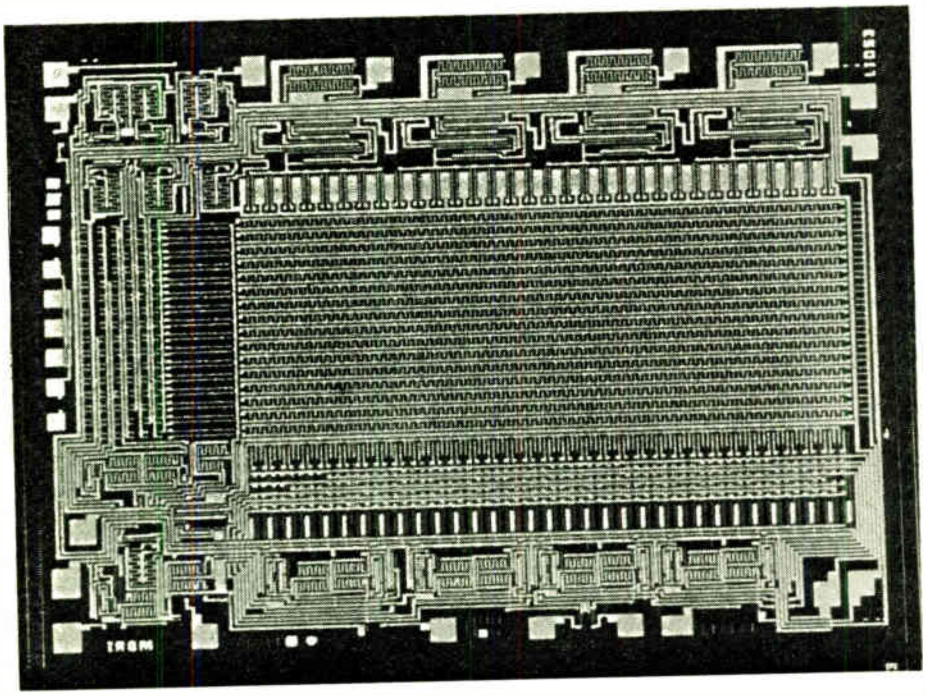


FIGURE 3. A 2048-bit MOSFET read-only memory made by ion implantation. Illustration shows more than 3000 MOS devices on a 0.25- by 0.25-cm chip. Insets show techniques for implanting the MOS device and resistors.

by digital signals to expose selected areas of the desired pattern. Each pattern element is exposed by sequentially stepping the beam along a row of spots, with enough rows used to make up the pattern element. The deflection current is a staircase with the beam blanked off between step levels. Four units of time influence the total writing time of this system. The path time from the last spot on one element to the first spot on the next element is related to the slewing rate of the scan coil and amplifier in the beam deflection system; typically it will be about a microsecond. The time required to output an address word from the computer to the D/A converter in order to direct the beam to a new spot location on the same row will be about 2 μs per spot, and the time required for the deflection current to settle to its next level value on the staircase—so that the beam can be turned on—will be about 10 μs. The fourth element of time is that required for the beam to deposit the requisite charge on the surface in order to carry out the process desired; for example, to expose the resist material. It is assumed in the above that the exposure time is long enough to permit the calculation of the next spot location and the loading of this digital value into the I/O buffer register preceding the D/A converter.

These two techniques of sequential exposure can be combined in such a way that only the initial and final coordinates of a rectangular area need be addressed from the computer. The area of the rectangle is filled in by raster-scanning the beam so as to start and end at the desired coordinates. This combined digital–raster method should provide faster writing speed at the sacrifice of some additional electronic complexity.

It is noted that the bandwidth required for this sequential digital means of writing will be one megahertz or less. The resolution attainable will be limited by either the digital-to-analog converter or the spot diameter, depending on the system design. The technique of beam writing illustrated in Fig. 4(A) offers several advantages in that the beam is scanned only over the areas where it will write, the required bandwidth is low for a given resolution, and the pattern programming can be done very accurately and changed easily in the digital computer.

One parallel-exposure technique involves the use of a focused beam that is caused to pass through a transmission mask so as to create a demagnified image of the mask on the target.[15] Since such a mask must be self-supporting, the type of patterns that can be created is limited (e.g., an annular ring cannot be exposed). A parallel-exposure method that does not suffer from this

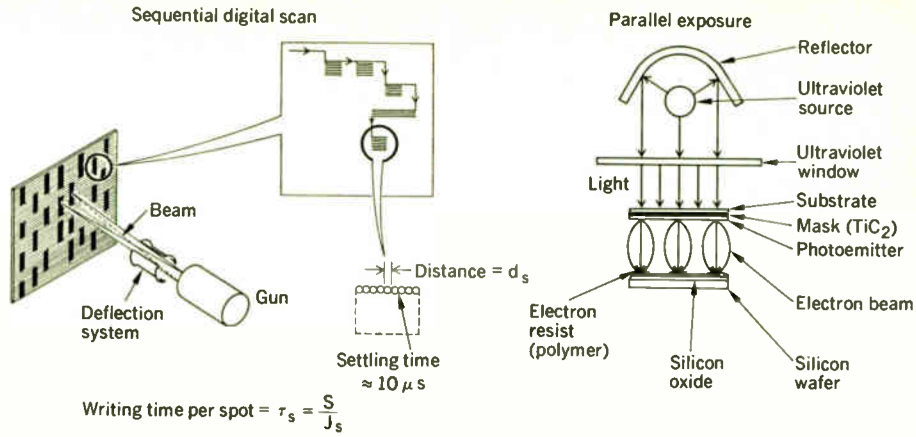


FIGURE 4. Two classes of technique for electron-beam pattern exposure. In the parallel-exposure method, the pattern in the mask is created by an electron beam using a sequential writing technique.

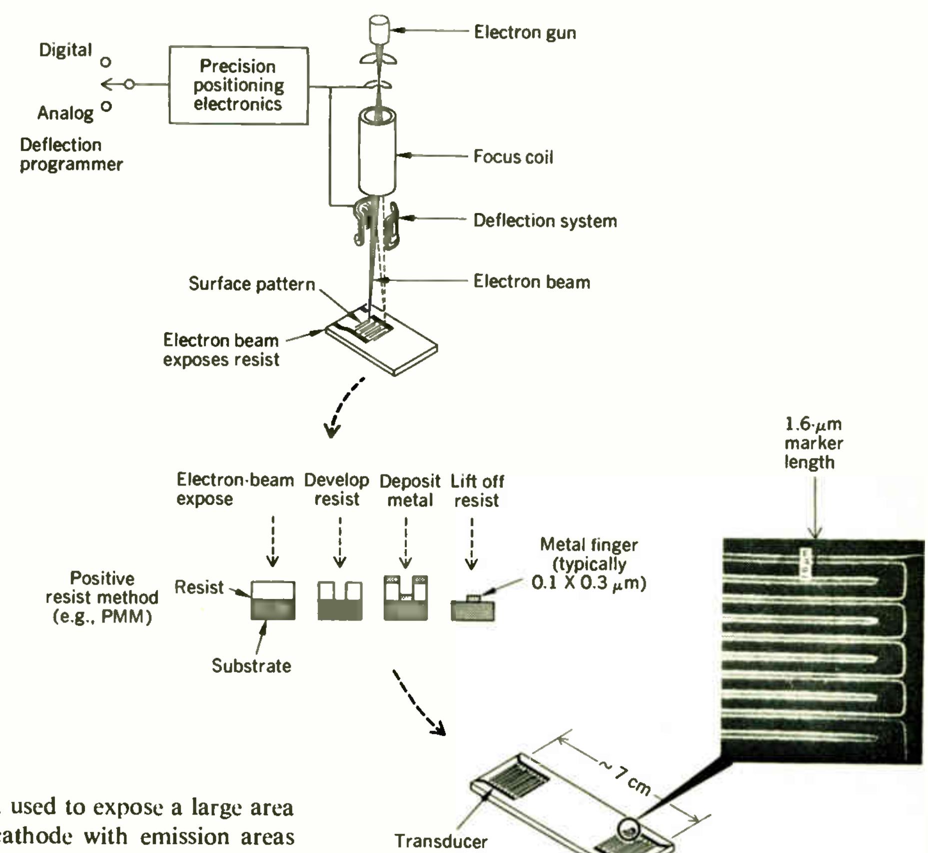


FIGURE 5. Technique and steps in the fabrication of a surface acoustic wave transducer by use of a programmed electron beam.

restriction and that has been used to expose a large area (5-cm diameter) utilizes a cathode with emission areas made to define the pattern; this object is then imaged onto the target by an image-orthocon-type lens system.[16,17] The technique is illustrated in Fig. 4(B). A new cathode is made for each new pattern; to attain high resolution, the pattern on the cathode is written with a tiny electron beam as in Fig. 4(A). A 2-mm by 2-mm field can be written, with the total cathode pattern created by step and repeat. The parallel-exposure technique in general provides shorter pattern exposure time, for a given area, than the use of sequential exposure, but it is not as flexible or rapidly programmable for different patterns. Minimum line width of the exposed resist of 0.5 μm has been achieved. Registration between successive masks of $\pm 2\ \mu$m has been demonstrated, with potential improvement to $\pm\ 0.5\ \mu$m.

Multiple-field parallel exposure of resist by electron beams focused by a multiple-lens "fly's-eye" system has been demonstrated[18] to yield patterns with about a micrometer resolution. This technique, however, also uses a transmission mask, for which the pattern geometry is limited, as mentioned previously.

Examples of electron-beam patterns

Let us look at two examples of the type of high-resolution patterns that have been generated by electron beams, using the sequential digital scan technique shown in Fig. 4(A). Figure 5[19] shows the programmed electron beam used to expose a positive electron resist (polymethyl methacrylate) that is developed to yield a mask through which metal is evaporated or sputter-deposited to yield the desired device. The device illustrated here is

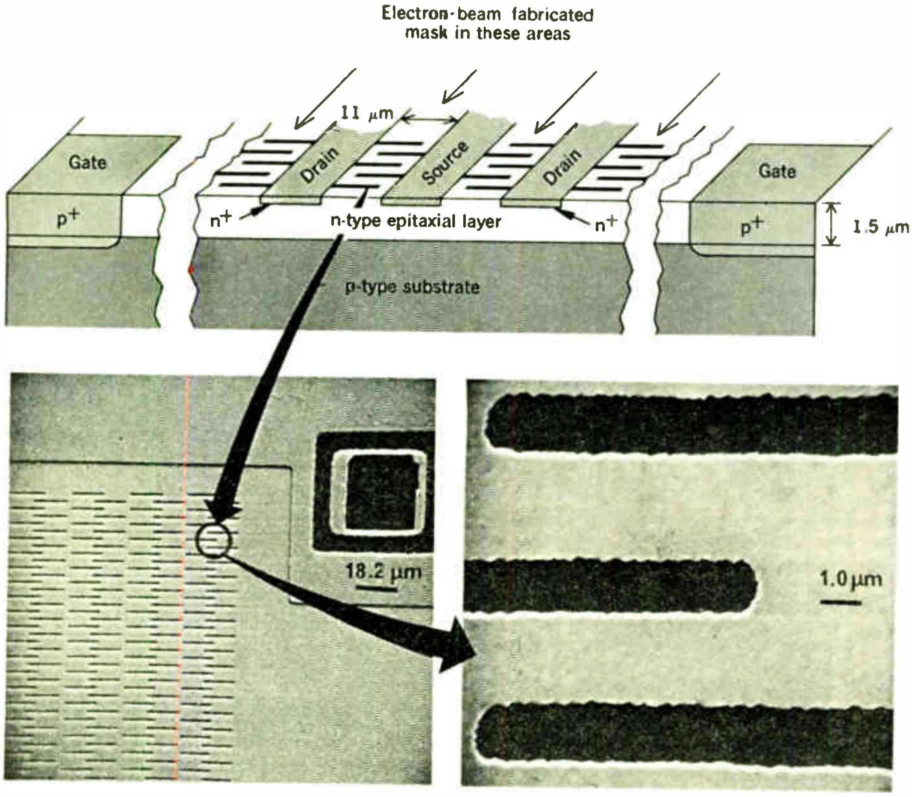


FIGURE 6. Exposure of resist by an electron beam to yield windows in a gold layer through which fingers are ion-implanted to make the high-conductance FET device shown.

an interdigital transducer used to couple to a surface acoustic wave on a piezoelectric crystal at a frequency of about one gigahertz. The transducer is about 0.4 mm square, composed of fingers about 0.5 μm wide and 0.1 μm thick. Similar patterns with finger width as narrow as 0.1 μm have been generated by a specially designed electron beam using the raster-scan method, with a flying-spot-scanner input.[20,21]

The second example is a somewhat more complex pattern to create narrow ion-implanted fingers for a junction field effect transistor (JFET) microwave switch[22]; see Fig. 6. In this device, the source and drain are diffused into the epitaxial layer that acts as the channel. The device conductance is increased by extending the source and drain electrodes by means of an array of interlocking fingers that are fabricated starting with electron-beam exposure of resist that covers a gold layer.

The developed resist is used as a mask against ion-beam sputtering to define the finger pattern in the gold, which is then used as a mask for ion implantation of the fingers into the epitaxial layer. The finger pattern and an enlargement—scanning electron microscope (SEM) picture—of the fingers are shown in Fig. 6. The implanted fingers are about 1.3 μm wide, with an edge resolution of 0.1 to 0.2 μm. A single device involves a finger pattern about 0.4 mm square, with 1180 fingers.

Resolution limits

One of the principal advantages of electron beams as a fabrication tool for microelectronic devices lies in the higher resolution of the written pattern, which permits fabrication of devices that will operate at higher frequencies, with higher packing density, and so forth. The resolution, however, is dependent not only on the diameter to which the electron beam can be focused, but on any scattering suffered by the electrons as they enter the resist material. The process or writing speed depends on the current density in the beam. A number of physical effects limit the attainable spot size and current and thereby the maximum attainable current density in a beam of charged particles; these include lens aberrations and transverse thermal velocities. Other effects limit the total size of pattern that can be created (deflection distortions) and restrict the minimum resolution of pattern (electron scattering in the resist). These problems will be discussed briefly to show some of the limitations inherent in the electron-beam fabrication process.

It is well known that the electrons of a beam impinging onto a resist material will be scattered through large angles by the atoms of the resist. Under some conditions these backscattered electrons can expose the resist through a roughly spherical volume that is larger in diameter than the beam spot diameter. This effect is illustrated in Fig. 7, which shows a SEM photomicrograph[23] of a cleaved layer of thick electron resist (PMM) into which a 500-Å electron beam has been scanned perpendicular to the plane of the photograph, producing a neck at the surface about 2000 Å wide and a cylinder about 4500 Å in diameter, or about nine times the diameter of the beam. Since the thickness of the resist is usually about 0.2–0.5 μm, the exposed line actually will be about 2000 Å wide, or four times the beam diameter. Although Fig. 7 shows an extreme example to illustrate the effect, electron scattering can result in an attainable resolution that is much poorer than that of the electron beam. This scattering problem has been studied experimentally and analytically by several workers.[24,25] A recent paper[26] provides curves, supported by experimental data, from

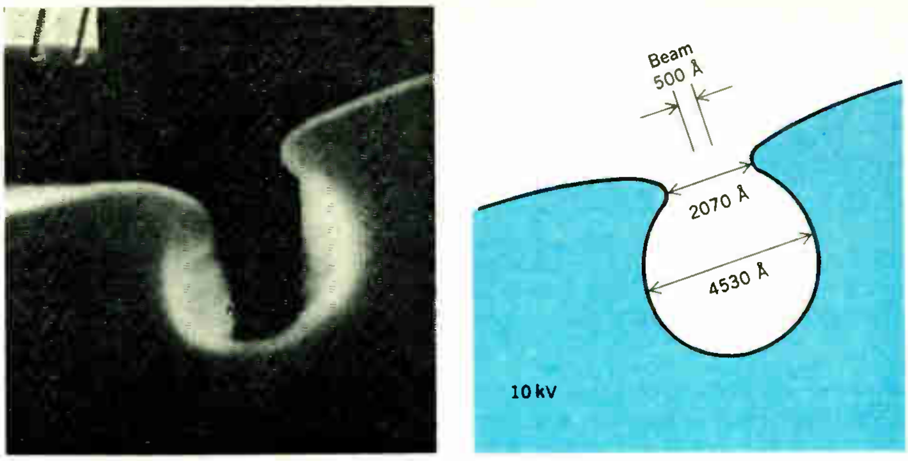


FIGURE 7. Electron-beam spreading in resist material creates an exposed volume much larger than the beam diameter, as shown here. An electron beam was scanned across a thick resist, as shown, to create the cylindrical volume.

which the spreading can be calculated easily. These calculations show that for a thin resist (so that the backscattering occurs in the substrate) on silicon, the current I_0 in a small-diameter beam should be in the range

$$S\left(\frac{\pi}{4}\right) d_b v_s \leq I_0 \leq 2S_t R v_s$$

where S and S_t represent the electron charge density necessary for full exposure and the threshold value, respectively, d_b is the beam diameter, v_s is the electron-beam linear scan velocity, and R is the electron range. This inequality indicates that the current must be great enough to expose the resist but less than the value at which transverse spreading occurs.* Under such conditions electron-beam-exposed resist materials have been used to create lines as narrow as 0.1 μm.

The spot current density J_s is related, of course, to the total current in the beam and the spot diameter at the target. The diameter of an undeflected spot is limited by three principal effects: the spherical and chromatic aberrations of the focusing lenses and the transverse thermal velocity of the particles. Each of these effects can be considered to give rise to a minimum value of beam diameter, usually called the disk of least confusion. These effects are treated extensively in the literature on electron optics (see, for example, Ref. 27, chap 7). Briefly, spherical aberration results from the fact that the focusing fields of a lens are almost always stronger at the outside of the beam than on the axis; electrons entering the lens at different radiuses are therefore directed toward a focus at different points along the axis. Chromatic aberration refers to the sensitivity of the lens to particle energy; since the electrons always have a spread in energy, due to such factors as thermal velocity of emission, power supply ripple, etc., this effect will also defocus the beam.

In addition to beam-distorting effects caused by the electric or magnetic fields used to focus or deflect beams, the beams themselves contain, in a sense, the cause of two other forms of distortion. The first of these is the beam expansion resulting from the beam space charges.

*For example, for a 500-Å 10-kV beam onto PMM resist (R = 2.7 μm, $S = 5 \times 10^{-5}$ C/cm², $S_t = 1.7 \times 10^{-6}$ C/cm², v_s = 0.03 cm/s), 2.4×10^{-11} A $\leq I_0 \leq 2.8 \times 10^{-9}$ A. If $I_0 < 2S_t R v_s$ ($< 2.8 \times 10^{-9}$ A in this example), electron scattering will be minimized or eliminated altogether.

However, this space-charge-expansion effect is usually not the limiting factor in determining spot size, since the current values are too low and the beam voltage too high in the types of beams of interest here.

A second property of the charged particles in a beam that usually causes a first-order effect on both spot size and current density is the *transverse* random component of velocity due to the emission of the particles from a hot surface. Electrons or ions emitted from a solid surface or from a plasma sheath at temperature T_c in general will possess a Maxwellian velocity distribution $[f(v_r) = k \exp(-mv_r^2/2kT_c)]$. In passing through the subsequent acceleration and focusing systems, this transverse velocity distribution is transformed into a Gaussian distribution of current density in the focused spot. If the convergence angle of the beam toward this focus is denoted by θ, it can be shown[28-30] that the maximum possible current density J_m at this spot is related to the emitter current density J_c by the following equation, which represents the Langmuir limit on current density:

$$J_m \approx J_c \frac{11\,600 V_0 \theta^2}{T_c} = \frac{4}{\pi} \frac{I_0}{d_g^{\,2}}$$

where d_g is the "Gaussian" spot diameter; that is, the diameter if the spot is limited by this transverse thermal velocity effect.

The equations relating the diameter to the beam and lens parameters attributable to these three effects are

$$d_s = \frac{1}{2} C_s \theta^3 \qquad \text{(spherical aberration)}$$

$$d_c = C_c \frac{\Delta V}{V_0} \theta \qquad \text{(chromatic aberration)}$$

$$d_d = \frac{7.4}{\sqrt{V_0}\theta} \text{ (Å)} \qquad \text{(diffraction)}$$

$$d_g = \frac{1}{\theta} \frac{I_0}{\sqrt{\frac{\pi}{4} F J_c \left(\frac{11\,600}{T_c}\right) V_0}} \qquad \text{(transverse thermal velocities)}$$

where C_s and C_c are the spherical and chromatic aberration coefficients, respectively; θ is the half-angle of convergence of the beam at the target; I_0 and V_0 are the beam

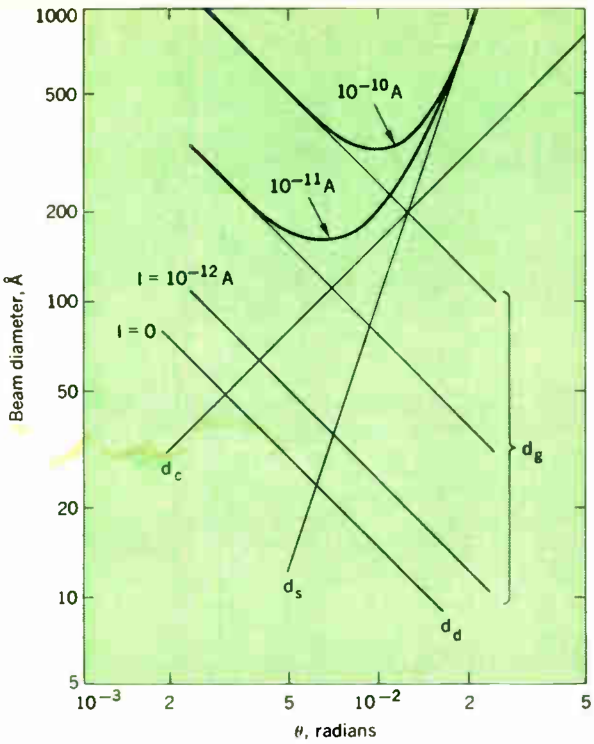


FIGURE 8. Calculated spot diameter versus convergence angle for an electron beam in a SEM. These curves illustrate the importance of the several types of aberrations present in beams and lenses. C_s = 2 cm; C_c = 1.6 cm; J_c = 5 A/cm; $\Delta V/V = 10^{-4}$; V = 10 kV.

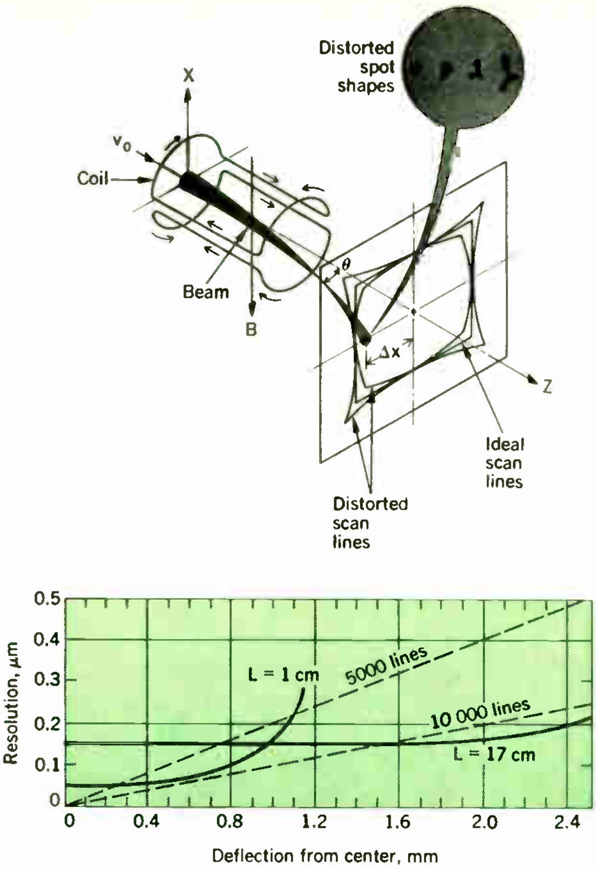


FIGURE 9. The several forms of distortion of the spot and the pattern as a result of beam deflection. Curves show the change in spot diameter with deflection distance for two values of the lens–target distance L.

current and voltage at the target; and $e\Delta V$ represents the spread of initial energy of the electrons from the source, where e is the electron charge. The term d_d refers to the minimum spot diameter as limited by diffraction in the aperture. F is a factor determined by the definition of the diameter of a beam with Gaussian density distribution; if 80 percent of the current is to be enclosed within a circle of diameter d_g, $F = 0.62$.

Since the independent parameter in these equations is θ, it is useful to plot the total beam diameter (the rms sum of these four quantities) versus θ; see Fig. 8. The values of the parameters used in the calculation of the curves are typical of modern commercial scanning electron microscopes.[31] It is seen that the principal effects influencing spot size are the transverse thermal velocities—d_g, dependent on beam current—and spherical aberration d_s. The effects of chromatic aberration d_c and diffraction d_d are relatively insignificant. This graph (and the inset in Fig. 10) show that electron-beam diameters less than 500 Å can be achieved for useful ($\sim10^{-9}$–10^{-10} ampere) currents; thus beam creation is not the principal resolution limit.

Limitations on electron-beam pattern area

Although electron beams can produce fine patterns, with line width and edge resolution considerably finer than possible with photolithography, the area of pattern that can be written is limited, and is related, as in the photographic process, to the resolution. There are two principal limitations on the area over which an electron beam can be scanned.

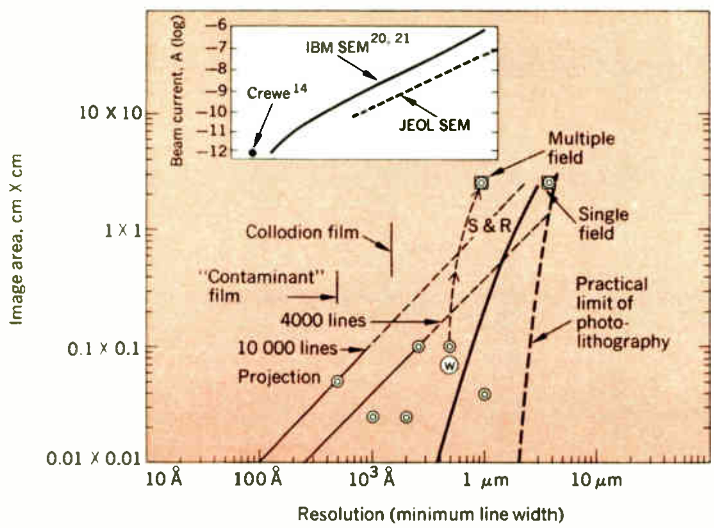
One limit on pattern area lies in the electron optical distortion suffered by a beam of electrons in passing through the complex electric or magnetic fields used to deflect them. That is, if a limit is placed on maximum allowable spot distortion, the scan distance is effectively limited. It is well known that the fields used to deflect an electron or ion beam will cause a distortion of an initially round beam into a noncircular spot. Also, as electrons are deflected through different angles they will be brought to a focus at different distances from the gun; thus, the spot size seen on the screen will change with deflection angle. Further, if deflecting current is applied so as to create a rectangular pattern on the screen, this pattern can appear distorted. These effects are illustrated in Fig. 9, which also shows experimental data[32] on an electron beam in a scanning electron microscope with distortion correction. It is seen that a distance of roughly 1 mm each side of center, depending on the working distance L, can be scanned before the beam diameter doubles (an extreme case). For a 0.05-μm spot, this corresponds to total deflection of about 10 000 spot diameters. For a longer working distance (lens to target), a larger (0.15-μm) beam was deflected 10 000 spot diameters with very little distortion. It appears therefore that with a high-quality electron optical system a deflection distance of 10^4 spot diameters can be achieved. For reference, the deflection of a typical television raster is 500 lines.

The second limit on pattern area results from the finite resolution of the digital-to-analog converter. The digital signals from the controlling computer must be converted into analog form in order to drive the beam-deflection amplifier. However, this D/A converter can produce only a certain finite number of discrete voltage levels; for example, 4096 for a 12-bit converter or 1000 for a 12-bit binary-coded decimal converter. The linearity is usually limited to approximately 0.01 per cent of full scale. Fifteen-bit D/A converters are available commercially, so it is reasonable to expect a scan field limit of about 10^4 lines. Therefore, the actual usable spot diameter, or bit density, per line scan will be limited either by the deflection aberrations discussed in the foregoing or by the electronic linearity of the D/A converters.

Area versus resolution

Let us combine the resolution and pattern-area limits on programmed electron beams to see what kinds of patterns can be created. Naturally, the IC manufacturer would hope that a new technology could create patterns of 0.5-μm lines or less, with almost perfect registration, over a 5-cm (or even 7-cm) slice in a few seconds' time! The actual state of the art in total outside pattern area is shown in Fig. 10 plotted versus resolution, defined here as the minimum line width, with several processes included for comparison. The heavy solid line shows the area-resolution limit of photolithography due to diffraction and lens aberration.[38] Because of practical tolerances such as registration, etching, undercut, and diffraction from the optical mask, the practical optical limit is given approximately by the heavy dashed line. Several points showing the area and resolution that have been attained in laboratory samples of devices fabricated by electron beams are shown by the circles.[18,19] In one case, marked *W*, the basic pattern was created by a scanned electron beam; it was then reproduced by a step-and-repeat process to result in a larger total pattern, represented by the left square point, with only slightly degraded resolution.[34] The right square point represents a single field pattern drawn by an electron beam.[35]

FIGURE 10. Image area versus resolution for electron-beam-exposed patterns (circles), compared with the limits of photolithography. The upper left graph shows, to the same abscissa scale, the electron-beam current versus spot diameter for three electron-beam systems. (JEOL = Japan Electron Optics Laboratory brochure.)



It is seen from Fig. 10 that electron-beam patterns have been created with resolution that is better by about an order of magnitude than that of photolithography. It should be noted that the circles represent typical devices that have not been made with resolution near the ultimate limit of the electron-beam process. The resolution limits of collodion[36] and hydrocarbon "contaminant" films[36] are shown by the vertical lines so marked. The number of lines representing the ratio of pattern height to line width is indicated by the light diagonal lines. It was shown in the foregoing that the number of lines over which the beam can be deflected is limited, in the sequential digital scan technique, by the resolution of the D/A converter and by deflection distortion of the beam; 10 000 lines should be possible in the near future. Thus, the area of greatest applicability of electron beams for writing high-resolution single-field patterns is given roughly by the triangle with its left side on the 10 000-line line and right side at the heavy dashed curve. Larger areas, for a given resolution, can be exposed by a precision-controlled stepwise movement of the target and repeat of the electron-beam-formed pattern.

The direction of further technology development is represented by movement upward (larger area) and to the left (better resolution) from the existing circled points. Digital-to-analog converters exhibiting better performance than indicated plus improved deflection systems should make this possible.

The graph in the upper left of Fig. 10, which is plotted to the same abscissa scale as the larger graph, shows the beam current that can be focused into a given spot in typical SEMs. As higher resolution is sought, lower current in the electron beam is available to carry out the fabrication task.

It is clear from Fig. 10 that electron beams can serve a very useful fabrication function by providing a means of extending the resolution range available to microelectronics processes. Electron-beam techniques can be used to form two classes of patterns. First, they can create a large-area (slice-size) mask for direct optical exposure of photoresist. Such a mask made by multiple field exposure would exhibit a resolution perhaps a factor of two or three better than one made photographically—a significant gain from the point of view of integrated circuits. The second class of pattern lies at the left side of this "accessible" triangle in Fig. 10. Such patterns represent structures for which the ultimate resolution is required but which are of limited area. An example is the interdigital surface acoustic wave transducer, in which the upper limit on operating frequency is directly dependent on pattern resolution. In this class of device the increased performance must justify the probable requirement of individual exposure of each pattern by the electron beam.

Beam process time

The economics of a beam process is obviously an important consideration in any comparative evaluation. The cost of fabricating a pattern or device will be determined partly by the time required to write the desired pattern with the electron beam; the rest of the necessary process time will be required by vacuum pumpdown, resist applications and development, metalization, etc. The writing time should be short enough that it does not form a limitation on the economics of fabricating complex, high-resolution patterns. Therefore, it is worthwhile to examine the factors affecting pattern-writing time.

When an electron (or ion) beam strikes a target, it deposits a charge, given by the current density multiplied by the time of exposure. The useful effect of a beam on a target surface can be expressed in terms of the number of electrons or ions deposited per second per unit area necessary to produce the desired effect, and thus each target material, such as a resist, is characterized by a value of this "sensitivity" factor S, measured in coulombs/cm². The time necessary to expose a spot of diameter D_s with a beam providing a current density J_s at the surface is then simply $\tau_s = S/J_s$. For example, in order to expose an electron resist material that exhibits a value of $S = 8 \times 10^{-5}$ C/cm², approximately 5×10^{14} electrons/cm² are required to break the chemical bonds of the polymer so that it can be developed to the desired degree of contrast. The smaller the value of S, the more sensitive the process; that is, fewer charged particles are required to effect the desired surface interaction. The value of S varies widely for different beam processes; some representative numbers are given in Table II. It should be borne in mind that the resist values are quite dependent on the development process. The exposure time for PMM resist will be $\sim 6 \times 10^{-5}/J_s$ seconds per spot. The electron-beam current-density value used should be calculated on the basis of the exposed area of the resist; that is, including the scattering factor discussed previously. Typically, $J_s \approx 0.1$–1 A/cm² using this definition; thus, $\tau_s \approx 60$–600 μs per spot. The total exposure time for a given area composed of N spots is just $N\tau_s$. In order to attain 0.1-μm edge resolution with a 1-μm line, about ten spots are required per line width, or 10^2 spots per square; roughly 6×10^3 such squares are needed to form a typical surface wave transducer as shown in Fig. 5. Thus, complete exposure requires 6×10^5 spots. If $\tau_s \approx 60$ μs, the exposure time will be about 36 seconds. As electron-beam densities as high as 10^2 A/cm² have been focused into a spot with existing equipment, shorter exposure times can be attained if the scattering factor is low.

The current density from the photocathode of the parallel-exposure system [Fig. 4(B)] is from 10^{-4} to 10^{-5} A/cm². Thus, in an ideal situation, the parallel-exposure method will be faster than the sequential method, with a spot current density of 1 A/cm², when the pattern involves more than 10^4–10^5 spots. The wafer exposure time with parallel exposure is quoted[15] as 20 seconds or less, so the target interchange and pumpdown times will be determined by the manufacturing speed limitations.

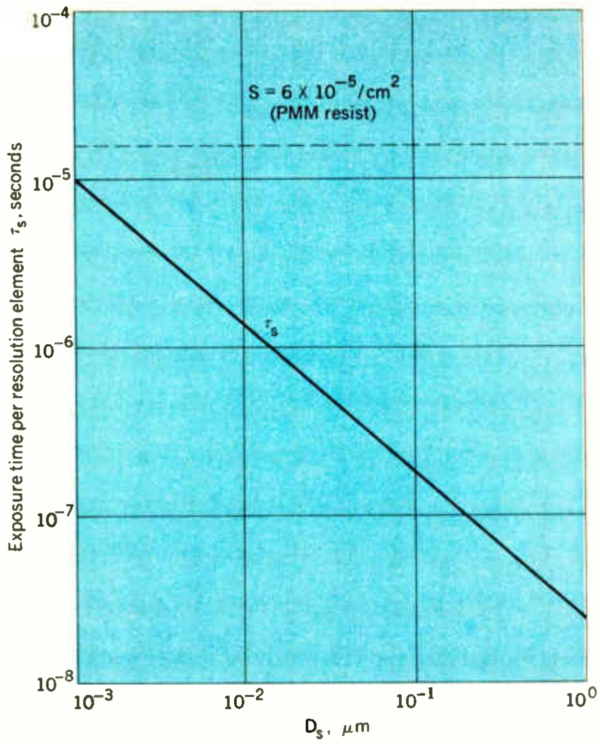


FIGURE 11. Tradeoff between spot exposure time and resolution. Dashed line shows the limit due to computer calculation and address time.

Due to limitations on the emission density obtainable from a cathode and to electron-optical aberrations, the current density in a focused spot generally decreases as the spot is focused to a smaller diameter.[37] By using the best SEM beam performance attained to date (see inset curves in Fig. 10), a relation between J_s and D_s can be used together with the spot-writing-time equation to find τ_s as a function of D_s. Such a plot is shown by the solid line in Fig. 11. Since the best beam performance has been used in plotting it, and beam spreading in the resist was not taken into account, this curve represents a lower limit on spot exposure time (for $S = 6 \times 10^{-5}$ C/cm²).

Perhaps of even more importance than the exposure time is the writing-speed limit placed by the address and settling times discussed earlier for the sequential digital scan mode. The sum of these two times per spot is usually about 10–15 μs; this level of time is shown in Fig. 11 as the horizontal dashed line. It is clear that in this mode we will not gain in writing time per spot by increasing the current density. With the scan technique shown in Fig. 4(A), τ_s cannot be reduced below the dashed line; this is a serious limitation imposed by the computer and the D/A converter. It should be noted that if the exposure time τ_s per spot is less than the computer time, the 12-μs limit shown here is not even valid, because there may not be time enough to calculate and address for the next spot. This situation shows the need for advanced beam programming techniques that do not require each spot to be calculated and addressed, but still retain the advantages of the digital computer for spot programming. A way out of this limitation may lie in

II. Variation in sensitivity for different beam processes

Process	Particle	S (C/cm²)
Ion-beam machining (0.2 μm of SiO_2)	ion	0.1
PMM (positive) resist	electron	6×10^{-6}
Implantation (dose = 10^{14})	ion	1.6×10^{-5}
Negative resist (e.g., KTFR)	electron	5×10^{-6}
Silver halide film	electron	$\sim 10^{-8}$

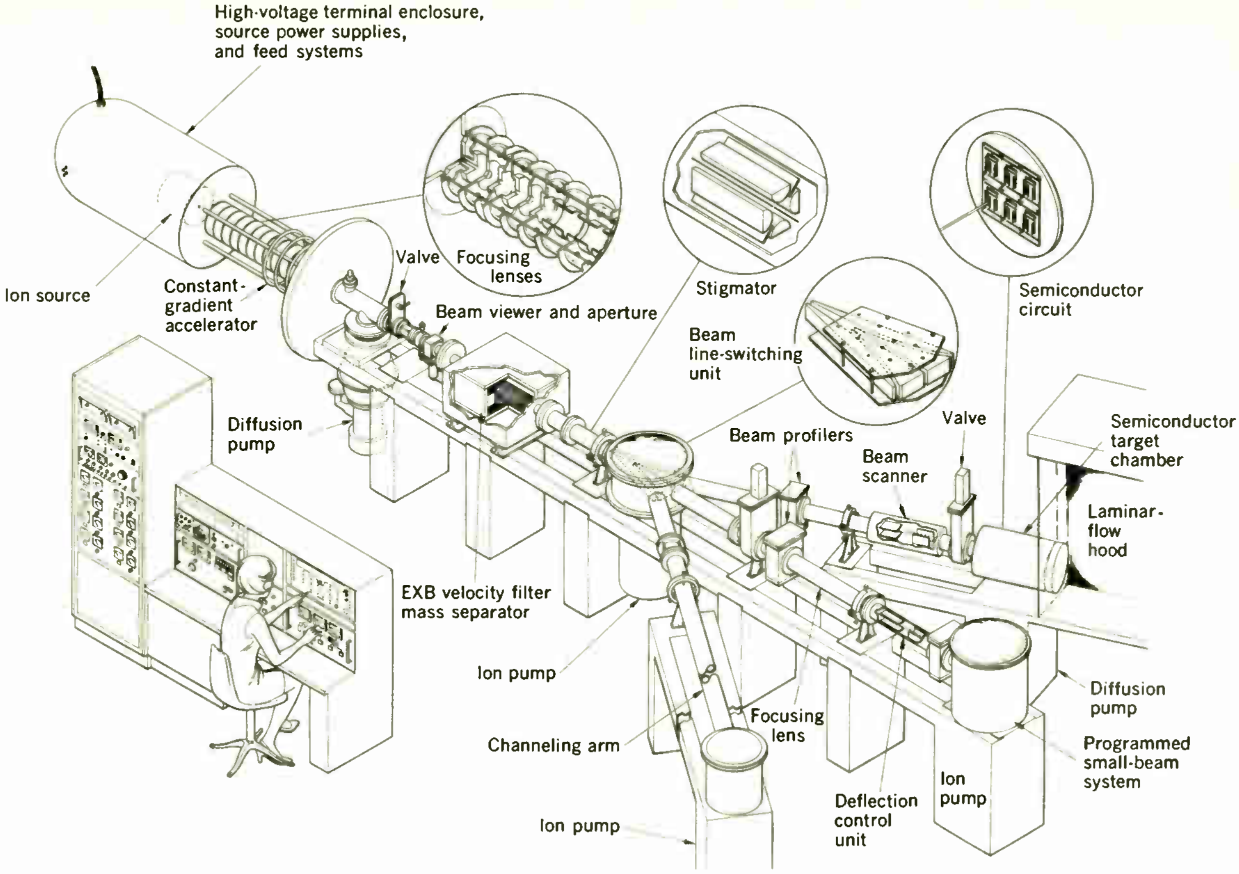


FIGURE 15. Research-type 300-kV ion implantation system. Design permits beam to be switched into one of three target arms for channeling, tiny beam focusing studies, and normal implantation respectively.

FIGURE 16. Ion implantation system for carrying out the implantation step in the manufacture of MOSFET integrated circuits. The terminal housing, including ion source, is shown at the top and the rotating multiple-target chamber at bottom right. (Figures 15 and 16 courtesy Hughes Aircraft Co.)

laboratory ion implantation system that provides a very broad range of implantation parameters and functions. Capital requirement to construct this system was about \$120 000. Finally, Fig. 16 presents an ion implantation system custom-made for the batch production of *p*-channel MOSFET integrated circuits.* The capital cost of this system was about \$45 000.

Conclusions

Electron and ion beams provide a qualitatively different process technique for fabricating microelectronic devices and circuits. Ion implantation beams are currently the more advanced in application. In fact, this doping technique is being used as a manufacturing process step for MOSFET integrated circuits, where it is compatible with and complements contemporary planar processes.

Electron beams offer a means for creating higher-resolution patterns, a capability that can be applied to a wide variety of microelectronic elements. Multiple-field parallel-exposure methods can provide large-area patterns with somewhat better resolution than photolithography and with economical process times. Higher resolution over small areas will provide a means for fabricating those types of devices requiring this resolution, for which the performance gain is worth the slower process speed of single-field exposure. However, although high-resolution electron-beam-pattern fabrication has been demonstrated in laboratory prototype devices, much development

* Developed partially under Air Force Contract No. F33615-70-C-1238.

remains to convert it to a manufacturing process.

Both types of beam processes offer some unique features; some very useful devices just cannot be made any other way! The potential yield improvement by all-vacuum processing, using electron and ion beams, together with automated electron-beam testing, present goals worth working toward. Beams are presently not widely used or fully developed, and involve relatively sophisticated processes involving generally unfamiliar and expensive equipment. Nevertheless, in the writer's opinion, they form the next major process technology for microelectronics. In some cases beam processes may replace contemporary fabrication processes; more othen they will be applied in ways that take best advantage of their unique features.

This article was presented in part as an invited paper at the IEEE Conference on Electron Device Techniques, New York, N.Y., Sept. 22, 1970.

The author expresses appreciation to several of his colleagues, especially Dr. E. D. Wolf, for considerable help in the preparation of the article.

REFERENCES

1. Johnson, H., "The anatomy of integrated circuit technology," *IEEE Spectrum*, vol. 7, pp. 56–66, Feb. 1970.

2. Hochman, H. T., and Hogan, D. L., "Technological advances in large-scale integration," *IEEE Spectrum*, vol. 7, pp. 50–58, May 1970.

3. Suran, J. J., "A perspective on integrated circuits," *IEEE Spectrum*, vol. 7, pp. 67–79, Jan. 1970.

4. Petritz, R. L., "Current status of large-scale integration technology," *IEEE J. Solid State Circuits*, vol. SC-2, pp. 130–147, Dec. 1967.

5. McCaldin, J. O., *Nucl. Instr. Methods*, vol. 38, p. 153, 1965.

6. Mayer, J. W., and Marsh, O. J., *Applied Solid State Science, Vol. 1.* New York: Academic, 1969, pp. 239–337.

7. Gibbons, J. F., "Ion implantation in semiconductors—Part I. Range distribution theory and experiments," *Proc. IEEE*, vol. 56, pp. 295–319, Mar. 1968.

8. MacDougall, J., Manchester, K., and Palmer, R. B., "Ion implantation offers a bagful of benefits for MOS," *Electronics*, vol. 43, pp. 86–90, June 22, 1970.

9. Bower, R. W., *et al.*, "MOS field effect transistors formed by gate masked ion implantation," *IEEE Trans. Electron Devices*, vol. ED-15, pp. 757–761, Oct. 1968.

10. Kramer, N. B., Hughes Aircraft Co., Private communication, 1970.

11. Wauk, M. T., "Microwave acoustic surface wave amplifier," presented at 1970 IEEE Internat'l Electron Devices Conf., Washington, D.C.

12. MacDougall, J. D., Manchester, K. E., and Koughan, P. E., "High value implanted resistors for microcircuits," *Proc. IEEE*, vol. 57, pp. 1538–1542, Sept. 1969.

13. Dill, H. G., *et al.*, paper presented at Internat'l Conf. on Ion Implantation in Semiconductors, Physics, and Technology, Thousand Oaks, Calif., May 1970.

14. Crewe, A. V., "High resolution scanning electron microscopy," *Proc. 2nd Annual Scanning Electron Microscope Symp.*, Apr. 1969.

15. Koops, H., Mollenstedt, G., and Speidel, R., *Optik*, vol. 28, p. 518, 1969.

16. O'Keeffe, T. W., "Fabrication of integrated circuits using the electron image projection system (ELIPS)," *IEEE Trans. Electron Devices*, vol. ED-17, pp. 465–469, June 1970.

17. O'Keeffe, T. O., Vine, J., and Handy, R. M., "An electron imaging system for the fabrication of integrated circuits," *Solid-State Electron.*, vol. 12, p. 841, Nov. 1969.

18. Heynick, L. N., paper presented at 1970 IEEE Conf. on Electron Device Techniques, New York, N.Y.

19. Wolf, E. D., Perkins, W. E., and Coane, P. J., *Proc. 3rd Annual Stereoscan Colloquium*, Chicago, Ill., Apr. 1970.

20. Broers, A., notes for U.C.L.A. short course, "Ion and electron beams in microelectronics," Jan. 1970.

21. Broers, A. N., Lean, E. G., and Hatzakis, M., "1.75 GHz acoustic-surface-wave transducer fabricated by an electron beam," *Appl. Phys. Letters*, vol. 15, pp. 98–101, Aug. 1969.

22. Wolf, E. D., Bauer, L. O., Bower, R. W., Garvin, H. L., and Buckey, C. R., "Electron- and ion-beam fabricated microwave switch," *IEEE Trans. Electron Devices*, vol. ED-17, pp. 446–449, June 1970.

23. Wolf, E. D., Hughes Aircraft Co., Private communication, 1970.

24. Everhart, T. E., *J. Appl. Phys.*, vol. 31, p. 1483, Aug. 1960.

25. Archard, G. D., *J. Appl. Phys.*, vol. 32, p. 1505, Aug. 1961.

26. Schiller, S., and Neumann, M., Paper no. 237, Electrochemical Society Meeting, Los Angeles, Calif., May 1970.

27. Grivet, P., *Electron Optics.* New York: Pergamon, 1965, p. 424.

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

29. Moss, H., *Narrow Angle Electron Guns and Cathode Ray Tubes.* New York: Academic, 1968, p. 88.

30. Moss, H., *J. Electron. Control*, vol. 6, p. 403, 1959.

31. Everhart, T. E., Pease, R. F. W., and Pederson, S. R., "The Berkeley scanning electron microscope I," Rept. ERL-66-11, University of California, May 1966.

32. Chang, T. H. P., and Stewart, A. D. G., paper presented at 10th Symp. on Electron, Ion and Laser Beam Technology, Washington, D. C., 1969.

33. Petritz, R. L., "Technological foundations and future directions of large-scale integrated electronics," *Proc. 1966 Fall Joint Computer Conf.*, pp. 65–87.

34. Angello, S. J., "Should microelectronic processing be done with electron beams?" Minutes, 1969 WESCON, San Francisco, Calif.

35. Maekawa, A., *et al.*, Paper no. 254, Electrochemical Society Meeting, Los Angeles, Calif., May 1970.

36. Lenz, F., Private communication, 1970.

37. Brewer, G. R., *Proc. 14th Internat'l Conf. on Electron and Ion Beam Science and Technology*, May 1970 (note Fig. 8).

38. Lean, E. G., and Broers, A. N., paper presented at Ultrasonics Symposium, St. Louis, Mo., Sept. 1969.

39. Phillips, C. D., *et al.*, "Complex monolithic arrays: some aspects of design and fabrication," *IEEE J. Solid State Circuits*, vol. SC-2, pp. 156–172, Dec. 1967.

40. Dill, H. G., and Trerise, B., Hughes Aircraft Co., Private communication, 1970.

41. Moore, E. G., "What level of LSI is best for you?" *Electronics*, vol. 43, pp. 126–130, Feb. 1970.

42. Murphy, B. T., "Cost–size optima of monolithic integrated circuits," *Proc. IEEE*, vol. 52, pp. 1537–1545, Dec. 1964.

43. Septier, A., ed., *Focusing of Charged Particles, Vol. II.* New York: Academic, 1967, chap. 3.4.

44. Garvin, H. L., *Proc. 3rd Symp. on Deposition of Thin Films by Sputtering*, University of Rochester, pp. 4–11, Sept. 1969.

45. Septier, A., *Advances in Electron and Optical Microscopy, Vol. I.* New York: Academic, 1966, p. 219.

George R. Brewer (F) received the B.S. degree in electrical engineering from the University of Louisville in 1943, and the M.S. degree in electrical engineering, the M.S. degree in physics, and the Ph.D. degree in electrical engineering, all from the University of Michigan, in 1948, 1949, and 1951 respectively. He joined the Hughes Aircraft Company in 1951, and since then has been engaged in research, development, and technical direction in the area of microwave electron tubes, electron guns and beams, and ion propulsion. He was manager of the Ion Propulsion Program at Hughes from its inception in 1959 to 1968. He is now manager of the Ion Physics Department, Hughes Research Laboratories, Malibu, Calif. Prior to joining Hughes, Dr. Brewer was a research associate in the Engineering Research Institute at the University of Michigan (1947–51); from 1944 to 1947 he was associated with the Naval Research Laboratories; and in 1943–44 he was a staff member of M.I.T.'s Radiation Laboratory. Dr. Brewer is a member of Sigma Xi, Phi Kappa Phi, and the AIAA. He has written some 18 papers on electron and ion beams, holds 14 patents on electron optics devices, and is the author of a book entitled *Ion Propulsion*.

Using thyristors and diodes to improve commutation

One intriguing feature of diode-assisted and thyristor-assisted commutation systems is that they lend themselves to an arrangement whereby the commutator is completely separate from the dc machine

J. J. Bates **Royal Military College of Science**

Machines employing thyristor- and diode-assisted commutation (TAC and DAC) offer numerous advantages over conventional dc machines in applications where commutation is a problem. It is anticipated that the first practical developments will involve adaptations of present systems, followed at a later date by the introduction of more unconventional construction techniques.

Way back in the thirties there were suggestions for improving commutation in dc machines by incorporating rectifying devices in the commutating coil circuit. The simple object was that if it were difficult to ensure current zero just as a brush leaves a segment (as it undoubtedly is), then it would be better to aim for a reversal of the current from the brush to the segment, but to block the reversal by the presence of a rectifying device so that the current from the brush to the segment would be zero some time before the brush leaves the segment. There are a number of ways of doing this.

Figure 1 illustrates the simplest technique. The interpole-induced voltage e brings the current to zero in brush b_1 before the brush reaches the end of its segment, and thus the current must remain at zero by action of the reverse-biased diode. Figures 2 and 3 show similar possibilities. The general principle is clearly the division of a commutator (or segmented ring) into a number of "active" and "inactive" zones, with armature tappings only to the active zones and brushes narrower than the inactive zones. The arrangements of Figs. 2 and 3 allow for running in both directions, as also of course could the arrangement of Fig. 1 by the inclusion of a further diode.

The advantage of these schemes is that the reactance voltage is virtually unlimited compared with a conventional machine. In the first place, it is obviously possible to include a "factor of safety" in interpole design, since it is necessary only to calculate the maximum interpole strength to bring the current to zero just as a brush leaves a segment, and then ensure that the design gives a somewhat higher interpole flux than this. Nonsaturation of the interpole no longer has much importance. Second, there is no need to secure a precisely maintained arc of contact between a brush and the surface, and thus small irregularities in profile should be acceptable. Both of these considerations allow the commutation of larger currents at higher speeds and should lead to smaller dc machines in the larger sizes where commutation is a problem and to the possible use of dc machines in the power and speed ranges in which commutation is at present an impossibility.

An even wider implication is the ability to construct dc machines of totally different geometry, and it may be well to dwell on this for a moment as it is a significant point. Figure 4(A) shows the arrangement of a two-pole dc machine assumed to be at its commutation limit; that is, it passes a current and runs at such a speed that it just starts to spark. If more power is wanted than this two-pole machine can provide, there are two possibilities. One is to think of the two-pole machine "split"—that is, opened out into a semicircle, or a quarter circle, as shown in Fig. 4(B), with similar segments arranged alongside

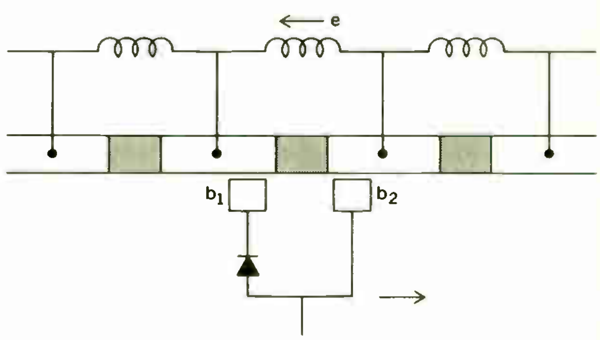
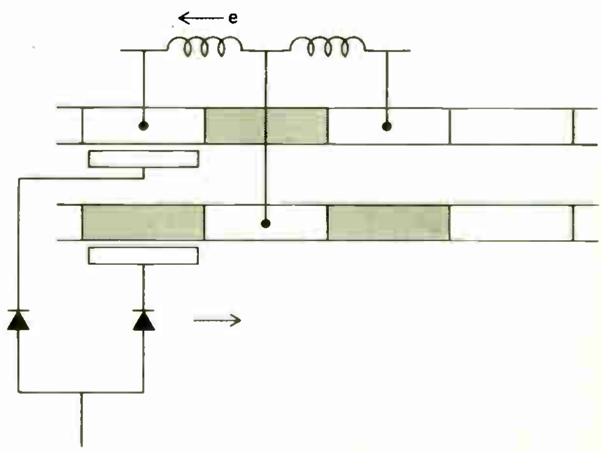


FIGURE 1. Simple single-commutator DAC scheme.

FIGURE 2. Two-commutator DAC scheme.



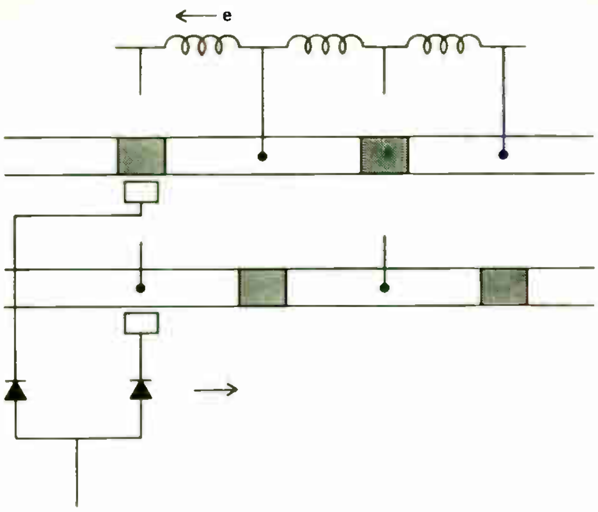


FIGURE 3. Alternative two-commutator DAC system.

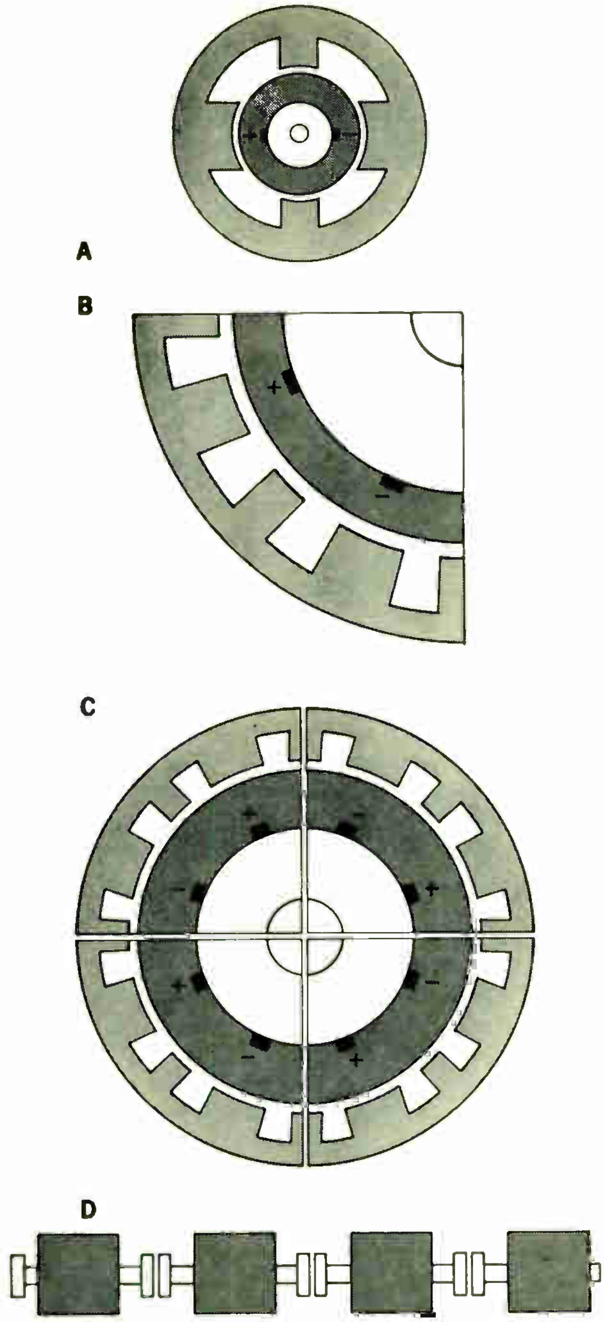


FIGURE 4. Two-pole dc machine. A—Basic arrangement. B—"Split" quarter-circle arrangement. C—Four-quadrant representation of quarter-circle arrangement. D—Mechanical coupling of four two-pole machines.

to complete the circle; see Fig. 4(C). If this new machine is to retain the same commutation limit the peripheral speed must be the same (a quarter as many revolutions per minute as in the two-pole machine), even though as the diameter is increased fourfold the total power will be four times that of the two-pole machine. The only other way of "stacking" the two-pole machine to get four times the power consists simply of four two-pole machines mechanically coupled together; see Fig. 4(D).

The simple conclusion is that high-power dc machines will have large diameters and run at low speeds, and that if high powers at high speeds are wanted, the only possibility is to couple machines together mechanically, with the commutator in each separate machine just coping with the commutation of the coils in that machine. In other words, a long, thin dc machine has to be chopped into lengths, with separate commutators coping with each discrete length. The disadvantages are obvious.

If diodes are used to assist the commutation it is perfectly possible to produce long, thin machines of high power and speed, akin to turboalternators in shape.

An obvious question to ask is why such machines are not commercially available if they have so many attractive possibilities. It is easy to see that when the principles were initially proposed, the thermionic nature of the available rectifying devices, which were of limited current and high voltage drop, might have prevented the introduction of this class of machine. But now that efficient, low-cost semiconductor diodes are readily available it is surprising that machines using diodes to assist commutation have not appeared quickly and in significant numbers, since the dc machine design is little affected.

The reason is rather difficult to find, but seems to lie in the fact that although small prototype machines are very successful, large, high-speed machines using the same principle do seem to suffer from unique contact problems of their own, associated with the presence of alternate active and inactive zones on the commutator.

An immediate reaction to the use of commutators working in such a mode is that unequal wear of the active and inactive sections is likely. But this would occur only after appreciable running and would be unlikely to prevent the introduction of the machines for certain special applications. A more serious immediate block to their application lies in the formation of segment marking, associated with pulse loading, on the brushes.

Some detailed consideration is thus given here to the nature of the sliding contact in such machines, the problems that have been encountered, and the way so far they have been overcome and are continuing to be overcome. Although the electromagnetic design of machines with large commutating sections in the armature (commutating in large lumps instead of little bits) is interesting, no difficulties of an electromagnetic nature have arisen.

Machines using both diodes and thyristors to assist commutation will be described and will be referred to as DAC and TAC machines respectively (diode-assisted commutation and thyristor-assisted commutation).

Problems associated with solid carbon brushes in DAC machines

One of the first effects noticed in DAC machines of any real size was a severe burning at the entering edge of a segment as a brush ran onto it. As a brush moves from an inactive segment to an active segment the brush may well take its support from the inactive segment, with a very uncertain low-pressure contact to the active segment. The number of taps to the armature must be much less than in a conventional machine, since the width of the inactive segments must be greater than the brush width, and if the commutator diameter is not to be excessive the number of taps must be low. This consideration—together with the fact that in any case the ultimate aim is for larger currents at higher speeds—means that a high voltage must be induced in the commutating coil. If the contact to the incoming segment is uncertain, arcs are readily established at the making contact and quickly lead to roughening and unacceptable sparking over an appreciable portion of the segment surface.

Such poor current-making effects have been investigated on the rig shown in Fig. 5. A solid brush moves back and forth relative to two excentric rings, and by varying the resistance R the voltage between the brush and the incoming surface can be varied. The nature of the current transfer is shown in Fig. 6. A brush can be nominally well on a surface before current flows at all, and the flow of current is abrupt and accompanied by arcing. A potential of only a few volts is necessary between the brush and the incoming surface to cause arcing, which occurs at low as well as high speeds. The arcing is severe when the voltage between the brush and surface is more than 15 volts.

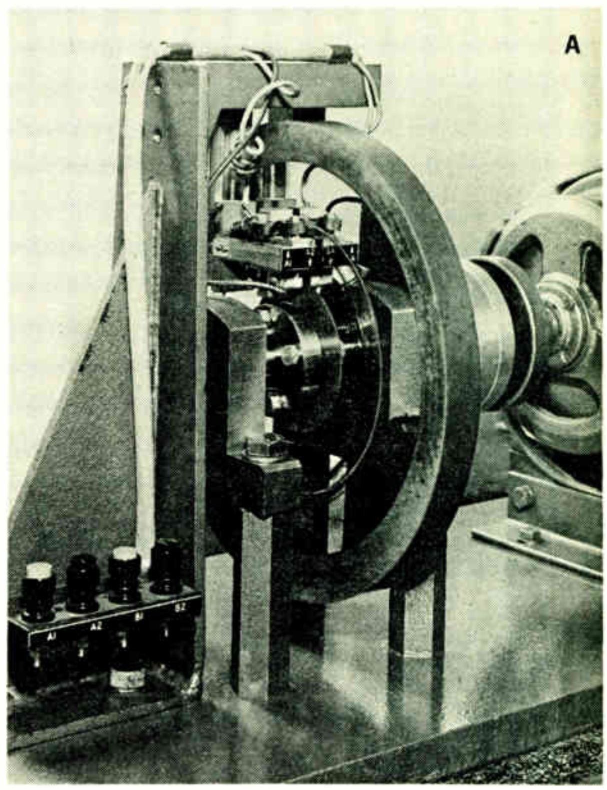
Replacement of diodes by thyristors

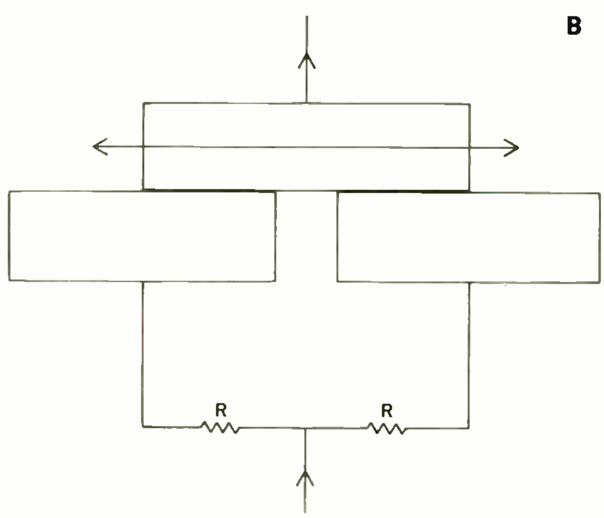
Surface damage at brush entry in DAC machines can be overcome by replacing the diodes by thyristors so that an incoming brush is not allowed to pass current until it is fully on an active segment. In addition, the interpole strength is made sufficient for commutation to be completed while the outgoing brush is still fully in contact with an active segment. This situation leads to the scheme shown in Fig. 7, which should be compared with Fig. 2. The brushes involved are now known as "part brushes." Thyristor T_1 is not triggered until part brush b_{12} is fully on active segment 1, and commutation has to be complete before brush b_{21} reaches the end of active segment 2. This way the surface damage produced by allowing the making contact to be entirely a brush-to-surface action should be avoided.

It must have been with some dismay that an even worse marking of the active segments was observed in large TAC machines. This marking largely took the form of brush imprints on the surface, which tended to occur mainly at the point at which a part brush was asked to carry current by the firing of its associated thyristor, but also occurred in a rather arbitrary way at other points on the surface. Although the surface damage did not take place as quickly as the entering-edge damage in DAC machines, it generally resulted within a matter of hours. The general conclusion was that, if the surface was clean and fresh, TAC was successful; however, after a period of running and the formation of surface films, the contact deteriorated to an extent that it was unable to cope with the shock of a brush suddenly being required to pass current. As soon as damage does occur in TAC machines it rapidly becomes worse. In the arrangement of Fig. 7 this effect is due to some extent to the fact that as brush b_{12} leaves the active segment 1 the current must transfer to brush b_{11} by the action of the volt drop in brush b_{12} only, and if the contact is poor this transfer will not take place satisfactorily. This is equivalent to an interruption of the current supply to the armature as a whole, giving rise to very high induced voltages and very severe arcing.

No grade of solid brush running on a copper surface has been found satisfactory in TAC machines with ratings of a few hundred kilowatts.

FIGURE 5. Current-transfer measuring rig (A) and its schematic representation (B).





Minimum taps per pole pair on TAC machines

In an arrangement such as Fig. 7, the position of the commutating coil at the start of commutation is an overlap of the interpole centerline by an arc equivalent to the brush width (i.e., a brush is just fully on an active segment). If the main pole is always to be outside the commutating zone, then the number of taps obviously affects the permissible main pole width and an expression can be deduced relating the pole-arc: pole-pitch ratio to the number of taps per pole pair and the brush-width: segment-width ratio. This relation indicates that if "conventional" pole-arc: pole-pitch ratios are retained, the minimum number of taps per pole pair is of the order of eight.

It can in fact also be shown that eight is the optimum number of taps; that is, for a given interpole strength (fluxwise) the maximum current can be commutated with eight taps per pole pair.

There are basically three commutator arrangements possible for the condition that a part brush must be fully on a segment during commutation. These are shown in Fig. 7 and in the corresponding arrangements of Figs. 1 and 3. A comparison of these three basic schemes will be made to establish the most advantageous.

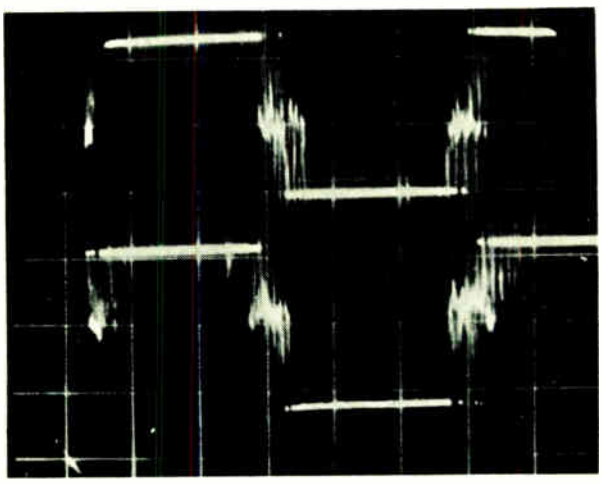
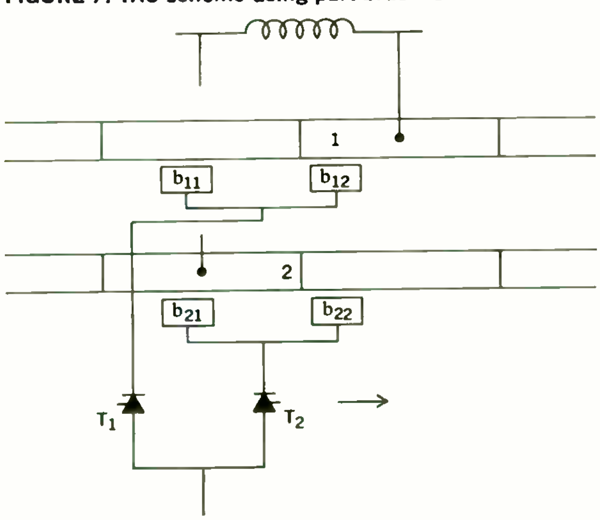


FIGURE 6. Solid-brush transfer properties.

FIGURE 7. TAC scheme using part brushes.



An important factor is the permissible brush-current density. Brushes do not carry current continuously and the limiting factor is the peak density possible. In Fig. 7 the current flows equally, in turn, in brushes $b_{21} \rightarrow b_{12} \rightarrow b_{11} \rightarrow b_{22}$ and experience has shown that the peak density possible with conventional brushes is of the order of 47 A/cm², so the average density of 11.6 A/cm² is comparable to that in a conventional machine. This is not surprising, inasmuch as the current is passing quickly between part brushes and, since it is well known that current only flows through a small part of a solid brush at any one time, it is not surprising that the four part brushes behave as if they were effectively in parallel.

Consider a four-pole machine running at a speed of 3000 r/min. The commutator must then be not more than about 25 cm in diameter, or 80 cm in circumference. Thus for the Fig. 7 arrangement the segment width will be about 5.2 cm. Assuming brushes 1.3 cm wide, the movement available for commutation in a TAC machine, with brushes fully on during commutation, is about 2.6 cm. Now consider the arrangement of Fig. 3. Here there are only two brushes, so each must be 2.6 cm wide for the same total commutator length as Fig. 7 if the average current density is to be the same. The inactive segments must thus be at least 2.6 cm wide and the active segments will be three times as wide. If a brush has to be fully on a segment during commutation, then there is no movement in which commutation can take place. If the 2.6-cm brush is replaced by two 1.3-cm brushes, with commutation starting with one 1.3-cm brush on and finishing when the other 1.3-cm brush is still on, there will be the same 2.6-cm movement possible as for Fig. 7. But the 1.3-cm brushes are difficult to accommodate and individually spring if trailing brush holders are used. Although split brushes with rubber tops are possible, these are not likely to give the same satisfactory contact as two brushes that are independently mounted and sprung.

The scheme of Fig. 7 is thought to be better than that of Fig. 3. But it must be remembered that the latter does avoid the current transfer from a leading part brush to a trailing part brush when both pass over an active segment. The advantage of one over the other may be somewhat marginal, and further investigation is indicated.

Next consider the arrangement of Fig. 1. Here there must be 16 active segments on the single commutator for eight tappings per pole pair. Each inactive segment must be 1.3 cm wide to keep the same density at the brushes as for Fig. 7, with a commutator having an axial length twice that of each commutator of Fig. 7. Therefore, the width of an active segment is limited to 3.9 cm. Study of Fig. 1 will now show that with brushes and inactive segments 1.3 cm wide and active segments 3.9 cm wide there is no commutation movement possible if commutation is to occur while brushes are sitting fully on the active segments. The 1.3-cm brush might be split in two to secure reliable contact, but an additional factor will then enter into the comparison. To provide adequate insulation between segments, in the presence of the high intersegment voltage of these machines, wider gaps than those in a conventional machine must be left between active and inactive segments if both active and inactive segments are of metal. Difficulties then arise with the running of brushes over wide gaps, where a fairly wide brush is necessary for satisfactory performance.

The single-commutator schemes seem to have definite disadvantages relative to double-commutator schemes. In the latter case, if one commutator is mounted at each end of the armature, constructional difficulties do not arise.

Surface damage and carbon-fiber brushes

The possibility of entering-edge damage in DAC machines is predictable and understandable. Although surface damage in TAC machines is much more difficult to explain, it is not entirely unexpected in view of the intermittent current-carrying nature of these units.

A considerable amount of experimental work has gone into elucidating the cause of surface damage in TAC machines. A first clue to a cause came from observations of the change of current in the thyristors and part brushes. These currents are as shown in Fig. 8(A) and, with more detail revealed by high-speed oscillographic techniques, in Fig. 8(B). There are very rapid changes in current at various parts of the commutation cycle, and solid brushes do not seem to have the ability to handle such rapidly increasing current, probably because of the need for time in the adaptation of a solid brush to its contact spots on the surface. So far, the rapid transients are thought to be attributable to capacitive effects in the machine winding resulting from the very high rate of change of an incoming thyristor from its blocking to its conducting state. There is considerable evidence that capacitive transients are a contributory factor—as, for example, by the deliberate addition of small capacitors between the armature tappings, leading to enhanced transients of the same general form as without the capacitors. Moreover, there is a marking of sliding contact surfaces when thyristors are triggered onto brushes through capacitances, this marking being very intense with peak currents as low as 1 ampere but having very short rise times. There is, however, some evidence to suggest that capacitive transients may not be the only cause of surface marking. It does not, for instance, seem possible to prevent the surface damage by the inclusion of inductance in the external part of the commutating loop as a means of reducing the high rate of rise, although admittedly it is difficult to construct a capacitance-free inductance able to pass the full load current of the machine. There is scope for considerable further investigation into the cause of the surface marking.

Although no solid carbon brush has been found to be capable of coping with the rapid current changes, investigations have shown that a brush made up of a bundle of carbon fibers does not give rise to surface marking. The construction of such a brush is shown in Fig. 9. If such a brush is tested in the current-transfer rig described earlier, the current-transfer behavior is remarkably different from a solid brush, as shown in Fig. 10. There is evidence of real reliability of contact and the current-transfer curves indicate a brush-surface resistance that is truly inversely proportional to the contact. The area transfer is done with no arcing, in the presence of much higher voltages between the brush and entering surface.

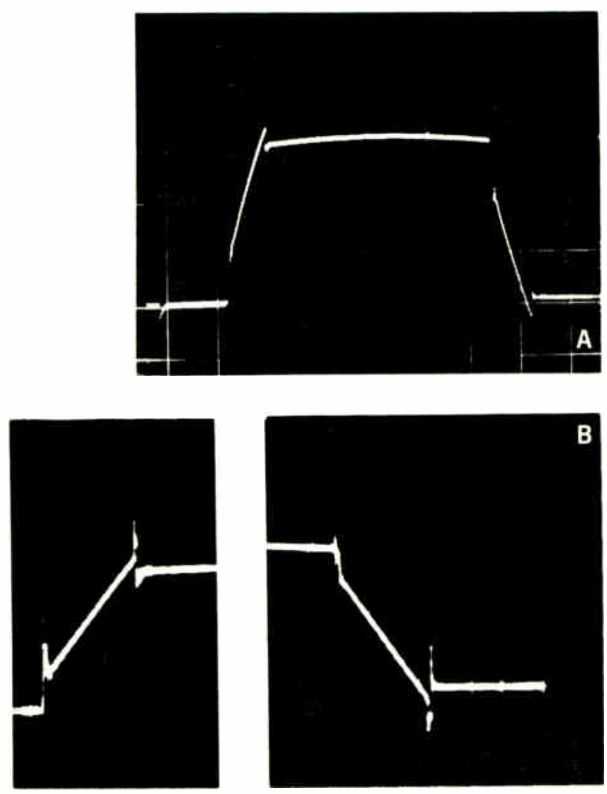
A number of small machines have been built using all-carbon-fiber brushes, and a DAC machine using them is shown in Fig. 11. Note that a commutator construction is used consisting of metal segments inserted in an insulating (resin-bonded laminate) ring, giving a smooth surface on which the soft, flexible fibers run ideally. The permissible current density for the fiber brushes is similar to that for conventional solid brushes, but their voltage drops and friction coefficients are somewhat higher. These small DAC machines have been very successful, since the flexibility of the fiber brushes prevents the entering-edge damage that usually occurs in DAC machines even of this small size.

The possibility that large machines could also work extremely well with all-fiber brushes is currently being explored. There is a limited working length to the brushes—namely, that existing outside the metal box—and they have at present an element of fragility. Once these brushes do spark they can be rapidly destroyed by the concentration of the sparking at individual fibers or bundles of fibers.

One particularly promising way in which carbon-fiber brushes have been applied to DAC and TAC machines is as protective brushes, in parallel with conventional solid brushes. Figure 12(A) shows how this would be done for a TAC machine, the fiber brush being in parallel with a solid brush through a small resistance r. When current tries to pass through the combination and the solid brush contact is unreliable, the current will flow through the fiber brush and build up to such a value that the voltage drop in the resistance r progressively, but not violently, forces the current to flow through the solid brush. Thus the fiber brush may absorb high transients, or otherwise carry currents when the solid brush will not; on the average, however, the current carried by the fiber brush is but a small fraction of the whole.

Figure 12(B) shows typical current waveforms resulting when fiber and solid brushes are used in this way. Following the firing of a thyristor, the current increases through the combination as commutation takes place. The reluctance of the solid brush to carry current in the early

FIGURE 8. Thyristor currents in TAC scheme of Fig. 7.



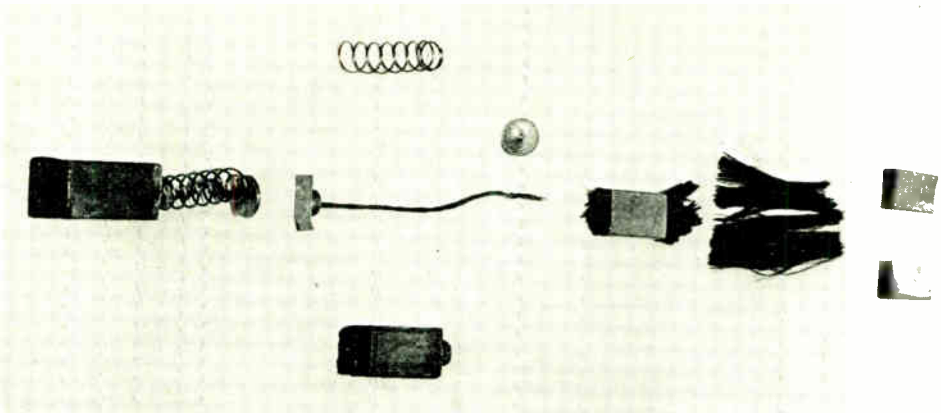


FIGURE 9. Carbon-fiber brush and its construction arrangement.

FIGURE 10. Transfer properties of a carbon-fiber brush.

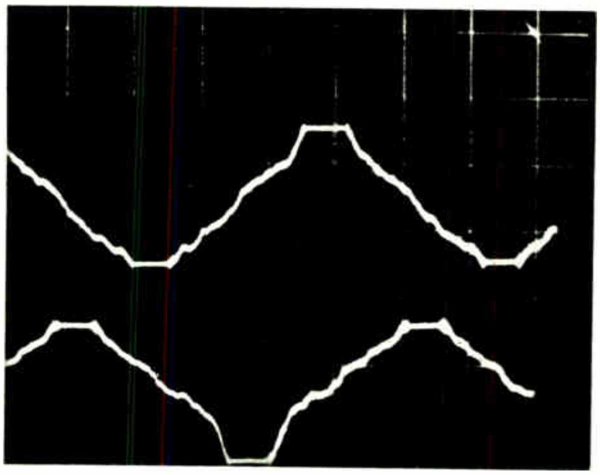
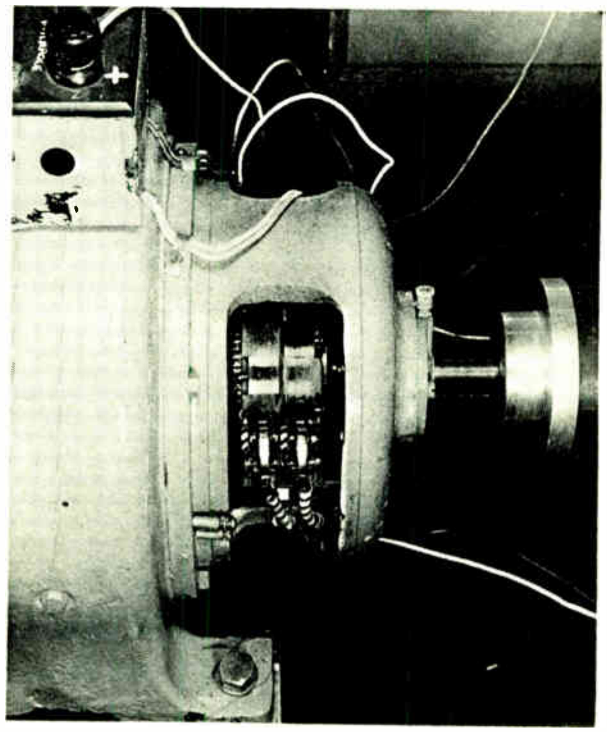


FIGURE 11. DAC machine with fiber brushes.



stage is apparent, since it can be seen that most of the current flows through the fiber brush.

Figures 13(A) and (B) show surface damage on a substantial TAC machine with solid brushes only and with solid brushes protected by fiber brushes. In the latter case, a good skin forms on both active and inactive segments, which are indistinguishable by visual inspection, and no surface damage is apparent after a prolonged period of running. Without the fiber brushes the damage indicated occurred in a few hours.

Fiber brushes can be used for protection in DAC machines, employing the principle shown in Fig. 14(A). In this scheme they lead solid brushes onto the active segments and prevent entering-edge damage by "clamping" the voltage that can exist between the solid brush and the surface to a low value. Again, the average current carried by the fiber brush is small compared with the total current. Figure 14(C) shows the current taken by the fiber brush; the total diode current is shown in Fig. 14(B).

The largest machine using fiber brushes known to be under development at the present time is a TAC machine of some 400-hp rating.

All-thyristor schemes; TAC and DAC starting

Techniques have been proposed that do away with sliding contacts entirely, except for field excitation, by inverting a dc machine and using an all-thyristor system to switch the tapping around the armature. Such an arrangement for a two-pole machine is shown in Fig. 15. Note that 16 thyristors are needed (a four-pole machine would require 32). The fairly obvious procedure involves the formation of a commutating loop by firing adjacent thyristors and commutating this loop by the interpole-induced voltage.

Apart from the economics of any system using so many thyristors, there is one very serious drawback compared with a TAC machine, in that it would not be self-starting as a motor. In other words, commutation fails at low speeds. Figure 16 shows a commutation-time speed curve for a TAC machine with eight taps per pole pair. If the commutating coil had no resistance this curve would be a flat line down to standstill. But with so few sections in the armature winding (and all-thyristor schemes can hardly be otherwise), the resistance of a commutating section can be appreciable. Hence, there is a minimum speed below which the interpole-induced voltage would be insufficient to complete commutation and an outgoing thyristor would not turn off.

In an all-thyristor machine, as in Fig. 15, this means the development of a short circuit as all thyristors successively fire with none turning off. In DAC and TAC machines, however, if the commutation is not completed by the thyristors it must be completed by the brushes, so such machines are self-starting and can be used in motor applications. Here again, fiber brushes have been found to be very advantageous. With solid brushes there will be some sparking at low speeds because of the relatively large inductance of the commutating sections and the sudden-transfer properties of solid brushes. Fiber brushes, with their much superior resistance-switching properties (as shown in Fig. 10), allow the low-speed band to be passed without sparking. This band is quite low, typically 50 r/min in 3000 at full current, and in many cases no auxiliary starting mechanism is necessary in DAC and TAC machines.

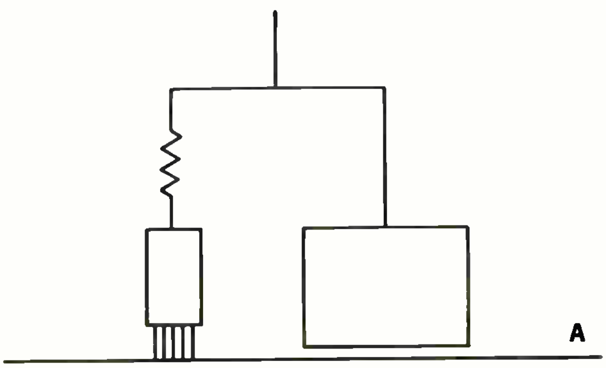
There are ways of starting at low speeds completely sparklessly by introducing commutating voltages into the external part of the commutating loop. Figure 17 shows an arrangement using a commutating transformer, injecting an unsynchronized power-frequency voltage into the commutating loop. It can be shown that, provided the average ac voltage is sufficient to commutate in one half-cycle and the speed is sufficiently low for three half-cycles to be available, commutation by injected unsynchronized ac will always be successful. At speeds where the time available is less than three half-cycles, the interpole-induced voltage is invariably sufficient to take over.

Figure 18 shows a system in which the voltage drop across the small starting resistor is used to charge a capacitor when the leading part brush enters a segment so that when the commutating loop is completed by the firing of the incoming thyristor the outgoing one is turned off by the stored energy in the capacitor. This system is particularly attractive for an all-dc system.

Novel contact systems for DAC and TAC machines

So far, the major use of DAC and TAC machines has been basically in terms of assisting commutation on a fairly conventional type of commutator—admittedly with wide segments, but generally recognizable as a commutator. Also, these commutators occupy the same position, and are of the same size and shape, as in a conventional machine. In the early stages of the development of an

FIGURE 12. Carbon-fiber brushes used in a protective mode. A—Schematic diagram. B—Current waveforms.



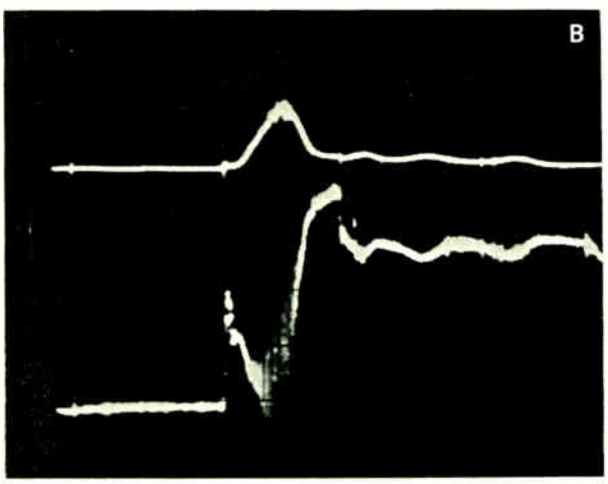
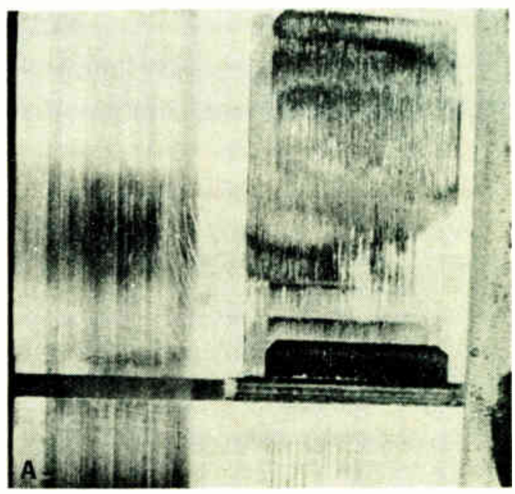
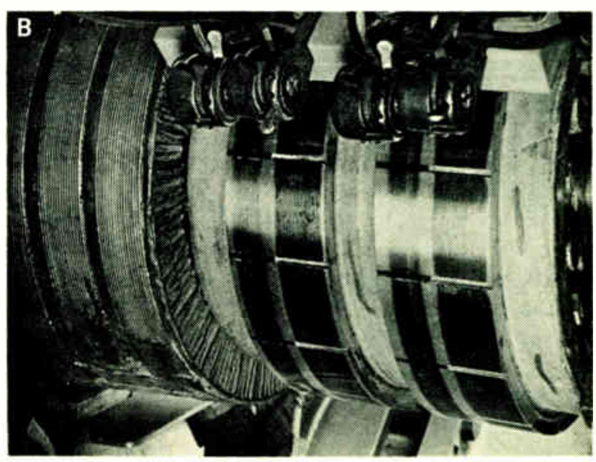


FIGURE 13. Surface conditions with and without the protective action of fiber brushes. A—Solid brush only. B—Solid and fiber brushes.

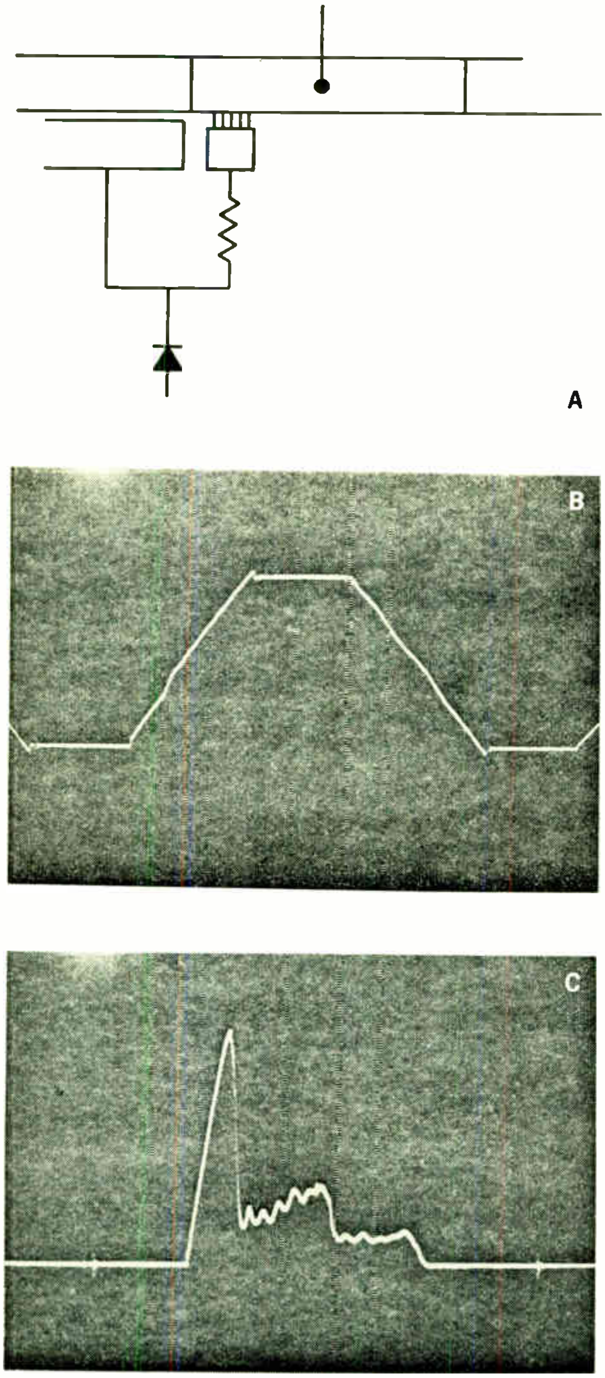




idea this is a desirable feature, since it means very little basic design change in the machine as a whole and is psychologically desirable in introducing a new idea as no revolutionary or expensive change is proposed.

But thinking only in terms of conventional commutators is perhaps shutting one's eyes to the most promising possibilities with DAC and TAC systems. In the TAC systems especially, the role of the commutator is that of a rotating off-load distributor that need not resemble a conventional commutator closely in construction. This concept can lead to very novel and useful results.

FIGURE 14. The use of fiber brushes to lead in carbon brushes in DAC machines. A—Circuit arrangement. B—Diode current. C—Fiber-brush current.



One idea that has been exploited is to separate out the commutator or distributor from the machine. Because few armature taps are involved there is no need for the distributor to be physically close to the armature winding; it can, in fact, be completely separate from the machine, with taps to the armature brought in under the bearings. This arrangement allows the full length of the machine between bearings to be used for torque production; that is, a much greater core length is possible. The output for a given machine weight is thus much increased, if increases in core length are considered together with increases in speed and commutatable current. Also the maintenance, cleanliness, and cooling of this distributor is a totally different matter from that of a conventional commutator. It is removable from the machine without dismantling the machine, and can receive its own clean cooling air remote from the cooling of the machine. Because it is physically separated from the winding it receives little heat from the winding and thus runs cooler.

This principle of physical separation (Fig. 19) can also be applied to a small machine with DAC commutation to provide a motor capable of running immersed in water, by having the distributor at the end of a long shaft, running

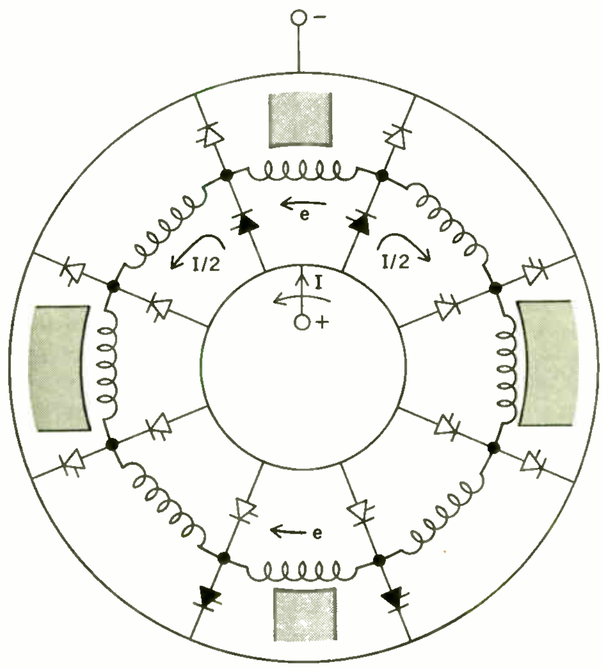
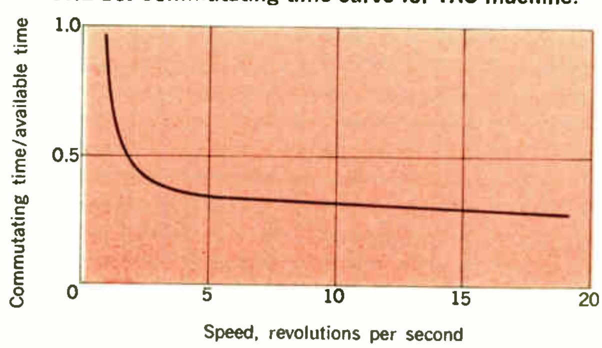


FIGURE 15. All-thyristor inverted dc machine.

FIGURE 16. Commutating-time curve for TAC machine.



in air with the motor itself immersed. This arrangement can be used for portable boat propulsion, the motor running immersed in outboard fashion with the distributor at the top of the shaft sticking out of the water.

Some form of sliding contact is probably essential in high-speed machines. But for low-speed machines (up to a few hundred revolutions per minute) many forms of off-load distributor are possible. Figure 20 shows a rolling-contact system designed for an inverted (armature stationary) dc machine in which current is distributed from a continuous ring to a segmented ring by a pair of rollers,

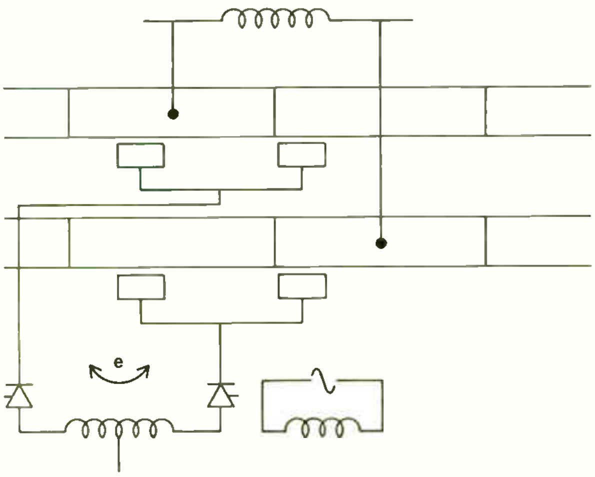
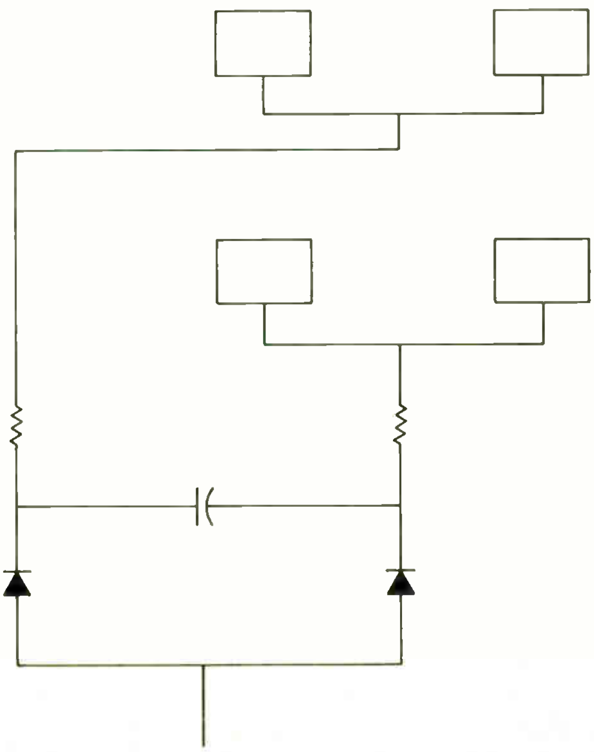


FIGURE 17. Commutating transformer employed for starting TAC machines.

FIGURE 18. Capacitor starting for TAC machines.



with thyristor switching so that the rollers make and break off-load. For such low-speed commutation, interpoles can be omitted and a commutating transformer, as described in Fig. 17, can be used. This rolling-contact system could be applied to steelworks drives to give long, thin dc machines capable of high reversal rates.

Cam-operated distributors are another possibility. In fact, the field is wide open for ingenuity.

Moreover, if few segments are involved on a commutator, why should the commutator construction follow the conventional pattern in which many segments support each other by pressure? With few segments, many alternative and often cheaper constructions are possible.

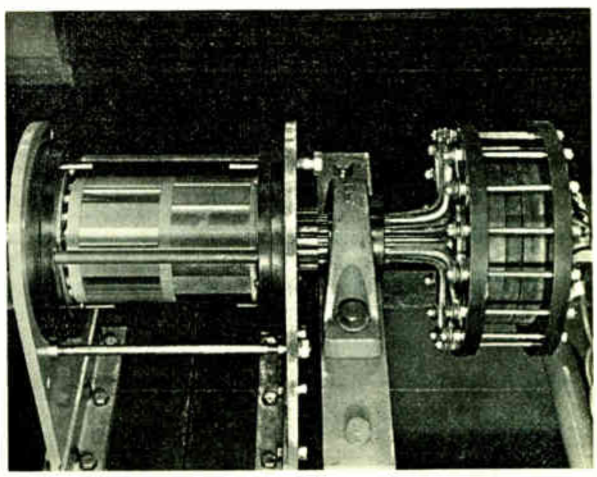
Interpole excitation and some electromagnetic design possibilities

A conventional machine has interpoles in series with the armature, and it is essential for the interpole flux to follow the variations in armature current. This situation often leads to bad commutation under shock-loading conditions, or when the supply is not pure dc, as in a rectifier-fed motor.

In DAC or TAC machines other methods for interpole excitation are possible. If the interpoles are excited by a mixture of separate plus series excitation, so that the series excitation balances the armature ampere-turns, and the flux is then provided by the separate excitation, the interpole flux can always be held at maximum strength no matter what the armature current is, or how rapidly it is changing. This gives the well-proved ability of DAC and TAC machines to absorb shock loads without sparking, and to work extremely well on rectified ac supplies.

Because of the large section of the commutating coil with few armature tappings, a considerable part of the inductance can be largely eliminated and the commutable current still further increased. Since there are few corporated on the interpole face, these components of inductance can be largely eliminated and the commutable current still further increased. Since there are few armature tappings, fewer slots are needed (in small or modest-size machines at any rate) and there is a possible economy here in mass-produced machines that, with the other advantages of DAC systems, may outweigh the extra cost of the diodes.

FIGURE 19. External distributor for TAC machine.



Conclusion

What is the probable future of DAC and TAC systems? First, they are likely to appear as developments of conventional machines that have commutation problems and are wanted for special applications, such as very high speeds, or shock loading use.

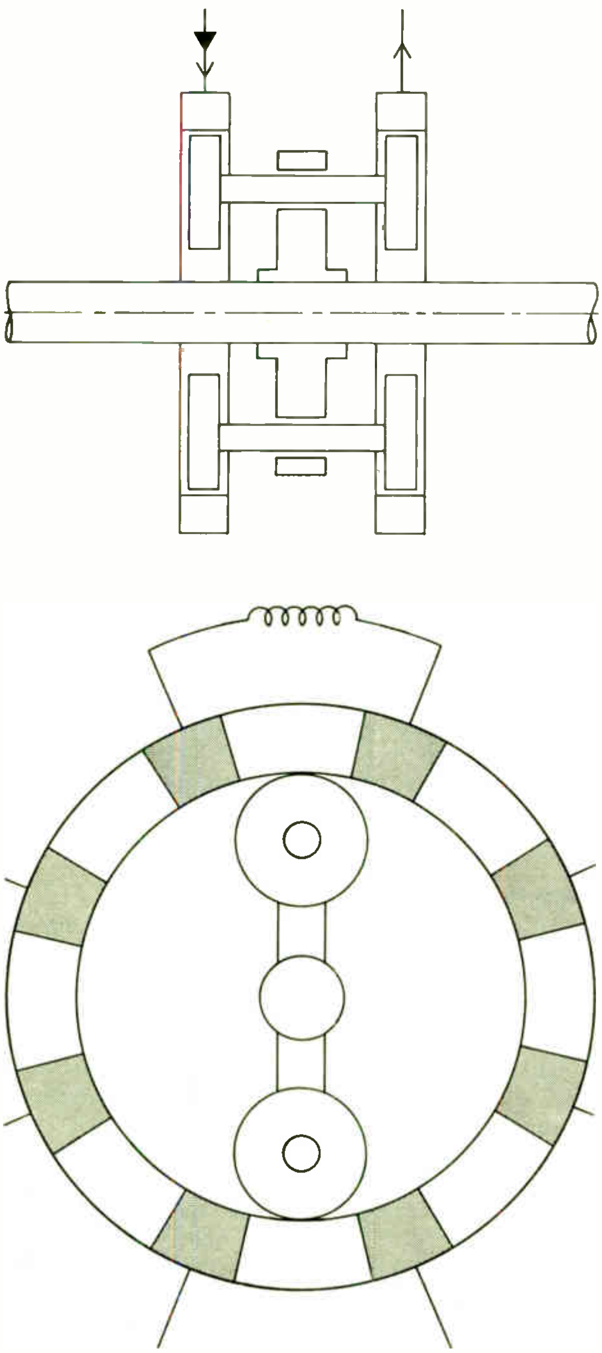
The reliability of the diodes or thyristors is very important. But it must be remembered that they are very conservatively used from a voltage standpoint in these machines. The voltage across a diode or thyristor cannot exceed that in the commutating coil, which is much less than the machine voltage. TAC and DAC machines could thus well lead to the development of high-voltage dc machines with a terminal voltage much greater than the safe working voltage for the diodes and thyristors incorporated.

Unconventional construction techniques are likely to follow on the established success of modified conventional machines. In particular, the movement of the commutator or distributor to the outside of the machine is a probable early development. Perhaps this piece of rotating switchgear, for that is what it is, may be remote from the machine and servo-driven to rotate in step with the machine.

Carbon-fiber brushes are likely to play a significant part by enabling commutators to be constructed of metal segments set (molded) in an insulating disk. The superior switching properties of the carbon-fiber brushes is, incidentally, likely to lead to much work on their use in general, with or without semiconductor devices to assist commutation.

The future for DAC and TAC machines could be very bright with a sufficiently bold approach to the development of machines embodying novel constructional techniques. The principle may not be new, but the scope for inventiveness and originality is now wide, recognizing that what was proposed many years ago was a sound principle and that modern devices and materials make its application very promising.

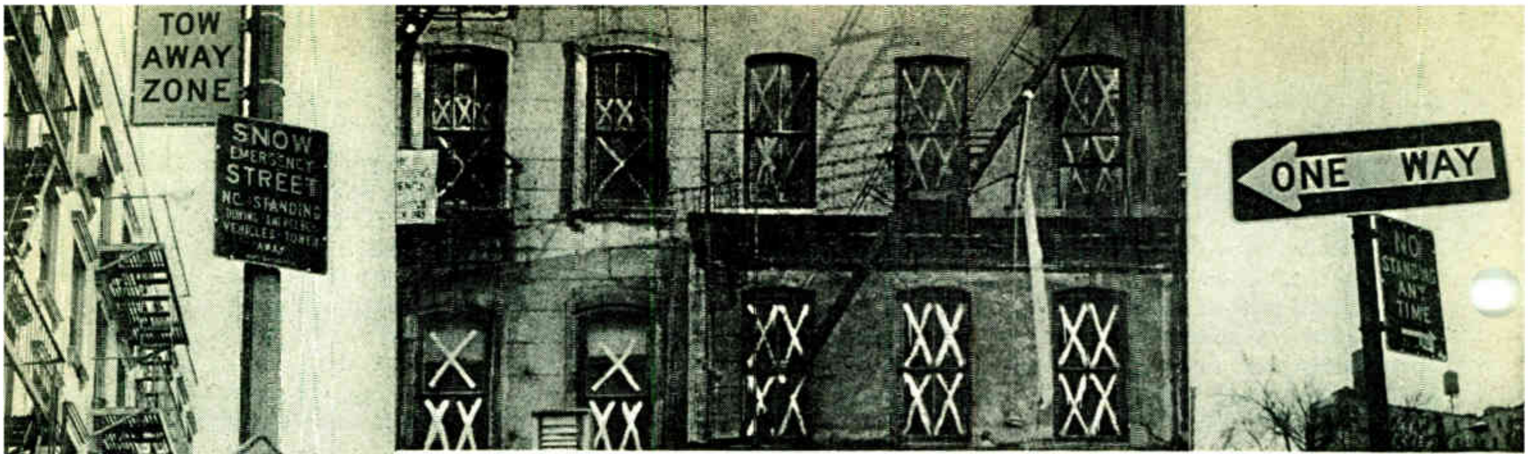
FIGURE 20. Principle of rolling-contact systems for use with TAC machines.



REFERENCES

1. Slepian, J., U.S. Patent 1 810 033, 1931.
2. Westendorp, V. S., U.S. Patent 2 231 607, 1941.
3. Klimov, N. S., "Improvement of commutation conditions in dc machines," *Elektrichestvo*, vol. 8, 1959.
4. Bates, J. J., "Thyristor assisted commutation in electrical machines," *Proc. IEE (London)*, vol. 115, June 1968.
5. Bates, J. J., Stanway, J., and Sansum, R., "Contact problems in machines using thyristor assisted commutation," *Proc. IEE (London)*, vol. 117, Feb. 1970.
6. Bates, J. J., "Diode assisted commutation," *Proc. IEE (London)*, vol. 117, May 1970.
7. Sone, S., "Improvement of commutation of ripple current motors," *Elec. Eng. Japan*, vol. 89, no. 1, 1969.
8. Bates, J. J., British Patents 1,079,792, 32571/65, 3845/67, 10624/69, 43481/69, 47304/69, 54686/69.

James J. Bates is head of the Electrical Engineering Branch at the Royal Military College of Science, Shrivenham, Swindon, England. He received the B.Sc. degree in 1948 from Birmingham University. After apprenticeships with the English Electric Co. and the British Electricity Authority (now the Central Electricity Generating Board), he returned to Birmingham University in 1950 to do postgraduate work on temperature rises in electric machines; he holds the Ph.D. degree from Birmingham. After a short period with the British Iron and Steel Research Association, London, he joined the Royal Military College of Science, Shrivenham, in 1954 as a lecturer in the Electrical Engineering Branch. In 1959 he moved to Imperial College, London, as a lecturer in the Electrical Engineering Department, specializing in electric machines and electric traction. In 1965 he became the first Reader in Electrical Engineering Science at the University of Durham. He was appointed to his present position at Shrivenham in 1967.



Systems analysis as a tool for urban planning

With the use of a computer model to simulate the growth, decline, and stagnation of a city, we can see how system structures and policies interact to create the urban ills surrounding us

Jay W. Forrester **Massachusetts Institute of Technology**

Ways for analyzing social systems result in new policies for improving the behavior of systems in which we live. Such policies can change urban slums into areas designed for self-renewal. Already, studies into the relationships between monetary policy, interest rates, and foreign exchange have thrown new light on the processes of corporate growth, product stagnation, and loss of market share, and on the growth and decline of cities. At one time the engineer's task was simply to balance financial cost against the economic performance of his technology. Now, psychological stress, ugliness, and crowding have become part of the cost. Engineers who fail to realize this broadened role will be vilified by a society that views them as insensitive to the needs of the times.

Most of the traditional steps taken to alleviate the conditions of our cities may actually be making matters worse. This is one of the conclusions of my book, *Urban Dynamics*,[1] which shows the city as an interacting system of industry, housing, and people. By presenting a computer model that interrelates these components of the city, the book shows how interacting processes produce urban growth, then cause growth to give way to stagnation. Job training programs, job creation by bussing to suburban industries or by the government as an employer of last resort, financial subsidies to the city, and low-cost-housing programs—these presently popular proposals are tested and shown to lie between neutral and detrimental in their effect on a depressed urban area.

The evolution of an urban area from growth into stagnation creates a condition of excess housing compared with the population and the availability of income-earning opportunities. Reducing the inherent excess housing of depressed areas and converting part of the land to industrial use appear necessary to reestablish a healthy economic balance and a continuous process of urban renewal. Such actions can produce a large enough wage and salary stream from the outside economy to make the area self-sustaining.

These results are controversial but, if right, they mean that many policies for remedying urban troubles may be turning growth into decline. Although I present here some results from the book, my principal emphasis is on the importance of systems analysis as a bridge between engineering and the social sciences.

Industrial dynamics

Over a decade ago at M.I.T. we began to examine the dynamic characteristics of managerial systems. The field known as *industrial dynamics* resulted.[2] Industrial dynamics belongs to the same general subject area as feedback systems, servomechanisms theory, and cybernetics. Industrial dynamics is the study of how the feedback loop structure of a system produces the dynamic behavior of that system. In managerial terms industrial dynamics makes possible the structuring of the components and policies of a system to show how the resulting dynamic behavior is produced. In terms of social systems it deals

The first IEEE publication of this article was in IEEE TRANSACTIONS ON SYSTEMS SCIENCE AND CYBERNETICS, vol. SSC-6, pp. 258–265, October 1970. The article appears here with the consent of the author and the TRANSACTIONS.

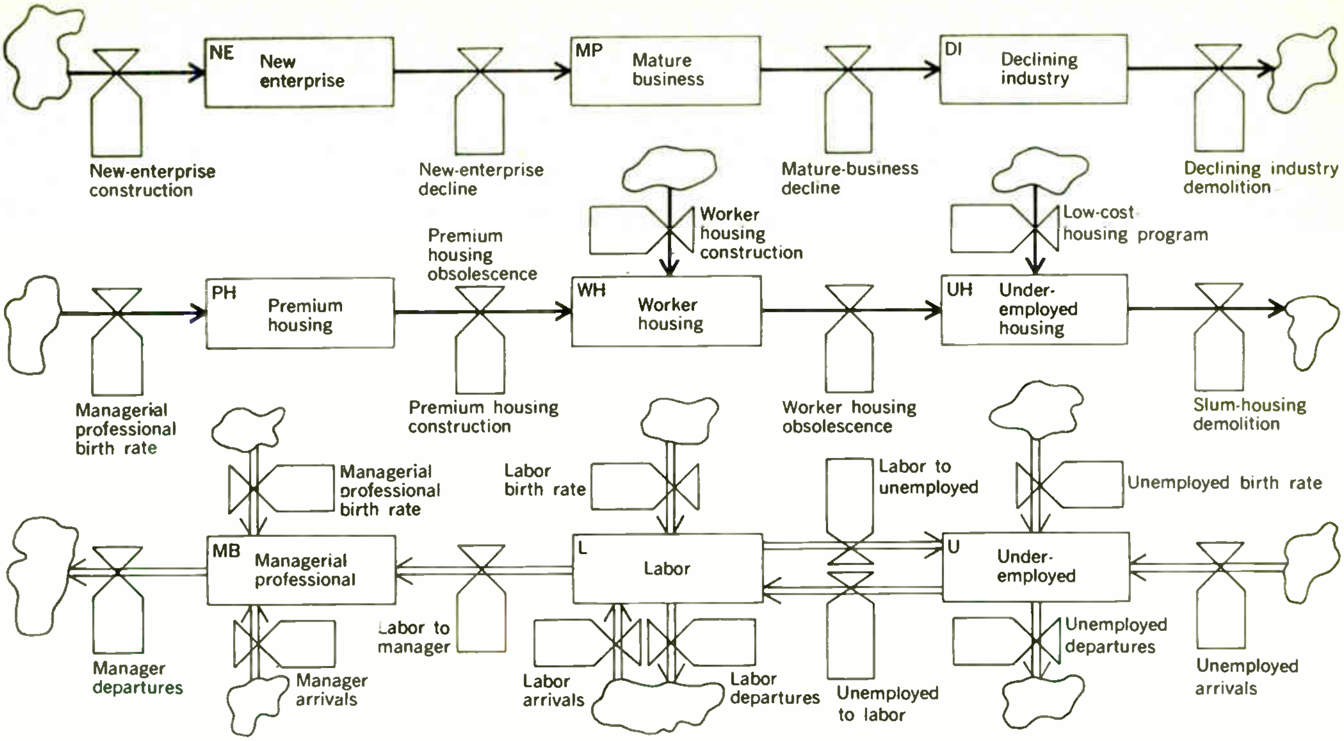


FIGURE 1. In this simplified urban system, nine levels are grouped into three subsystems. Across the top the industrial sector contains commercial buildings in three categories distinguished primarily by age. Across the center are residential buildings in three categories, also distinguished by age and condition. Across the bottom are three economic categories of population. Because of their complexity, the information linkages connecting the systems levels to the system rates are not shown. In this figure, the levels (rectangles) and rates (valves), along with the information linkages, represent the system components within the dynamic boundary. The "cloud" symbols are the sources or destinations of flows going from or to the outside environment. The flows from or to the outside are controlled only by conditions within the system.

with the forces that arise within a system to cause changes through time.

A design study of a social system seeks changes in structure and policies that will improve the behavior of the system. Some people recoil at the thought of designing social systems. They feel that designing a society is immoral. But we have no choice about living in a system that has been designed. The laws, tax policies, and traditions of a society constitute the design of a social system. Our only available choice is between different designs. If we lament the functioning of our cities or the persistence of inflation or the changes in our environment, we mean that we prefer a social system of a different design.

The six steps in the design process are: (1) observe the behavior modes of a system to identify the symptoms of trouble; (2) search for the feedback structures that might produce the observed behavior; (3) identify the level and rate variables making up that structure and explicitly describe them in the equations of a computer simulation model; (4) using the computer model, simulate in the laboratory the dynamic behavior implicit in the identified structure; (5) modify the structure until components and the resulting behavior agree with the observed conditions in the actual system; (6) introduce modified policies into the simulation model to find usable and acceptable policies that give improved behavior.

Surprising discoveries come from this combination of theory and laboratory experimentation: relatively simple structures produce much of the complex behavior of real-life systems, and people's skills in perception appear very different from those commonly supposed. It is often asserted in the social sciences that people are unreliable in analyzing their own actions, yet we find time and again that the policies and practices that people know they are following are the ones that interact to produce the most troublesome consequences. Conversely the vaunted powers of judgment and intuition usually deceive the person who tries to guess the time-varying consequences that follow even from a completely known system structure. We find that the modes of behavior that are most conspicuous in managerial, urban, and economic systems are produced by nonlinearities within those systems. The linearized models that have been used in much of engineering and the social sciences cannot even approximate the important modes of manifestations of nonlinear interactions. Including the so-called intangible factors relating to psychological variables, attitudes, and human reactions is relatively straightforward. Again, if the influences can be discussed and described, they can be inserted in the policy structure of a model. Any person who discusses why people act the way they do, or explains a past decision, or anticipates a future action is relating the surrounding circumstances to the corresponding human response. Any such discussion is a description of decision-making policy. Any such policy statement can be put into a system model.

Urban systems

A body of dynamic theory and principles of structure is emerging that allows us to organize and understand

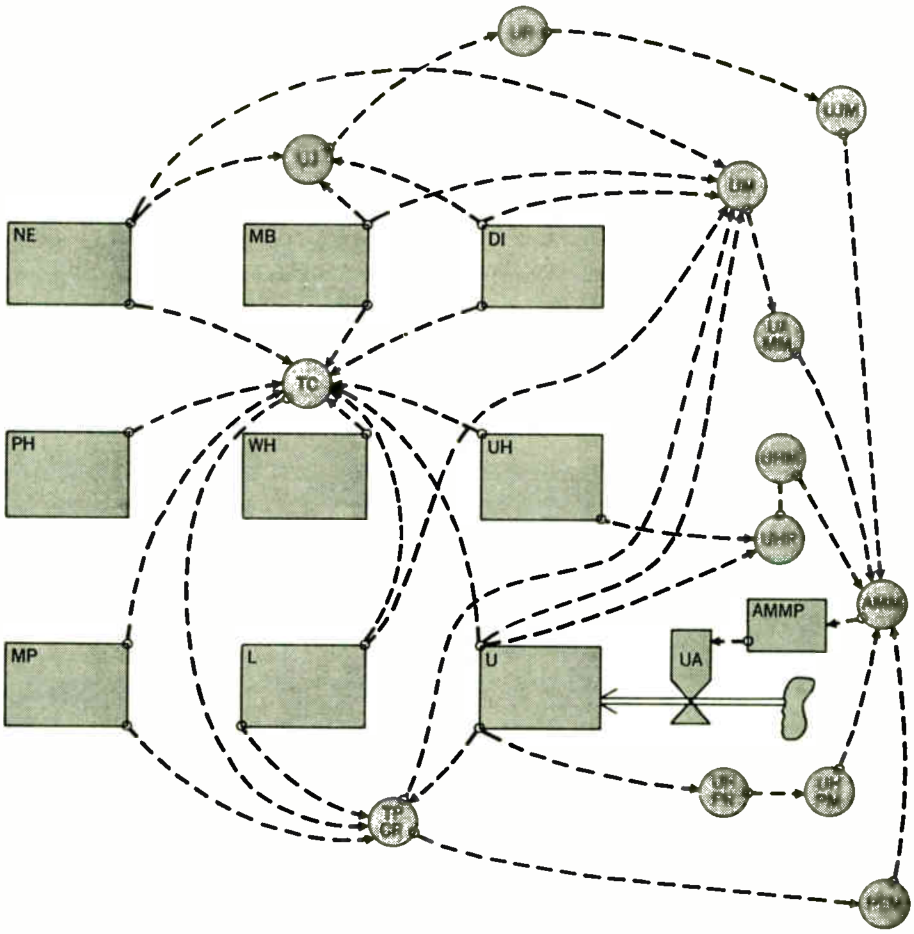


FIGURE 2. Information links to the underemployed-arrival rate. Five components of attractiveness are shown here. In the upper right corner UJM (underemployed/job multiplier) relates the population to available jobs and represents the income earning attractiveness of the area. The circle UAMM generates the attractiveness created by upward economic mobility. In other words, an area with high upward economic mobility is more attractive than one offering no hope of advancement. The circle UHM relates the underemployed population to the available housing. The area becomes more attractive as housing becomes more available. UHPM represents the attractiveness of a low-cost-housing program, if such exists. In the lower right corner PEM is the influence on attractiveness of the public expenditure per capita. As per-capita expenditure rises, it means better public services, better schools, and higher welfare budgets.

complex systems.[3] For example, the feedback loop becomes the basic building block of systems. Within the feedback loop there are two, and only two, kinds of variables. One is the *level variable* produced by integration; the other is the policy statement or *rate variable* that governs the changes in a system. The level variables are changed only by the rates of flow. The rate variables depend only on the levels. Any path through a system network encounters alternating level and rate variables. These and many other principles of structure are universal in the entire sweep of systems that change through time. Furthermore, the structure of a system determines its possible modes of behavior. Identical structures recur as one moves between apparently dissimilar fields. These identical structures behave in identical ways wherever they are found.

The same principles of structure and the same relationships between structure and behavior apply to a simple swinging pendulum, a chemical plant, the processes of management, internal medicine, economics, power politics, and psychiatry. A universal approach to time-varying systems that seems capable of dealing with systems of any complexity is emerging. Students, as they master the principles and practice of dynamic analysis, develop a remarkable mobility between fields of endeavor. The same person can clarify the dynamics of how a transistor functions, organize the processes of a public health epidemic, design new management policies to avoid stagnation in product growth, discover the sensitive factors in ecological change, and show how government policies affect the growth and decline of a city.

Figure 1 shows the central structure of an urban area. The nine rectangles represent the selected level variables. The twenty-two valve symbols represent the rates of flow that cause the nine system levels to change. Engineers often refer to these level variables as the state variables of a system. The distinction between level and rate variables is also familiar to anyone who examines financial statements. Balance sheet variables are always separated from variables on the profit-and-loss statement. They are separate because they are conceptually quite different. The balance sheet variables are system levels. They are created by accumulating financial flows. The profit-and-loss variables are system rates. This sharp distinction is found in all systems.

In Fig. 1 one can begin to detect the reasons for urban decline. The age of a building tends to determine the character of its occupants. A new commercial building is occupied by a healthy, successful commercial organization that uses relatively more managers and skilled workers than those who are unskilled. As the building ages, it tends to house a progressively less successful enterprise with lower employment skills. In addition to the changing employment mix as the industrial building ages, there is a tendency for total employment per unit of floor space to decline. On the other hand, as residential buildings age there is a tendency for occupancy to increase as well as to shift to a lower economic category of population. One perceives then a condition where the aging of buildings in an urban area simultaneously reduces the opportunities for employment and increases the population. The average income and standard of living decline.

Figure 2 shows the same nine system levels and one of

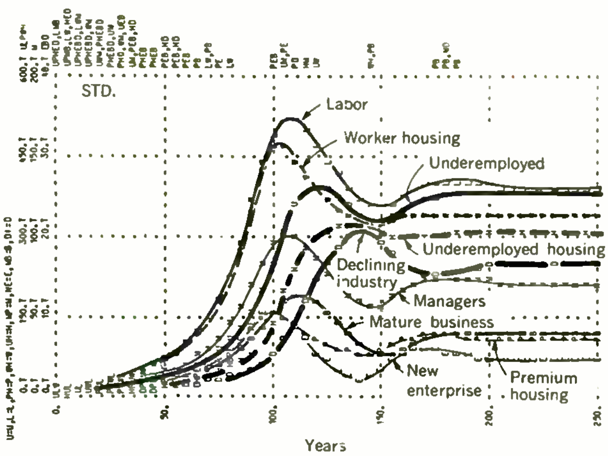


FIGURE 3. Growth and stagnation.

FIGURE 4. Changes in housing and employment.

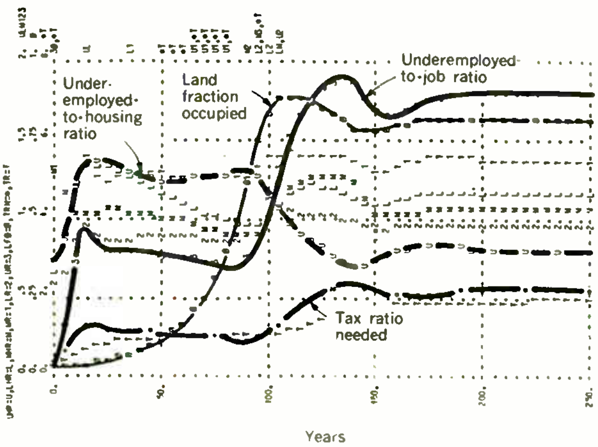
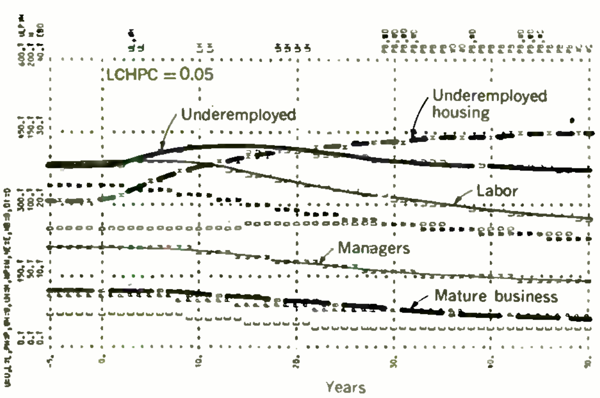


FIGURE 5. Decline of urban area caused by low-cost-housing construction each year for 2.5 percent of the underemployed population of a city.



the twenty-two flow rates. The dotted lines are the information linkages from the system levels to control the one flow rate—here the arrival of underemployed population into the urban area. The various levels of the system combine to create a composite *attractiveness,* which determines the inflow rate to the area. If the area is more attractive than those from which people might come, a net inward population flow occurs. If the area is less attractive, an outward flow dominates.

The concept of attractiveness is fundamental to the population flows. All of the characteristics of an area that make it attractive, these five and many more, combine to influence migration. An attractive area draws people. But almost every component of attractiveness is driven down by an increase in population. If there is an excess of housing, the area is attractive but a rising population crowds the housing. If there is an excess of jobs the area is attractive but the incoming flow of people fills these jobs. In other words, migration continues until the attractiveness of the area falls and becomes equal to any other places people might come from.

An important idea follows from examining these components of attractiveness. In a condition of population equilibrium, all areas must be equally attractive to any given population class, otherwise net migration would occur. If one component of attractiveness is increased in an area, other components must necessarily fall to establish a new equilibrium. Compensating changes in the components of attractiveness explain many past failures in our cities wherein we attempt to improve one aspect of the city only to discover that other aspects have become worse.

In making a laboratory model of a social system one should not attempt straightaway to solve a problem. Instead one should generate a model that will create the trouble symptoms. Only if one fully understands the processes whereby difficulties are created can he hope to correct the causes. This means that we want a model of an urban area that can start with empty land, grow a city, and show the processes whereby economic health falters into stagnation and decay.

As another guide to modeling, one should start, not by building a model of a particular situation, but instead by modeling the general class of systems under study. This may seem surprising, but the general model is simpler and initially is more informative than a model of a special case. Here we wish to model the general process of urban growth and stagnation. It should be a model that, with proper changes in parameters, is good for New York, Calcutta, a gold rush camp, or West Berlin. These all seem to have very different characteristics but they have certain elements in common that describe their urban processes. There are fewer concepts common to all than are to be found in any one. The general model can strip away the multitude of detail that confuses any one special situation. The general model identifies the central processes and is a statement of the theory for the entire class of systems.

Figure 3 shows the behavior of the laboratory model of an urban area. It presents the nine system level variables over 250 years. The first 100 years is a period of exponential growth, but then the land area becomes filled, growth ceases, and the aging process begins. At year 100 near the end of the growth phase the labor population is almost double the underemployed population. But by year 150, the labor population has fallen and the underemployed population has risen until these two groups are almost equal. Business activity has declined and the area has taken on the characteristics of a depressed city. This has occurred because of the way that the industry, housing, and populations in Fig. 1 have interacted with each other.

Figure 4 shows other variables during the same 250 years. Notice especially the underemployed/job ratio and the underemployed/housing ratio. During most of the first 100 years of growth these two ratios were almost constant. The underemployed/housing ratio was high

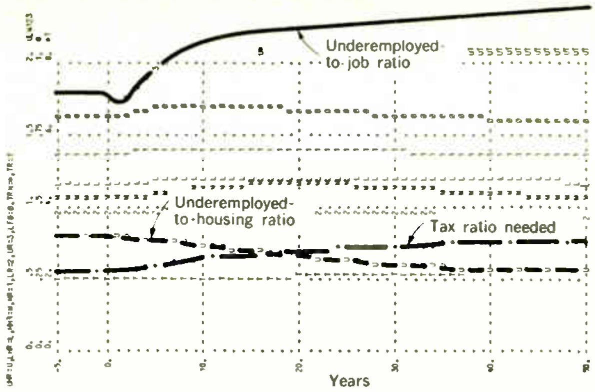
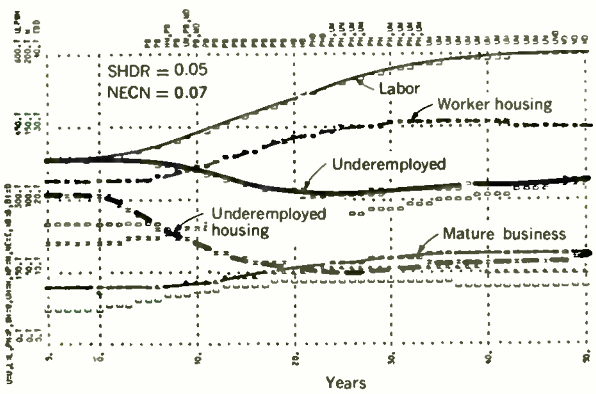


FIGURE 6. Rising unemployment and falling occupancy.

FIGURE 7. Revival caused by removing 5 percent of underemployed each year and encouraging business construction to generate jobs.



(above the center of the figure), meaning that the population is large compared with the housing. In other words, during the first 100 years there was a housing shortage for the underemployed population. On the other hand, the underemployed/job ratio was low, meaning that the population was below the job opportunities, jobs were readily available, economic opportunity was good, and upward economic mobility was high. During this early period of growth and high economic activity, the underemployed population was being effectively adjusted in relation to other activity by balancing good economic opportunity against a housing shortage.

But between 90 and 140 years, notice the sharp reversal of the curves for underemployed/job ratio and underemployed/housing ratio. Within this fifty-year span, the underemployed have increased while available jobs decreased; the result is a precipitous rise in unemployment. But in this same period, the housing that is aging and becoming available to the underemployed is rising even more rapidly than the underemployed population. Jobs have become scarce while housing has become surplus. The model is behaving the way our cities do.

Many people do not seem to realize that the depressed areas of our cities are areas of excess housing. The economy of the area is not able to maintain all of the available housing. Because of low incomes, people crowd into some dwelling units while other buildings are abandoned, stand idle, and decay.

Recall the earlier comments about compensating movements in the components of attractiveness. Here, as housing becomes more available, jobs become more scarce. The stagnating urban area has become a social trap. Excess housing beckons people and causes inward migration until the rising population drives down the standard of living far enough to stop the population inflow. Anything that tends to raise the standard of living is defeated by a rise of population into the empty housing.

Figure 5 shows fifty years beginning with the conditions found at the end of Fig. 3. At time zero, a low-cost-housing program is introduced which each year builds low-cost housing for two and one half percent of the underemployed population. Observe what happens. Underemployed housing, which is being actively constructed, rises forty-five percent, but premium housing falls thirty-five percent, and worker housing falls thirty percent. New enterprise declines fifty percent and mature business declines forty-five percent, all in the fifty-year period. Economic conditions become sufficiently worse so that even the underemployed population, although it rises initially, eventually falls to slightly less than its beginning value. *These changes are a result of the low-cost housing program.*

In Fig. 6, the corresponding underemployed/job ratio has risen thirty percent, indicating substantially higher unemployment, while the underemployed/housing ratio has fallen thirty percent, indicating a still higher excess of housing. Again, the two components of attractiveness compensate for one another with better housing and a falling standard of living. In the long run, the low-cost-housing program has not served the interests of the low-income residents. Instead, it has intensified the social trapping characteristic of the area. Over the period, the tax levies rise thirty-five percent. The area has become worse from almost all viewpoints.

In this same manner job training programs, job creation programs, and financial subsidies were examined. All lie between ineffective and harmful. The low-cost-housing program was the most powerful in depressing the condition of a stagnant urban area.

The depressed areas of our cities seem to be characterized by excess housing compared with jobs and by excessive concentration of low-income population. These conditions, created by aging industrial and dwelling buildings, interact to drive out the upper-income population and business activity, and to reduce the tax base. Once the decline starts, it tends to accelerate. Unless one can devise urban management policies that produce continuous renewal, difficulties are inherent.

Figure 7 shows an urban condition that begins with stagnation and then changes toward revival. Here five percent of the slum housing is removed each year and the incentives for new enterprise construction are increased somewhat. The result is a cascading of mutual interactions that raise the economic activity of the area, increase upward economic mobility for the underemployed population, and shift the population internally from the underemployed to the labor class. This is done without driving the existing low-income population out of the area. Underemployed housing is reduced. Initially this reduction comes largely from the empty housing. The resulting housing shortage restrains the population inflow that would otherwise defeat the revival of the area.

Figure 8 shows the same fifty-year span as in the preceding figure. Here again, employment and housing move in opposite directions. The underemployed/job ratio falls, which means more jobs and lower unemploy-

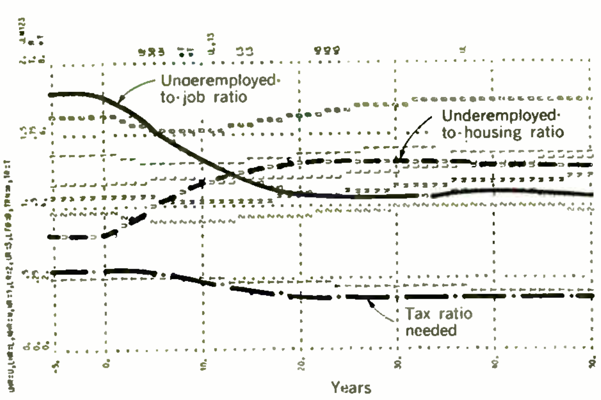


FIGURE 8. Falling unemployment and rise in housing occupancy for the same 50 years of Fig. 7.

ment. On the other hand, the underemployed/housing ratio rises, which means a tighter housing situation. If the economic circumstances are to be improved, we must accept some compensating change in other components of attractiveness. Here it is the increased tightness of housing that allows job opportunities to increase faster than population until a good economic balance is reached. I stress economic revival as the first stage of rebuilding a depressed area because it appears that an economic base must precede social and cultural development.

It is simply not possible to increase all of the attractiveness components of an area simultaneously. Attractiveness is here defined in a very broad sense. For example, legal restrictions like an immigration barrier into a country can produce enough "unattractiveness" to inward migration so that components might be maintained at a high level. But wherever one component of attractiveness is high others will be found low. Engineers, especially, should consider the compensating changes that will occur in the attractiveness components of an area because engineers tend to deal with economic considerations and technology. Economic and technical factors are more concrete than the intangible *quality of life* variables. The economic and technical aspects of a city are the ones we most easily see how to improve. Our technological society tends, therefore, to observe, react to, and improve the economic and technical aspects of a city. Such improvements increase the technical and economic components of urban attractiveness. But as a result, population density rises until the urban area once again reaches attractiveness equilibrium with its environment. The burden of forced reduction in other components of attractiveness falls on the quality of life variables—crowding, pollution, and psychological stress. These less tangible variables have been weak and hard to measure, and have been defenseless against the persuasiveness and the certainty of improvement shown by the technical and economic considerations. But we are entering a time when a reversal will occur between the formerly weak and strong variables. For a substantial fraction of our population, the standard of living is already high enough so that more gain in the economic and technical areas will come at too high a price in the quality-of-life components of our environment. The engineer, if he continues to serve society, must balance a greater number of social needs against one another. At one time his task was simply to balance financial cost against economic performance of his technology. Now the product and also the medium of payment are both expanding. Social value and quality of life become part of the product. Psychological stress, ugliness, and crowding become part of the cost. Engineers who fail to recognize this broadened role will be vilified and castigated by a society that perceives them as narrow and insensitive to the demands of the times.

When a system misbehaves, we should ask ourselves what policies within that system cause the undesirable characteristics. If we examine the laws under which a city operates, we see a structure of regulations that could hardly be designed better to create stagnation and decline. The aging and decay of buildings is central to the urban decline process, yet we see throughout our tax laws and regulations numerous incentives to keep old buildings in place. As the value of a building decreases, so do the assessed taxes. The reduced expense makes it possible to retain the old building longer. Under some circumstances the value of a building can be depreciated several times for income tax purposes. This produces incentives to keep an old building in place. This is not the place for detail, but it seems clear that a different set of tax laws and city regulations could be devised to produce the individual incentives necessary for continuous renewal. As an example, I recently saw a suggestion that each building have a mandatory trust fund into which the owner must pay a levy each year. At any time, whoever owns the building can draw out the money if he demolishes the building and clears the land. This, you see, would create an earlier incentive for replacement. Property tax levies and income tax accounting could both be changed to produce pressures in the same direction.

These studies of managerial, urban, and other social systems have uncovered complex systems characteristics that serve to identify potential detrimental modes of behavior. First, complex systems are counterintuitive. They behave in ways that are opposite to what most people expect. They are counterintuitive because our experience and intuition have been developed almost entirely from contact with simple systems. But in many ways, simple systems behave exactly the opposite from complex systems. Therefore, our experience misleads us into drawing the wrong conclusions about complex social systems.

Second, complex systems are strongly resistant to most policy changes. A new policy tends to warp the system so that slightly changed levels present new information to the policy points in the system. The new information, as processed through the new policies, tends to give the old results. There are inherent reasons within complex systems why so many of our attempts at correcting a city, a company, or an economy are destined to fail.

But third, the converse is also true. There are points in systems from which favorable influence will radiate. Often these points are difficult to perceive, and the required action is the opposite of what is expected. But when these points are found, they tend to radiate new information streams in such a way that the new circumstances, when processed through the old attitudes and policies, produce a new result.

Fourth, complex systems tend to counteract most active programs aimed at alleviating symptoms. For example. Chapter 4 in *Urban Dynamics* shows how a job training program can increase the number of underemployed in a city. When outside action tries to alter the condition of a system, the system relaxes its own internal processes aimed at the same result and throws the burden ever more onto the outside force that is attempting to produce a correction. The internal need for action is reduced and the external supplier of action must work ever harder.

Fifth, in complex systems the short-term response to a policy change is apt to be in the opposite direction from the long-term effect. This is especially treacherous. A policy change that improves matters in the short run lays a foundation for degradation in the long run. The short tenure of men in political office favors decisions that produce results quickly. These are often the very actions that eventually drive the system to ever-worsening performance. Short-run versus long-run reversal processes are all around us. If an agricultural country is to industrialize, it must accumulate railroads, factories, and steel mills. This capital accumulation can only be done by foregoing consumption and reducing the standard of living first in order that the standard of living may rise at a later time. If a company faces declining earnings because its products are obsolete, it must invest more heavily in product research and incur even deeper short-term losses if it is to recover in the more distant future to a profitable product stream. A student forgoes short-term earning opportunities by attending college to increase his longer-term earning capability. This reversal between the short run and the long run occurs repeatedly.

Sixth, a system contains internal dynamic mechanisms that *produce* the observed undesirable behavior. If we ignore the fundamental causes and simply try to overwhelm the symptoms, we pit two great sets of forces against one another. In general, our social systems have evolved to a very stable configuration. If the system is troublesome, we should expect that the causes of the trouble are deeply embedded. The causes will outlast our persistence in overwhelming the symptoms. Furthermore, the internal pressures usually rise to counteract a corrective force from the outside. We can expend all our energy to no avail in trying to compensate for the troubles unless we discover the basic causes and redesign the system so that it spontaneously moves to a new mode of behavior.

As the last of these characteristics of complex systems, we must recognize that a certain ensemble of conditions goes with each possible mode of a system. More specifically, each mode of a system is accompanied by a set of pressures characteristic of that mode. We cannot sustain a particular mode unless we are willing to accept the corresponding pressures. For example, contrast the depressed mode of a city in Figs. 5 and 6 with the revived mode in Figs. 7 and 8. The depressed mode is one characterized by the pressures that come from decaying buildings, low incomes, and social disorientation. But the revived mode also contains pressures. The revived mode is sustained by the housing shortage and the legal and tax pressures that generate a steady demolition and replacement of old buildings. But everyone in the system will want to alleviate the pressures. Active industry will want more employees; residents will want more floor space; and outsiders will want housing so they can move to the attractive job opportunities. Rents will be high. These pressures are easy to relieve by increasing the fraction of the land area permissible for housing, by keeping old buildings in place longer, and by allowing taller apartment buildings. But such moves will start the area back toward the depressed mode. We must decide the kind of system we want with knowledge of and acceptance of the accompanying pressures. Instead, much of our social legislation of the last several decades has consisted of trying to relieve one set of pressures after another. The result is a system mode characterized by inflexibility, conformity, crowding, frustration, supremacy of the organization over the individual, and a choking of the environment. And the resulting pressures, acting through the counterintuitive and short versus long-term reversal characteristics of complex systems, may well move us further in the same direction.

I am suggesting that the time is approaching when we can design social systems to obtain far better behavior. Different policies could change our urban areas from ones that are designed to deteriorate into ones that are designed for self-renewal. One can foresee a time when we will understand far better the relationships between monetary policy, interest rates, unemployment, and foreign exchange. Already such studies have thrown new light on the processes of corporate growth, on the reasons for product stagnation and loss of market share, and on the growth and decline of cities.

To design new policies for social systems requires a level of skill that is rare. The kind of system modeling and policy design I have been describing requires a professional training at least as extensive as that in any of the established professions. The proper training requires theory, laboratory, case studies, apprenticeship, and practicing experience.

This article is based on a paper that originally appeared in *The Engineer and the City*, proceedings of a symposium sponsored by the National Academy of Engineering, October 22–23, 1969. Proceedings are available from the Academy, 2101 Constitution Avenue, Washington, D.C. 20418.

REFERENCES

1. Forrester, J. W., *Urban Dynamics*. Cambridge, Mass.: The M.I.T. Press, 1969.

2. Forrester, J. W., *Industrial Dynamics*. Cambridge, Mass.: The M.I.T. Press, 1961.

3. Forrester, J. W., *Principles of Systems*. Cambridge, Mass.: Wright-Allen Press, Inc., preliminary printing of first ten chapters.

Jay W. Forrester (F) is professor of management at the Massachusetts Institute of Technology, where he directs research and teaching in the field of industrial dynamics. Dr. Forrester was graduated from the University of Nebraska with high distinction in 1939. That same year he joined M.I.T. as a research assistant in electrical engineering. Two years later he became a member of M.I.T.'s newly formed Servomechanisms Laboratory. Prior to his present position, which he assumed in 1956, Dr. Forrester was head of the Digital Computer Division of M.I.T.'s Lincoln Laboratory. In this capacity he guided the military operational planning and technical design of the Air Force SAGE system for continental air defense. He holds the basic patent for "random-access, coincident-current magnetic storage," which is now the standard memory device for digital computers.

In 1967 Dr. Forrester was elected to the National Academy of Engineering. He is a Fellow of both the American Academy of Arts and Sciences and the Academy of Management. In 1969 he received the honorary doctor of science degree from Boston University in recognition of contributions to computer technology, air defense, and urban dynamics. Also in 1969, the Danish Academy of Technical Sciences awarded him the Valdemar Poulsen Gold Medal. Dr. Forrester is a director of the ALZA Corporation. He is also a member of the American Physical Society, the American Economic Association, and ACM.

Demand-assigned service for the Intelsat global network

In April 1970 Intelsat decided to allow earth-station owners to access the Atlantic Intelsat IV satellite via the SPADE system. It is expected that by mid-1971 another milestone will have been achieved with the introduction of the first operational demand-assignment system

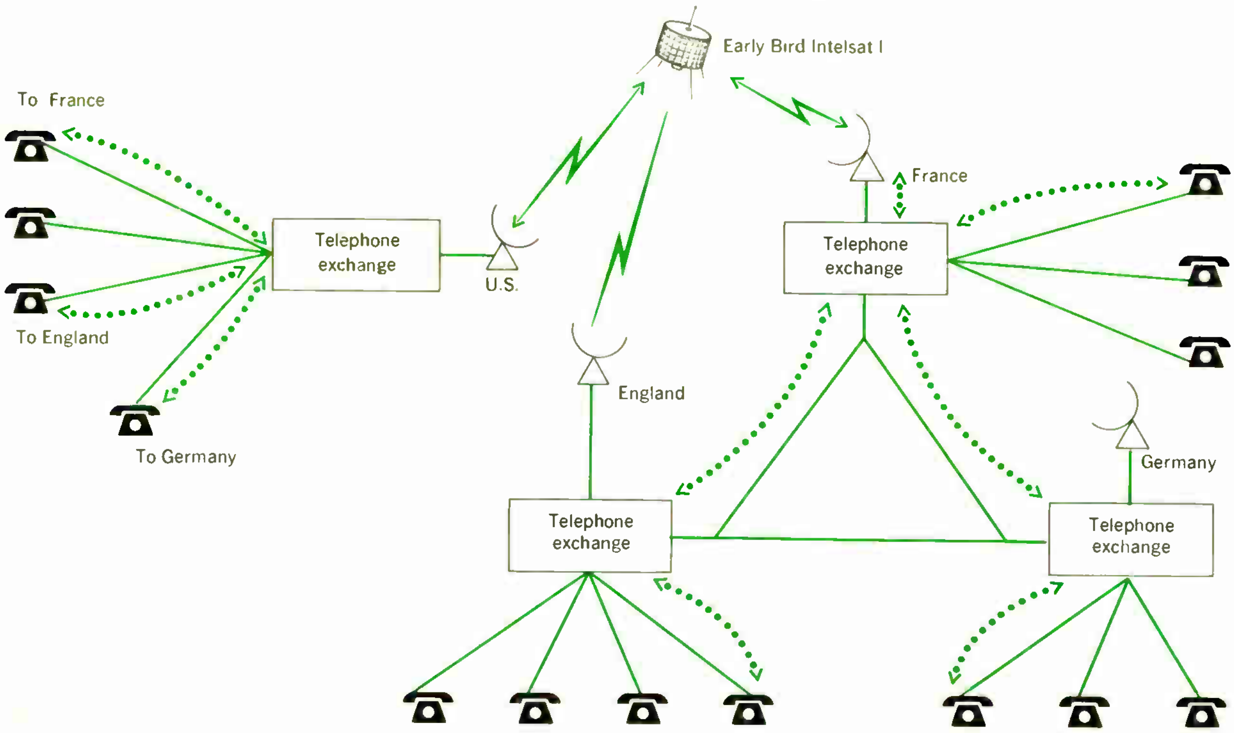
J. G. Puente, A. M. Werth Comsat Laboratories

Communications satellites have been in commercial operation for more than five years. With the growth of traffic requirements, particularly those of the developing nations, the early methods used in the Intelsat network for multiple access have had to be updated. The full potential of satellite communications can be realized by using the satellite to provide connections between any two users on demand. This article traces the development of the multiple-access techniques employed with Intelsat from the two access methods associated with Early Bird to the demand-assignment technique planned for Intelsat IV. This latter method, SPADE, is to be the first operational demand-assignment multiple-access system and is scheduled for use in 1971. Details of the SPADE system and its use for other applications are reviewed.

The launch of the Early Bird satellite in April 1965[1] initiated a new era in telecommunications. Its 240 voice circuits provided a greater capacity than all of the cables laid between the United States and Europe over the previous ten years. In addition, most of Western Europe was in sight of Early Bird and, for the first time, it was technically possible to provide direct, reliable, and instantaneous high-quality voice, television, and data communication between distant points.

Early Bird was a modified version of Syncom,[2] the first of the synchronous satellites successfully constructed by Hughes Aircraft Corp. and launched by NASA. Early Bird featured a hard-limiting transponder that used frequency-division multiple access, and thus restricted the number of earth stations accessing the satellite to two if channel capacity were to be maintained. Because of this

FIGURE 1. Early satellite operation. (Dotted lines show conversations in progress.)



characteristic becomes more linear. Other factors then become important, such as the conversion of amplitude to phase modulation in traveling-wave tubes.)

As the number of accesses increased, the satellite channel capacity still decreased, but not nearly as quickly as for a hard limiting satellite such as Early Bird. For this reason Intelsat II and III have normally been regarded as multiple-access satellites when operating with FM-FDMA. The objective in optimizing the satellite channel capacity was to have a minimum number of carriers entering the satellite, that is, one carrier from each earth station, carrying traffic to a number of distant destinations.

Thus evolved the multidestination FM carrier diagrammed in Fig. 2. The figure has been simplified to bring out the salient points. Long-distance telephone calls originating in country *A* enter a telephone exchange or transit center (CT) and are multiplexed as shown in the 60-channel baseband. Country *A* transmits on a single FM carrier f_A with a 60-channel capacity. It has preassigned the 60 channels in groups of 12 channels to be received by five other countries. Therefore, countries *B*, *C*, *D*, *E*, and *F* require an FM receive chain operating on frequency f_A in order to receive telephone traffic from *A*. Conversely, country *A* must have a receive chain for each of the countries in its baseband. Thus a receive chain is required for every country with whom communications are desired. This will become more important in our later discussion.

In tracing a telephone call from country *A* to country *F*, the CT switches the call to the proper slot in the baseband—i.e., one of the 12 channels allocated to station *F* in the baseband. The baseband continuously modulates an FM carrier on f_A and the carrier is transmitted to the satellite to be amplified and retransmitted to the earth. Stations *B*, *C*, *D*, *E*, and *F* receive the carrier f_A. At earth station *F* the received signal is demultiplexed, selecting only those 12 channels preassigned for it on the f_A carrier. The CT in country *F* then makes the connection to the party called via terrestrial facilities, and the connection is complete. Of course the return connection to country *A* from country *F* is accomplished in a similar manner via carrier frequency f_F.

Note that in this mode of satellite communications, if the 12 channels to country *F* in country *A*'s baseband are being used, a new call entering the CT would receive a busy signal from the satellite system even though the remaining 48 channels in the baseband might be unoccupied. We can see that multidestination carriers are a step toward improving the utilization of satellite capacity, but, as shown above, they are not fully utilizing the satellite's capability from the point of view of occupancy. Nevertheless, for large-capacity links between the two points, the statistics of call placement are such that very few calls are lost due to overload.

Consider next the case where connections are required between an earth station and many other users, but the total traffic between any paired earth station link is very low. To improve the efficiency of this model, the satellite capacity should somehow be shared in such a way that when a new call comes in and there is any unused capacity in the satellite, the connection can be established. The assignment of satellite capacity in preassigned, fixed blocks between two users prevents reallocation of unused channels. Therefore, for a network of light traffic links, satellite channels should be assigned on demand—that is, when a call comes into the CT. When the call is completed, the satellite channel is made available to any other pair of users. The introduction in the Intelsat network of demand assignment (DA)[4] represents the third major innovation in commercial satellite communications since 1965.

The need for demand assignment

Figure 3 illustrates the configuration of earth stations predicted for the Atlantic Basin in the early 1970s. The present projection of operating earth stations having access to an Intelsat IV satellite poses serious problems because of the increasing number of connections required among various earth stations. Potentially, a total of 820 links are possible for the 41-station Atlantic network. Considering that a large proportion of any such connections would carry light traffic, the satellite, operating as it does today, in an FDM/FM/FDMA (frequency-division multiplex, frequency-modulated, frequency-division multiple access) multidestination mode as described earlier, could not efficiently accommodate the increasing number of connections.

Figure 4 presents a summary of traffic data[5] and shows

I. Traffic for Argentina

Traffic Between Argentina and	Number of Circuits	Paid Minutes per Day
Brazil	**29***	3022
Canada	2	70
Chile	**35**	12 750
Colombia	4	15
France	7	350
Germany	6	350
Italy	**15**	824
Mexico	4	84
Panama	1	42
Peru	6	378
Spain	10	196
Switzerland	3	280
United Kingdom	10	220
United States	**52**	8882
Venezuela	4	70

* Boldface numbers indicate preassigned circuits.

II. Total traffic, Atlantic Basin, 1973

Preassigned voice channels	3618	
Preassigned record data (or alternate voice data) channels	1544	
Total channels preassigned	5162	
Voice-channel candidates for demand assignment	1158	
Total equivalent channels required	6320	
Percentage DA of total		18.3%
Number of paths (preassigned)	84	
Number of paths (demand-assigned)	266	
Total paths	350	
Percentage DA of total		76.0%

the density expected for each of the known predicted links. The black dots at the intersections of two countries represent traffic in excess of 12 circuits, the colored dots indicate traffic less than 12 circuits. (A figure of 12 circuits has been taken arbitrarily for illustrative purposes since it represents a standard group.)

It is evident from Fig. 4 that the majority of the presently predicted traffic between these links is relatively light. To illustrate this fact, Table I examines in more detail the traffic for Argentina.[6] Note that of 15 links for Argentina, only 4 have a requirement for more than 12 circuits. Similar analyses of traffic for most other nations result in nearly the same distribution.

Table II summarizes the total predicted traffic for the Atlantic Basin in 1973. Of the 6320 total channels required, 1158 (or 18.3 percent) are candidates for demand assignment. Of particular interest in these statistics is that there is a predicted total of 350 paths. Using the guideline that 12 circuits or less should be served by demand assignment, 266 (or 76 percent) of these paths fall into this category. Furthermore, it should be noted that in the matrix of Fig. 4 there is a sizable number of boxes with no dots. Although these blanks may represent paths for which no predictions are available, they generally represent paths over which traffic is so light that it is not economically attractive to service or even plan these links. It follows that the growth of the Intelsat network will occur not only between those areas indicated by dots, but also between those locations where economical communications links have previously been impossible.

Experience in satellite communications shows that there will be a need for accommodating the heavy trunks on a preassigned basis between certain locations. These links are economically justifiable because they provide efficient service. It is evident, however, that if all desired links were to be handled on this preassigned basis, the capacity of the satellite would quickly be exhausted. Therefore, it appears that the answer to the problem of providing efficient global service consists of implementing a mixture of preassigned and demand-assigned services.

Figure 5 illustrates methods of assigning multidestination satellite channels. In a preassigned (PA) system, if earth station A has traffic to stations B through F, it can multiplex this traffic on a single carrier and transmit it through the satellite. Then each destination earth station would receive station A's carrier and extract the communications addressed to that particular destination. The lower half of Fig. 5 shows that, using demand assignment, earth station A would connect to earth station B on a demand basis over channel 3, etc. After each of these communications, the channels would be available to other pairs of users. The fundamental difference between PA and DA is that for DA the channels are never permanently assigned between any pair of stations.

It is appropriate at this point to summarize briefly the key goals Intelsat expects to achieve with the implementa-

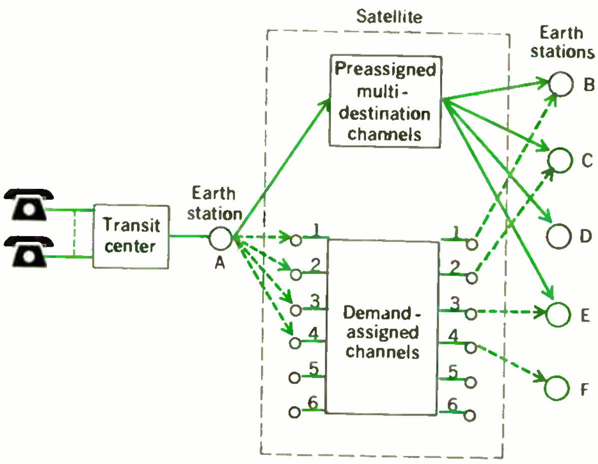
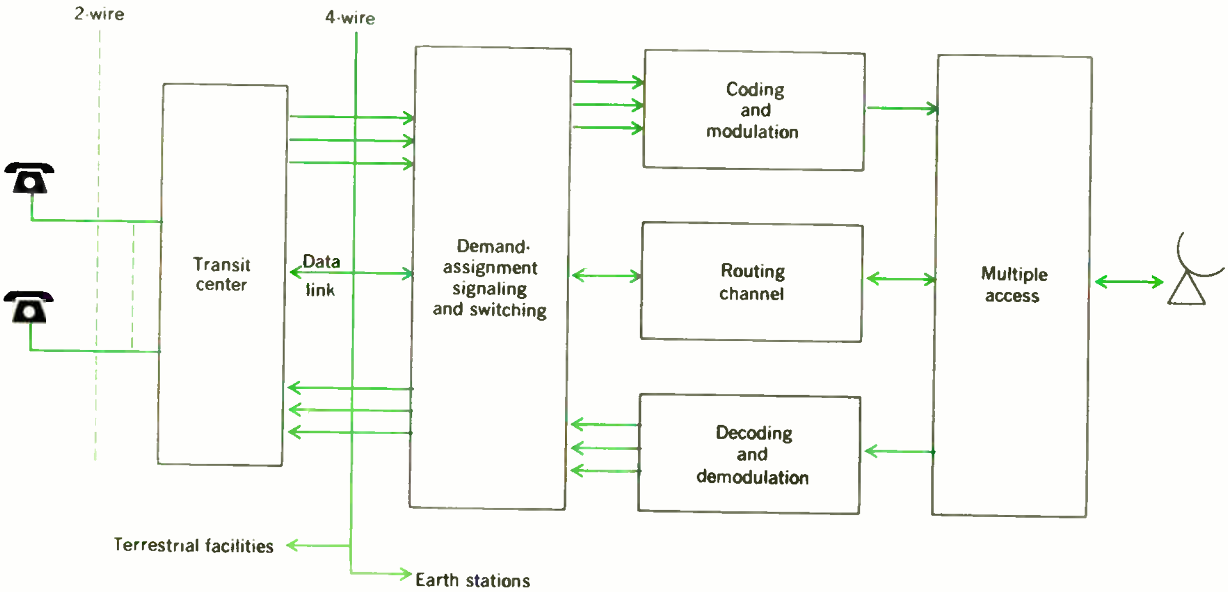


FIGURE 5. Preassigned and demand-assigned operation of satellite channels.

FIGURE 6. Functional earth-station requirements of demand-assigned satellite systems.



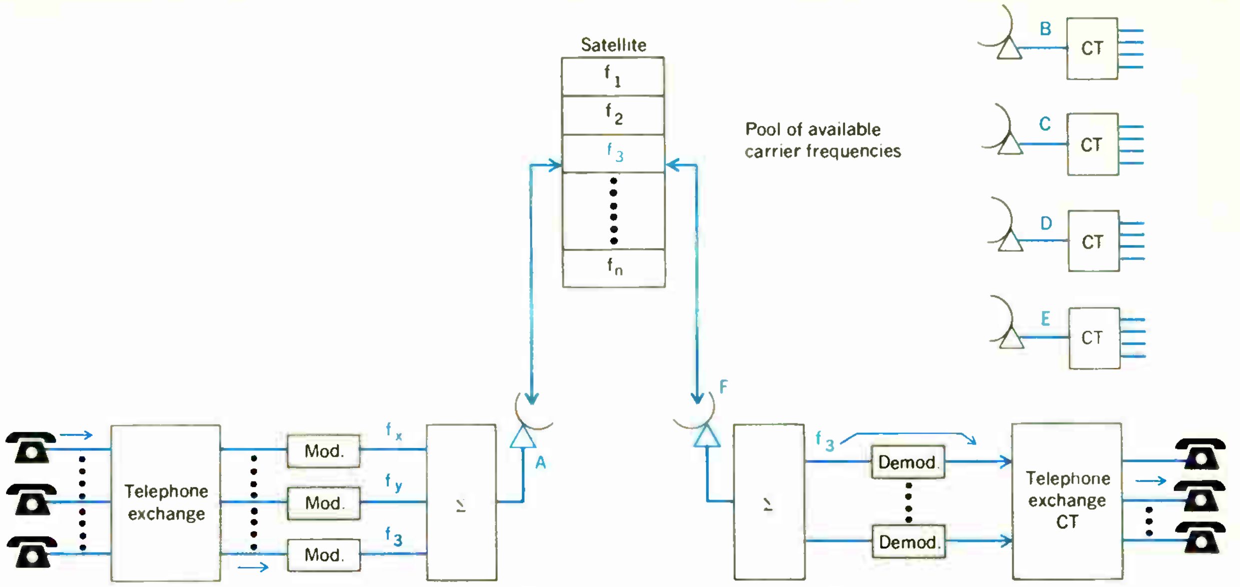
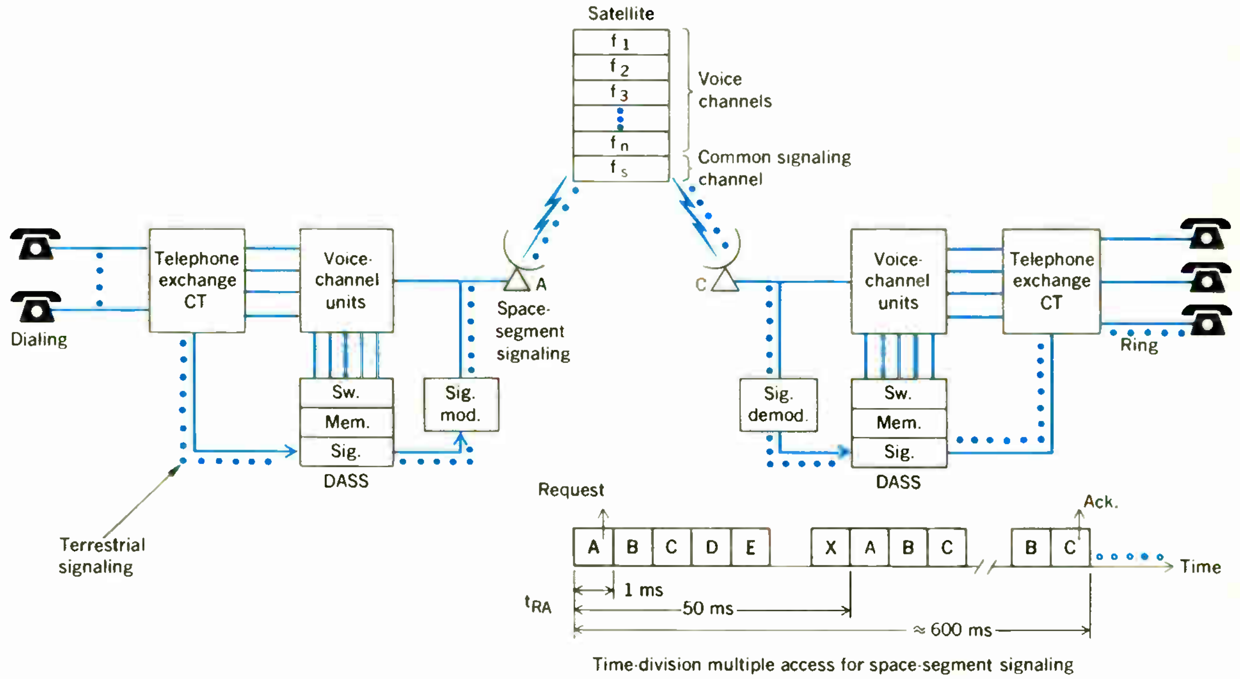


FIGURE 7. Voice-signal flow in a demand-assignment FDMA system. (f_x = carrier frequency; x = 1, 2, 3, . . ., n.)

FIGURE 8. Signaling flow in demand-assignment system.



tion of a demand-assignment system. They are:

1. To provide efficient service to light traffic links.
2. To handle overflow traffic from medium-capacity preassigned links.
3. To allow establishment of a communication link from any earth station to any other earth station within the same zone on demand.
4. To utilize satellite capacity efficiently by assigning circuits individually.
5. To make optimum use of existing earth-station equipment.

The increased total number of terrestrial circuits that might be served by a DA pool through a more efficient use of the space segment could delay the necessity of an additional satellite in a given region. This not only would provide a space-segment cost saving, but also could provide an additional saving to some earth-station owners by deferring for a substantial time period the need for an additional antenna.

It is also important to note that DA channels have a greater equivalent traffic-carrying capability than PA channels. From the point of view of traffic, DA channels,

III. Some technical characteristics of the SPADE system

Channel encoding	PCM
Modulation	4-phase PSK (coherent)
Bit rate	64 kb/s
Bandwidth per channel	38 kHz
Channel spacing	45 kHz
Stability requirement	±2 kHz (with AFC)
Bit error rate at threshold	10^{-4}
TDMA common signaling channel:	
Bit rate	128 kb/s
Modulation	2-phase PSK
Frame length	50 ms
Burst length	1 ms
Number of accesses	50 (49 stations + 1 reference)
Bit error rate	10^{-7}

since they are not permanently connected between two points, can serve a larger number of terrestrial circuits than can PA channels. It is estimated that a two- to threefold increase in circuit utilization can thus be effected.[7]

In an effective DA system it is possible to establish communications between any two stations in common view of a satellite; therefore, even the least dense traffic links can be accommodated with no penalty to the system. With this capability the loading on these links would be allowed to grow in gradual steps as the availability of the service attracts new traffic.

It should be recognized that the advantages of DA apply most effectively to light-to-medium-capacity multidestination traffic. As noted earlier, the efficiency, both in terms of equipment and spectrum utilization, of high-density limited-destination trunks requires that a mix of preassignment and demand assignment be maintained.

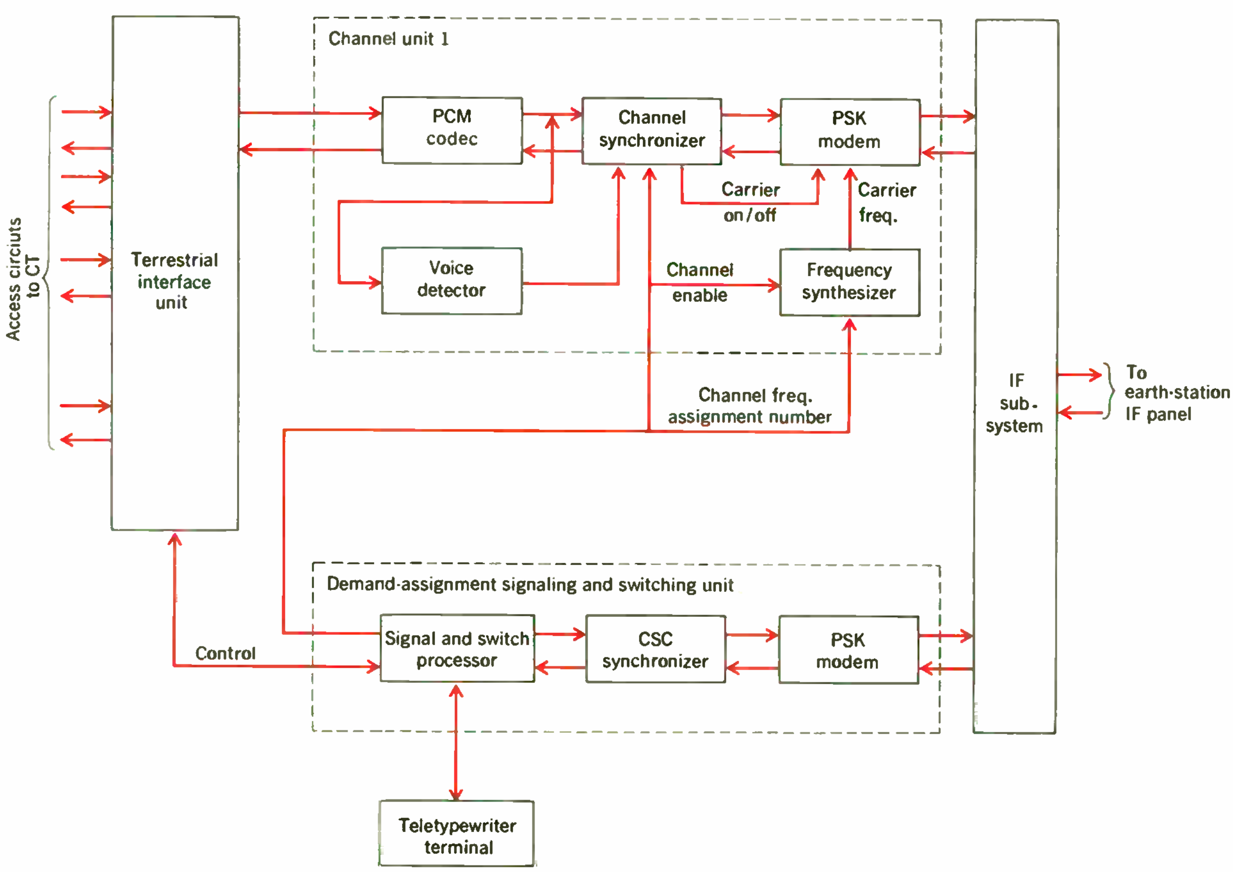
Characteristics of demand-assignment systems

Figure 6 illustrates the functional requirements for any type of demand-assignment system. There are three key functions:

1. *Multiple-access equipment.* The equipment necessary to provide access to the satellite—for example, FDMA and time-division multiple access (TDMA).
2. *Coding and modulation equipment.* The equipment necessary to modulate/encode and demodulate/decode the incoming and outgoing analog signals, respectively.
3. *Demand-assignment signaling and switching (DASS).* The equipment necessary to connect with existing terrestrial equipment and the other users in the demand-assignment pool through the routing channel. All channel requests, busy signals, terminating functions, system status reports, etc., are received and transmitted via the routing channel under the control of DASS.

The transit center is not part of the demand-assignment multiple-access system. It is normally located at major metropolitan areas and is connected to the earth station via cable or microwave links. Demand-assign-

FIGURE 9. SPADE terminal functional block diagram.



ment systems, through the flexibility of a stored programmed DASS,[8,9] can be designed to interface with all of the standard CCITT signaling systems, such as CT1, R-2, CT5, CT5 bis, and the forthcoming CT6.

Two demand-assignment multiple-access systems developed by Intelsat are described in Refs. 10 and 11. For the Intelsat system, where there are a large number of low-traffic sources desiring interconnectivity, a single-channel-per-carrier FDMA system is efficient. On the other hand, where there is a relatively lower number of accesses carrying medium-to-heavy traffic (12–60 circuits) per link, TDMA appears to offer a more efficient solution. As the number of channels per link increase, the effectiveness of the application of demand-assigned channels decreases.[12]

The particular demand-assignment system developed by Intelsat, which is to be implemented in mid-1971, uses FDMA and is described in the following section.

Principles of FDMA demand assignment[8]

The colored lines in Fig. 7 show a typical path allocation for a single voice link in a demand-assignment FDMA system. A subscriber requesting a call to a DA destination is connected at the transit center to one of a number of terrestrial access circuits terminating in DA equipment at earth station *A*. Note that, regardless of the ultimate call destination, any access circuit may be used, thus allowing more efficient use of the CT-to-earth-station link. In the earth-station DA equipment, the call is routed to modulating equipment and a particular carrier frequency pair (f_3 in this case) out of the pool of available network frequencies is assigned to the modulator and to the distant-end channel. For the duration of that connection frequency, f_3 remains assigned to that circuit and "is busied out" to other requesters. At the end of the communication, f_3 is returned to the network pool for subsequent reallocation. At earth station *F*, the call arriving on frequency f_3 is demodulated in the particular unit to which circuit f_3 has been assigned and is forwarded, via the CT, to the subscriber over the terrestrial link. Again, the receiving unit and its associated terrestrial connection are assigned arbitrarily, and only the allocated frequency determines which particular circuit the receiver will process. The important factor is that the equipment assignment is independent of either call source or destination.

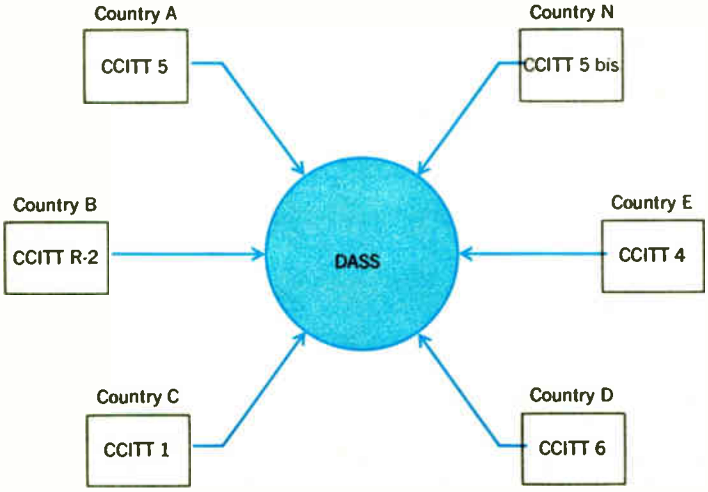
The method used to establish calls among users of the DA network will now be examined. Figure 8 illustrates a typical signaling and allocation method. Automatic dialing, which may originate either from the subscriber or from an operator at the CT, is assumed. The colored dots show the signaling proceeding from the subscriber through the CT to earth station *A*. There the DASS unit routes the signaling via a common signaling channel (CSC), operating in a time-division multiple-access broadcast mode, to all participating earth stations. At the initiation of this routing, station A requests allocation of a particular frequency for this call from a busy–idle table of frequencies continually updated via the CSC. Station *C*, for which the call is destined, monitors the common signaling channel and notes the arrival of station *A*'s request and the allocated frequency. If no other station has requested that frequency before station *A*, station *C* transmits an acknowledgment of the call request, assigns the frequency to the channel equipment, and proceeds to verify continuity and establish ultimate connection to the subscriber via the CT.

Station *A* monitors the common signaling channel between the time it requests a frequency and the time it receives its own request (about 240 ms). If, during that time, another station had requested that frequency, both stations *A* and *C* would register a busy to that allocation on their busy–idle tables and station *A* would initiate a new request. To minimize conflicting requests for simultaneous frequency assignment, each station chooses randomly from a list of available frequencies stored in DASS.

Note on the time scale of Fig. 8 that access to the common signaling channel occurs once every 50 ms for each station and the access duration is 1 ms. Station *A*, therefore, may initiate other calls while awaiting its own and station *C*'s response.

Having established the connection between the subscribers using stations *A* and *C*, the DASS and CSC proceed with processing of other incoming and outgoing requests. At the end of the call, the disconnect signal is broadcast over the CSC and all stations note that this frequency is again available for assignment. The DASS unit may also note the duration of the allocation and other

FIGURE 10. Signaling systems SPADE can interface.



IV. Intelsat IV transmission characteristics (global beam)

RF Bandwidth per Accessing Carrier, MHz	Channels per Carrier	Total Accesses per Transponder	Total Channels per Transponder
FDM/FM/FDMA (multichannel carriers):			
2.5	24	14	336
5	60	7	420
10/5*	132	4	456
36	900	1	900
PCM/PSK/FDMA (SPADE):			
0.045	1	800	800

* Three carriers at 10 MHz and one carrier at 5 MHz.

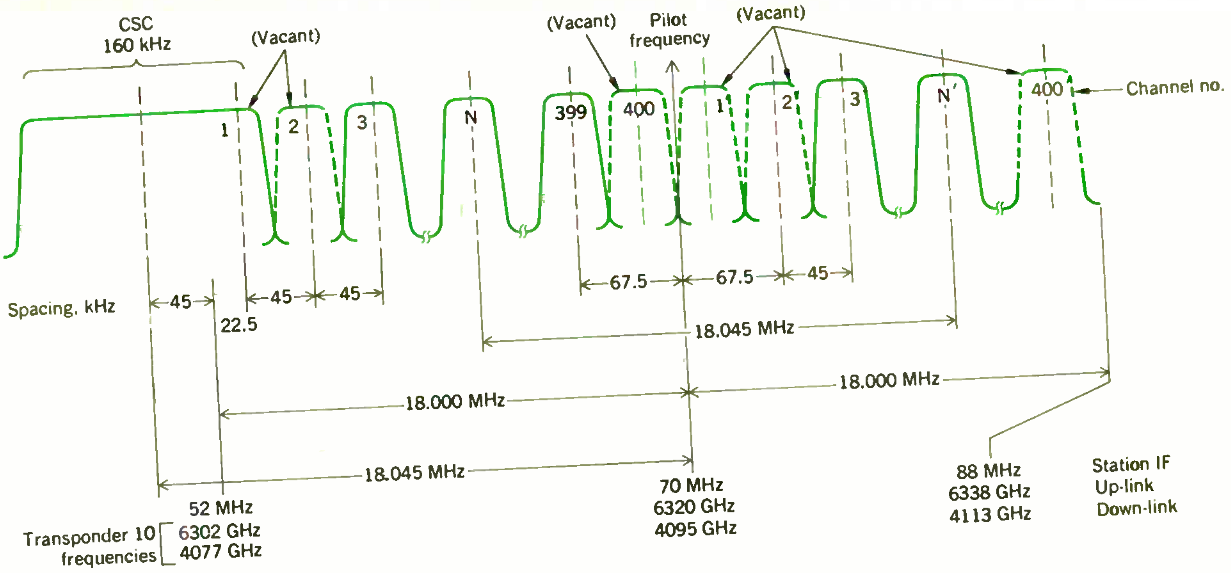


FIGURE 11. SPADE multichannel frequency-allocation spectrum.

vital statistics pertinent to each call.

This DA/FDMA method, which at present is being implemented by Intelsat, is known as the SPADE system. (The term SPADE is derived from *s*ingle-channel-per-carrier, *p*ulse code modulation; multiple-*a*ccess, *d*emand-assignment *e*quipment.) It is particularly timely in view of the rapid increase in the number of earth stations accessing Intelsat satellites, and the associated evolution of new traffic patterns. Further, it interfaces easily with existing earth-station facilities and is compatible with present methods of operations.

The SPADE demand-assignment system

Some of the technical characteristics of the SPADE system are summarized in Table III. This is a single-channel-per-carrier FDMA system providing demand assignment without central control. Pulse code modulation (PCM) is used for channel encoding and four-phase coherent phase-shift keying (PSK) is used for modulating each carrier. Other significant parameters of the system, including the characteristics of the common signaling channel, are also shown in Table III.

Figure 9 is a block diagram of a typical installation. Telephone circuits from the local transit center are linked to the SPADE terminal via the terrestrial interface unit. This equipment provides the interface for call signaling as required to initiate, supervise, and terminate all calls systematically. When a call request is received, the DASS unit automatically selects a frequency pair from the pool of available frequencies and alerts the destination station of an incoming call and the frequency assignment for response. All DASS units utilize the signaling information disseminated by the CSC to update a channel table such that the frequencies just assigned are unavailable for new calls.

The frequency selected is provided to the channel unit by means of a frequency synthesizer that is capable of generating any of the 800 discrete pool frequencies using digital codes provided by DASS. This frequency is used both for the outgoing carrier and the received-signal local oscillator. Channel pairings are based on the common use of the synthesizer for received and transmit signals.

Upon turn-on of the modem, the DASS unit conducts a two-way circuit continuity check. Once the call has been established, the analog signal received by the channel unit is sent to a PCM codec, which transforms it to a digital signal for outgoing transmission and transforms returning signals from digital to analog form.

The content of the voice channel coming from the CT is detected by a voice detector, which is used to gate the channel carrier on or off. This conserves satellite power as a function of talker activity. The digital bit streams in and out of the voice encoder/decoder are synchronized by the transmit–receive synchronizer, where timing, buffering, and framing functions are performed. The PSK modem modulates the assigned carrier frequency with the outgoing bit stream and coherently demodulates the incoming bursts by recovering carrier and bit timing associated with the received signals. The modulated carriers, both outgoing and incoming, are passed through a common intermediate-frequency (IF) subsystem, which interfaces with the earth station up- and down-converters at IF. The carrier used for the CSC modem is also passed through the IF subsystem.

When the call is completed, a control signal from the CT allows DASS to return that circuit to the frequency pool for reassignment. This information is passed to all stations via the CSC.

The design of the SPADE system has taken into account the fact that a variety of signaling and switching systems will interface with each SPADE terminal, as illustrated in Fig. 10. The terrestrial interface unit and the DASS in the SPADE terminal provide the required "universal interface" to each of these CCITT signaling and switching systems. Communications between countries having different signaling and switching systems can therefore be established using the SPADE DASS to provide compatibility.

System capacity with Intelsat IV satellites

The Intelsat IV satellite consists of 12 independent transponders. The allocated satellite frequency bandwidth

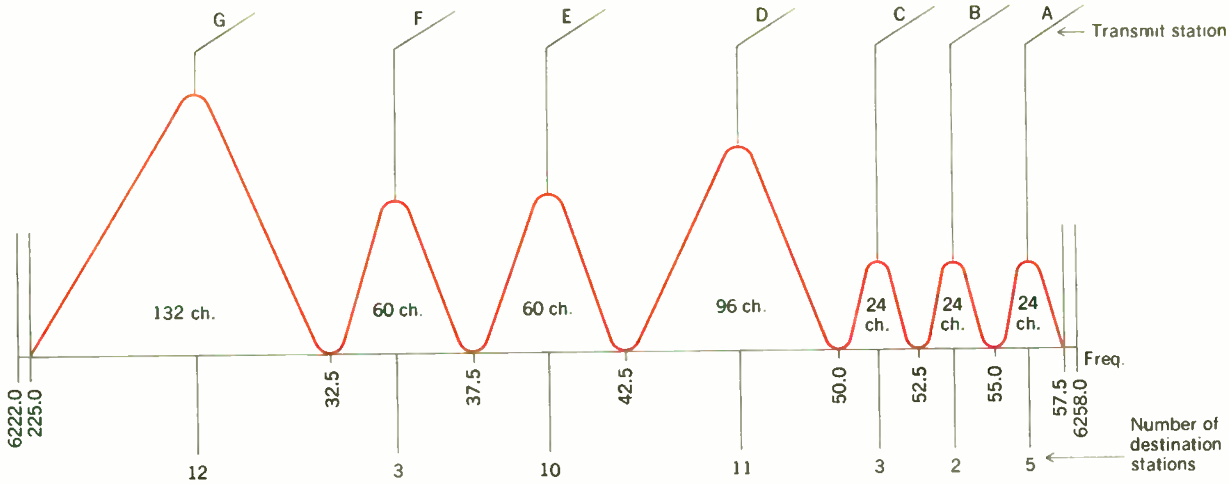
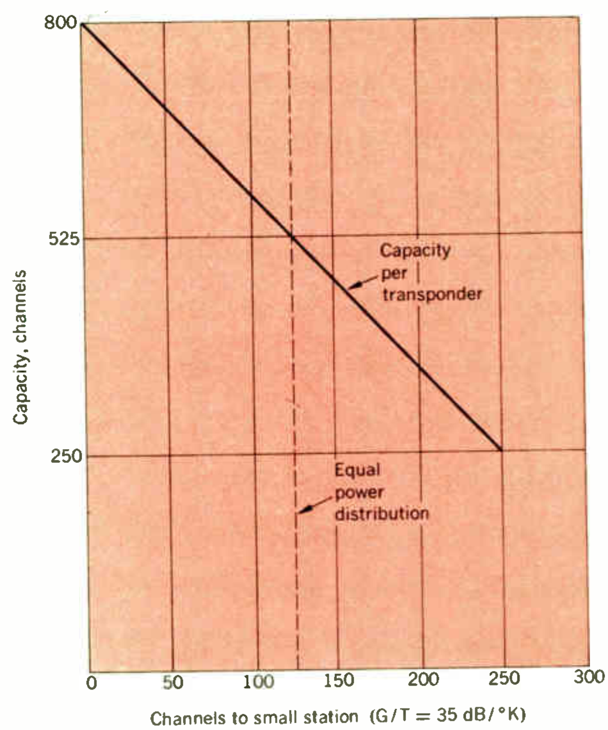


FIGURE 12. Typical FM/FDMA frequency-allocation spectrum.

FIGURE 13. Transponder capacity in mixed-size earth-station network.



of 500 MHz at 6 GHz and 4 GHz is distributed equally to each transponder so that each has an RF bandwidth of 36 MHz.

The SPADE system has been assigned to transponder 10, which lies between 6.302 and 6.338 GHz for reception and 4.077 and 4.113 GHz for transmission. Of the 12 transponders, eight can be selected to transmit either via narrow-spot beam antennas or a global coverage antenna. The four others are permanently assigned to the global earth coverage satellite antenna. Transponder 10 falls into the latter category.

The remaining transponders are accessed using conventional FDM/FM/FDMA multidestination, multichannel carriers. Intelsat has standardized these carriers to different sizes by specifying both the occupied RF bandwidth and the number of 4-kHz channels per accessing carrier. Some of these standard sizes are shown in Table IV. Since the smaller-capacity carriers are less efficient of bandwidth, the capacity of the transponder drops as the number of accesses increase. Thus, in Table IV, 14 accessing carriers, each occupying 2.5 MHz and each carrying 24 channels, fill the 36-MHz transponder bandwidth with a total of only 336 channels. Although various-size carriers can be transmitted from the same earth station, each carrier constitutes an access.

For SPADE, using PCM/PSK/FDMA, each RF carrier accommodates a single channel and occupies a bandwidth of 45 kHz. Thus, the 36-MHz bandwidth can be filled with 800 individual carriers, each representing an access to the satellite even though, again, several carriers emanate from the same earth station.

Since each carrier is associated with a voice channel, it need not be transmitted unless voice is present on that channel. Thus, there is sufficient satellite power to support 800 voice channels, which are on or off as a function of talker activity. This feature is not to be confused with time-assignment speech interpolation (TASI) where an idle channel is actually reallocated to another user during pauses in speech. In SPADE, the channel allocation remains fixed for the duration of the call, but during that call the carrier is turned on and off by talker activity to conserve satellite power.

Figure 11 shows the distribution of SPADE carriers over the 36-MHz band of transponder 10 and the corresponding band centered at the 70-MHz IF interface between the SPADE terminal and the earth station up and down frequency converters. At the lower end of the band, space is allocated for the common signaling channel and a system pilot is located at band center.

By contrast, Fig. 12 shows global beam transponder 8 being accessed by three 24-channel carriers, two 60-channel carriers, one 96-channel carrier, and one 132-channel carrier, resulting in a capacity of 418 channels. The number of destinations for each carrier is also given in this illustration.

The capacity of 800 channels[14] shown in Table IV assumes that the operating network is accessed by standard earth stations—i.e., earth stations with gain-to-temperature ratios (G/T)* of at least 40.7 dB/°K. At the expense of capacity per transponder, it is possible to operate the SPADE system in a network of mixed earth-station sizes. Figure 13 is a plot of the channels destined for small stations with G/T of 35 dB/°K. It can be seen from this illustration that if 125 channels are destined for the small earth stations, then the total network capacity is reduced to 525 channels. This represents the point at which half the available satellite power is being used to service the channels destined to the smaller stations and the remaining half is used to service channels terminating at the large standard stations. This curve is presented to illustrate the effect on network capacity occasioned by the introduction of small earth stations.[15] The main parameter being varied is the satellite-to-earth-station carrier power. All carriers destined to small stations must exceed the level of carriers destined to large stations by the difference in G/T between the large and small terminals. At the time the frequency is assigned for a call, each transmitting station is apprised of the size of the destination terminal and the level of the outgoing carrier is set accordingly. The implementation of the SPADE system, which allows the use of a single channel per carrier together with the DASS assignment feature, makes the penalties of mixing earth-station sizes less severe than for presently known frequency- or time-division multiple-access methods.

Conclusion

The technique of demand assignment will enhance the usefulness of satellite communications networks. In particular, it is advantageous for interconnection of countries requiring relatively few channels and for overflow traffic from medium preassigned trunks. The SPADE system, which is the specific system that Intelsat plans to use, is more efficient of power and bandwidth per channel than present methods.

* Gain-to-temperature ratio is the figure of merit applied to the receiving sensitivity of a given earth station without reference to its specific size or noise figure. Intelsat has ruled that the minimum G/T for a standard earth station in its network shall be at least 40.7 dB/°K.

This article is based upon work performed at Comsat Laboratories under Corporate sponsorship and under the sponsorship of the International Telecommunications Satellite Consortium (Intelsat). The views expressed herein are not necessarily those of Intelsat.

The authors gratefully acknowledge the assistance of the many members of the technical staff at Comsat Corporation during the course of this program. Further acknowledgment is due to Intelsat, under whose auspices the system was developed, and to the delegates of the various SPADE working groups, whose diligence contributed notably to the successful embodiment of an R & D concept into an operational system.

REFERENCES

1. Votaw, M. J., "The Early Bird project," *IEEE Trans. Communications Technology*, vol. COM-14, pp. 507–511, Aug. 1966.

2. Murphy, C. G., "A Syncom satellite program," Paper 63-264, presented at the AIAA Summer Meeting, Los Angeles, Calif., June 1963.

3. Berman, A. L., and Mahle, C. E., "Nonlinear phase shift in traveling-wave tubes as applied to multiple access communication satellites," *IEEE Trans. Communication Technology*, vol. COM-18, pp. 37–48, Feb. 1970.

4. Lutz, S. G., "A traffic study of multiple access satellite communication based on an Atlantic system model," Hughes Research Rept. 376, Aug. 1967.

5. CCITT/CCIR, "General plan for the development of the inter-regional telecommunication network (Mexico 1967-1970-1975)," ITU, Geneva, 1968.

6. Report of the Atlantic Region Traffic Subgroup Meeting, Washington, D.C., May 1969, ICSC Doc. 39-32.

7. Anguera, F., "Traffic considerations for demand-assigned telephone service through a satellite system," presented at the International Teletraffic Congress, Mexico City, Apr. 1967.

8. Dill, G., and Shimasaki, N., "Signaling and switching for demand-assigned satellite communications," *Proc. Intelsat/IEE Conf. Digital Satellite Commun.*, pp. 297–307, 1969.

9. Dill, G., and Shimasaki, N., "The terrestrial interface at SPADE terminals," Paper 70-413, presented at the AIAA 3rd Communications Satellite Systems Conference, Los Angeles, Calif., Apr. 6–8, 1970.

10. Werth, A. M., "SPADE: A PCM FDMA demand assignment system for satellite communications," IEEE Paper 70-CP-423-COM, presented at the 1970 International Conference on Communications, San Francisco, Calif., June 1970.

11. Schmidt, W. G., *et al.*, "MAT-1: Intelsat's experimental 700-channel TDMA/DA system," *Proc. Intelsat/IEE Conf. Digital Satellite Commun.*, pp. 428–440, 1969.

12. "BTR special study for Intelsat on demand assignment techniques for satellite communications," British Telecommunications Research Ltd. Rept. 97/69/04, Jan. 1969.

13. Morita, M., Tukami, T., and Yamato, S., "STAR system. I—General description," *Nippon Elec. Co. Res. Develop. J.*, Oct. 1966.

14. McClure, R. B., "The effect of earth station and satellite parameters on the SPADE system," presented at the IEEE Conference on Earth Station Technology, London, Oct. 1970.

15. Cacciamani, E. R., "The SPADE concept applied to a network of large and small earth stations," Paper 70-420, presented at the AIAA 3rd Communications Satellite Systems Conference, Los Angeles, Calif., Apr. 6–8, 1970.

John G. Puente (SM) received the B.E.E. degree in 1957 from the Polytechnic Institute of Brooklyn and the M.S.E.E. degree in 1960 from the Stevens Institute of Technology. From 1957 to 1960 he worked on the design and development of pulse instrumentation and radar equipment for Stavid Engineering, Plainfield, N.J., and ITT Laboratories, Nutley, N.J. In 1960 he joined IBM's Communications Department in Rockville, Md., where he was manager of communication device development, responsible for the development of data transmission equipment. He is now manager of the Communications Processing Laboratory, Comsat Laboratories, Communications Satellite Corporation, Washington, D.C., where he is responsible for the research and development of new satellite communications systems to handle voice, data, and television.

Andrew M. Werth (M) received the B.S. and M.S. degrees from Columbia University in 1955 and 1961 respectively. From 1955 to 1957 he served as an electronics officer in the U.S. Air Force. From 1957 to 1959 he worked on the development of servomechanisms for an airborne radar at Stavid Engineering, Plainfield, N.J. In 1959 he joined ITT Laboratories, Nutley, N.J., where he developed digital data links for the Lobar radar system and performed studies on satellite communications systems. He joined Comsat in 1964 and initially was responsible for telemetry, tracking, and command systems for Intelsat satellites. At present he is manager of the Modulation Techniques Branch at Comsat Laboratories, where he is serving as program manager for the SPADE demand-assignment system. He is also responsible for the development of PSK modems and applications of digital techniques to satellite communications.

Assessment, transfer, and forecasting of technology

The social implications of our rapidly developing technology must be brought to the attention of public policy makers if we are to achieve our social goals in a rational manner

Walter H. Kohl Consultant

The United States is finally facing up to the fact that we can no longer tolerate the uncontrolled application of technology. We are now seeing the results of what has seemed almost a national program of conspicuous consumption—depletion of our natural resources and the deterioration of our environment. Hopefully, the trend can still be reversed by planning and control. Over the years, a substantial body of literature has accrued on this vital topic, and an attempt is made here to provide the interested student with a guide.

Various studies have been conducted in the past on the impact of technology on the established patterns of society, but many of these were focused on particular industries, showing how technical innovations made some of them obsolete and created new ones in their place, thereby causing shifts in the demand for labor in certain localities. On the whole, such appraisals have been producer-oriented, rather than emphasizing the interests of the consumer. A review by Allen[1] lists the contributions made to this subject until 1958, and Kranzberg[2] shows, in a more recent review, that many of these concepts have their roots in ancient times.

During the past decade, the realization has grown, in both the public and private sectors of society, that the uncontrolled stimulation of demand for goods and services has dominated our economy too long. Ever-increasing production and concomitant consumption merely to satisfy the profit motive tend to have lower intrinsic value. It is now realized that not only will such uncontrolled production exhaust valuable and scarce resources but also that it contributes to the deterioration of the environment by causing pollution of air and water, accumulation of waste, health hazards, congestion of the traffic lanes on the ground and in the air, and the many other ills that now cry out for correction and control.[3]

A number of books have been published,[4-14] two journals on technological forecasting were launched in 1969,[15,16] and several symposiums have been held. A Seminar on Management of Technology Transfer[17] took place at U.C.L.A. in March 1968. The Working Group on Research Management of the Military Operations Research Society held its 22nd Symposium in Monterey in December 1968, at which the impact of forecasting techniques on R & D planning was considered at length.[18] A two-day Seminar on Technology Assessment was held in September 1967 by the Subcommittee on Science, Research, and Development of the Committee on Science and Aeronautics of the House of Representatives.[19] Further hearings on technology assessment were held before the House Subcommittee in November and December 1969.[2] (Former Representative Emilio Q. Daddario, until recently chairman of the Subcommittee, requested the National Academy of Sciences and the National Academy of Engineering in 1967 to examine the problems associated with an assessment of technology, and separate panels with a most distinguished membership were appointed. The reports of these panels[20,21] are now available, and a review article[22] by the chairman of the NAS committee and one of its members, puts the problems into focus. A third contract was placed with the Library of Congress and resulted in the comprehensive report prepared by F. Huddle.[23]) In Europe, the Organisation for Economic Co-Operation and Development (OECD) ordered a worldwide review of the methods used in technology forecasting that resulted in the comprehensive report by Jantsch[24] in which 413 references are listed.

Terminology

As indicated by the preceding overview of the literature, a general awareness of the need for a more formalized approach to technology assessment, transfer, and forecasting has developed only in recent years. These disciplines are still in a state of flux, and even the terminology has not been clearly established in all cases. As recently as 1968, for example, Cetron and Johnson[25] point out that technology assessment is not official jargon and is not found in the tables of contents and indexes of texts on management, nor is it identified and found in the general literature of management or in official planning, programming, and policy documents of the government agencies. Nevertheless, the term was firmly established after its first use for the aforementioned seminar and report.[19,26] A NASA bulletin[26] entitled "Assessing Technology Transfer" accidentally contributes to the confusion in the use of terms, but it must be borne in mind that this publication was issued at a time (1966) when none of these terms was used in any precise manner. NASA's Office of Technology Utilization aims at the transfer of space-oriented technology to civilian applications. Since a technology, in the generally accepted sense, does not exist unless it is being utilized, technology utilization is not a very happy choice of words; nevertheless, utilization and transfer are often used as synonyms.

For the purpose of our discussion, it is also necessary to define "technology" more broadly and not restrict it to the interaction of things. According to Gruber and Marquis,[6] "technology may be defined as the means or capacity to perform a particular activity," thus giving it a wider frame than that offered by Ayres,[7] who calls it "the systematic application of organized knowledge to practical activities, especially productive ones."

Carpenter (Ref. 2, p. 362) sees technology assessment as the "final step in a long sequence which could be termed the socialization of science and engineering." Similarly, Richman (Ref. 2, p. 491) remarks that one must not conceive of technology assessment in terms of technology alone. Rather, he says, it is important to define it and to execute it in terms of the overall social—including moral and ethical—political, and economical situation in the U.S. and the world. Technology merely becomes the tool to achieve the objectives set by the nation.

According to Bauer,[10] technology comprises the practical media by which man interacts purposefully with his environment to meet his needs. Its components may be tangible or intangible. Technology includes "hard" tools, such as the plow, the drill press, and the airplane, and soft tools, such as production scheduling procedures, double-entry bookkeeping, and computer programs. In either case, he tells us, the essence of technology is cognitive, not material.

Schon[27] defines technology as any tool or technique, any product or process, any physical equipment or method of doing or making by which human capability is extended.

In the sense of these broader definitions, a technology base may well exist without its being utilized; economic considerations, such as insufficient demand or adverse assessment, may militate against its use.

Assessment and control

The task of regulating commercial activities presents many difficult problems that have been discussed at length in the literature, and particularly in the Hearings before the Subcommittee of the Committee on Science and Astronautics of the House of Representatives.[2,19,20] Although some action by the government seemed indicated, it was not at all clear what form it should take, and whether a Control Agency should be affiliated with the Executive Office or with the Congress, or both (Ref. 2, pp. 85–111, 135). Weinberg (Ref. 2, p. 185) suggested that the large national laboratories, such as those of AEC, NASA, and DOD, could serve as centers of technology assessment in their respective fields and help in the development of new approaches toward the solution of environmental problems. Michaelis[28] proposed a "Technology Control Board" in which the management skill and technical expertise of these laboratories would be combined with those of industry, labor, and government to bring about a unified systems approach to the solution of the many problems that face our society.

These earlier thoughts and subsequent discussions have been crystallized in a bill submitted to the Congress by E. Daddario in which it is proposed to create an Office of Technology Assessment (OTA) under the Comptroller General of the United States but not subjected to the jurisdiction of the General Accounting Office. A concise summary of the objectives of the proposed OTA has been presented by Lear,[29] who lists the following findings and declarations that are pertinent to technology assessment:

1. That emergent national problems, physical and social, are of such a nature and are developing at such an unprecedented rate as to constitute a major threat to the security and general welfare of the United States.

2. That these problems are largely the result of and are allied to the rising pressure of expanding population, the rapid consumption of natural resources, and the erosion of the natural and social environment of the human species.

3. That widespread application of modern technology, existing and evolving, is a crucial element in the situation and either is or can be a pivotal influence in causing and solving the problems enumerated.

4. That the Congress is not now equipped to obtain adequate, independent, and timely information concerning the potential application or impact of technology.

5. That it is therefore imperative for the Congress to equip itself to assess technological developments, predict the probable consequences, and act to protect the people of the United States against adverse results while assuring the nation of all possible benefits.

The objectives of technology assessment may further be put into focus by quoting Huddle (Ref. 23, p. 481):

"Assessment includes forecasting and prediction, retroactive evaluation, and current monitoring and analysis. Measurements involve noneconomic, subjective values as well as direct, tangible quantifications. Above all, assessment requires that catastrophic consequences of each proposed new technology be foreseen and avoided before the new technology becomes entrenched in the socio-economic complex of human organization."

These findings and the proposals to set up a Board of Technology Assessment were to be incorporated in a Legislative Reorganization Bill (H.R. 17654) that was passed by the House of Representatives on September 17, 1970, and by the Senate in mid-October, but Daddario's motion to attach the OTA to an expanded Legislative Reference Service, to be known as the Congressional Research Service, was ruled out of order.[30] The likelihood of early legislative action to bring into being an OTA is therefore remote at the time of this writing. A similar fate may well befall Senator Caleb Boggs' proposal to create "A National Materials Policy,"[31] which is, of course, closely related to technology assessment. (The writer was a member of the *ad hoc* committee formed at Senator Boggs' request.[31])

The overriding question is how technology that is assessed as being harmful to the public can be effectively controlled or avoided when powerful interests lobby for its continued use or its introduction. Although the unscrupulous use of technology that is blatantly harmful should be suppressed by legal action, it is difficult in some cases to establish a clear-cut consensus, and the danger exists that precipitate action is taken before all factors have been carefully considered by all affected parties.

Recently (October 26, 1970), President Nixon issued an order restricting the use of leaded gasoline in federally owned cars and urged the state governors to do likewise in order to minimize air pollution. According to the *New York Times* of October 27, the federal government buys 0.5 percent of all gasoline purchased in the United States.

In the most recent issue of *Lead*,[32] the Lead Industries Association, Inc., on the basis of extensive investigations, takes a strong stand against proposals for the arbitrary elimination of lead additives to gasoline, since the harmful effect of such additives had not been clearly established. A U.S. Bureau of Mines study published in May (Report 7390) concluded that unleaded fuels increased eye-irritating, smog-producing emissions by as much as 25 percent over those of leaded fuels.

On the other hand, selfish motives often have prevailed and resulted in pollution of lakes and streams, for example, in spite of existing ordinances against such abuse, or in the manufacture and sale of hazardous products, or the performance of large-scale experiments that were known to have harmful consequences.

To be effective, control functions must be supported by the public at large, and all segments directly or indirectly affected by contemplated regulations be given an adequate hearing. As Wormuth[33] and others have pointed out, "the power to make rules should not be entrusted to those who will be governed by those rules." The Food and Drug Administration has been accused in the past of having had friendly relations with the pharmaceutical industry; the Interstate Commerce Commission has shown favoritism toward the railroads and discriminated against pipelines and water carriers; the Atomic Energy Commission, which promotes the construction of nuclear power plants, at the same time is entrusted with the establishment of radiation-tolerance levels; the Federal Aviation Agency, through its involvement in the promotion of the supersonic transport, has ties with the aircraft industry; the Department of Transportation finds itself coupled to the interests of those building highways and motor vehicles.

Technology transfer

The principles laid down for the technology assessment also apply to technology transfer and utilization, which are closely related. Great economic benefits often result from the transfer of a technology that has been developed in one field and is found useful in another, but frequently modifications are necessary before such utilization can become fully effective. The potential usefulness of a process, a device, or a management technique for another field of application is thus not self-evident and may have to be intuitively recognized. This requires, first, an awareness of the existing technology, and then the creative insight to realize its potential for transfer. If the occurrence of these conceptual events were left to chance, not much technology would ever be transferred.

Many people[26,34] have claimed that the transfer of experts is the most effective mechanism of technology transfer, assuming that these experts carry their knowledge from one field to another. However, the likelihood that such carriers enter industries far removed from their specialty is rather remote, and the question remains whether they are the innovative type. Much is therefore left to chance in this approach.

Managements of progressive corporations have become aware of the need for systematically cultivating technology transfer by forming a task force within their organization and giving all necessary support to their "search and innovation mission." The steps involved in the transfer process have been well described in the NASA publication[26] on this subject; they read as follows:

1. Finding the technical information.
2. Screening out that which has current relevance for possible special emphasis, but not abandoning what remains for it may have unrecognized value.
3. Organizing it in a manner that permits its rapid and efficient retrieval for a variety of potential users with different languages, interests, and orientations.
4. Bringing relevant parts of it, on a selective basis, to the attention of a variety of potential users.
5. Arranging for seemingly unrelated pieces originating in separate areas to be fitted together.
6. Encouraging its use on the basis of its value.
7. Relating it to ongoing efforts that may enhance its value.
8. Organizing it so that it can not only be called out to meet specific defined needs, but also be a source of ideas to the technical man "browsing" through it.
9. Permitting the full inventory to be examined in a way to allow the discovery of areas of knowledge convergency or potential breakthrough areas and areas of need.
10. Establishing and maintaining an economic and social environment conducive to change.

Behavioral scientists have addressed themselves to the question of how the organizational climate influences human motivation, and McClelland[34] discusses the role of achievement orientation in the transfer of technology. Psychologists refer to the "need for achievement" by the term "n-Ach" and point out that this factor is more common among innovator-entrepreneurs in the business world than it is among professionals. This well may be one reason why technologies developed at universities and in government laboratories have not found their way to the public domain as much as promoters of technology transfer programs would like. Or could it be that professionals have other channels that lead more easily to financial gain?

Technology transfer could be made more attractive if government agencies, such as NASA, AEC, DOD, and others who produce a substantial body of new technology in the pursuit of their missions, were to issue grants to industries interested in exploring new technology, so that the initial costs would be covered. The fact is often overlooked that the cost of evaluating and finally adopting a new technology is considerable and that in some cases, particularly in small industries, it may be a hurdle too difficult to overcome. If a development grant were issued and a successful transfer achieved so that an innovation were to result, the agency could collect a royalty from sales, or continued use, and thus actually share in the profits. Practical difficulties in realizing such a scheme no doubt exist, particularly since grants would have to be made available to a number of applicants at the same time, but such obstacles could be overcome.[27]

A much more deep-seated reason for the ineffectiveness of present-day technology transfer programs also exists. Bauer[10] and his coauthors at the Harvard School of Business Administration point out that the methodology of transfer had never been subjected to detailed analysis until they began their six-year study under the sponsorship of NASA. It is simply beyond the competence of a mission-oriented organization to pursue the many complex interrelationships that govern the transfer mechanism. Doctors[35] discusses the role of federal agencies in technology transfer and analyzes in detail the NASA Technology Utilization Program in which he was involved for a few years. Although the NASA effort is extensive and unique among federal agencies, Doctors points out that "the technology utilization program was founded primarily in response to political pressures and has continued to be used as a device for partial justification of NASA R&D funding, rather than a technical project in its own right. Therefore, instead of being an experimental program, as so often advertised, it is in many ways an expanded library operation with regional branch libraries located in many areas that have a low interest in aerospace technology." He goes on to say that "...hard as an agency may strive to establish what it calls an experimental transfer program, the need to produce a given set of visible results, relatively rapidly, militates against an objective, scientific approach."

The Office of the State Technical Services (OSTS), established in 1965 under a grant of the U.S. Department of Commerce and matching grants from the 47 states participating in the program, was unsuccessful for the same reasons, as long as it relied primarily on the passive dissemination of available information. However, it was very successful in those areas where individual field representatives searched out the problems that faced small industries and established contact with experts in federal agencies and universities and with consultants, who could then pursue the problem. In many cases, otherwise defunct, small enterprises have been revitalized and brought into the national economy. Unfortunately, OSTS was discontinued when the Senate-House Conference Committee on Supplemental Appropriations refused to allocate additional funds on December 20, 1969. Hamilton[36] describes the circumstances that led to this decision, in spite of a basically favorable report by Arthur D. Little, Inc.[37]

Many reasons may be cited for the unfortunate demise. The program was inadequately funded to begin with—$5 million per year compared with $100 million or more allocated to the Agricultural Extension Service. A much larger budget had originally been proposed for OSTS but, unlike AES, it was not backed by federal programs of research relevant to specific industrial needs. It also suffered from a lack of coordination between a number of federal programs aiming at technology transfer, such as those operated by NASA, AEC, and the Commerce Department. Finally, there had developed a personal feud between the Head of the House Subcommittee, Rep. John J. Rooney, and the former Assistant Secretary of Commerce for Science and Technology, J. Herbert Holloman, who was the enthusiastic sponsor of OSTS at its creation.

Glaser,[38] who had directed the Arthur D. Little survey, recently proposed a Federal Transfer Program that would take the place of OSTS and hopefully avoid some of the pitfalls to which the former organization succumbed. Meanwhile, some states are independently continuing the program.

Technology forecasting

Forecasting is of course widely practiced in everyday life. To give an example, most people plan their vacation trips on the basis of available resources and choose the time so that it fits into vacation schedules of others at their place of employment. On the basis of these given data, a person can forecast with reasonable accuracy how much money he is going to spend and how long he will be away, although unforeseen factors may intervene. This type of forecasting—an extension of given data into the future—is called "exploratory forecasting." On the other hand, if the ordinary working man makes a long-range plan to take a trip around the world, many uncertainties exist. After appraising the cost he may find that, in addition to embarking on a consistent saving program, he may have to take extra work to come anywhere near a realization of his goal. Thus, the future determines his planning for the present and his action. This sequence of events is described by the term "normative forecasting."

On a larger scale, economic forecasts have been used for a long time as a basis for planning in government and industry. Future trends, such as those of the availability of natural resources, the growth of consumption, or the influence of developing foreign economies on price levels in the home market, are estimated on the basis of available statistical data that have determined a pattern in the past. Such economic forecasts are fraught with uncertainties, since they do not take into account unforeseen technological developments or changes in sociological patterns. What is needed, therefore, is a model of the total socio-economic-technical system and a study of its behavior in the presence of a number of perturbations; in other words, a complete theory that has been thoroughly tested. Needless to say, such a theory does not exist, but a number of methodologies have been developed that greatly facilitate long-range forecasting and that take speculations about possible futures into account.[4,5,7,24]

During the early 1960s, concern with this state of affairs increased among public administrators, both in the U.S. and elsewhere. James Webb, then the Administrator of NASA, initiated in 1961 at the Sloan School of Management of M.I.T. an extensive program on "The Management of Science and Technology," which has made valuable contributions through the years to the study of the methodology of technology assessment, transfer, and forecasting. Other government agencies, apart from NASA, and private foundations continue to support this work. An Interagency Task Force that addresses itself to the role of the federal government in technology forecasting was established in 1966 under the chairmanship of D. A. Schon, who has given an overview of this work in Ref. 39. The Technology Forecasting Institute, Inc., conducted three-day seminars in New York and Los Angeles in November 1969 and April 1970, respectively, with the title "The Methodology of Technological Forecasting: The Exploratory and Normative Techniques and Their Utility in R&D Planning." North and Pyke[40] have described the application of forecasting techniques in an electronics company. The broader aspects of institutional forecasting in the framework of "joint systems" of society and technology have been elaborated by Jantsch,[41] who points out that the institution, or company, will become part of a dynamical pattern and must envision its own change to encompass societal objectives.

The underlying concept of matching technology to social goals was further developed by Jantsch in a discussion of the restructuring of the university. The author calls for an overriding emphasis on socio-technological systems engineering in universities whereby the basic functions of education, research, and service are brought into balance. Universities will then be able to assume leadership in determining the future course of society, rather than remaining the tools of a fragmented technology.

Quite apart from these long-range objectives, it is obvious that educational institutions and professional societies will have to play an important role in establishing a favorable public climate for the support of proposed legislation by creating among their members an awareness of the social implications of science and technology. The program established by the Georgia Institute of Technology for the "Study of the Impact of Science and Technology," as described by Kelly (Ref. 2, p. 485) is a good example of projects that are being developed at a number of universities; there are now over 50 of them in operation across the U.S.

Implications for the future

Granting that we should no longer tolerate the uncontrolled application of technologies, both soft and hard, forecasting must be closely tied to the assessment of the future. This has vast implications for the decision-making process in the private and public sectors of society. It calls for the establishment of broad social goals, an answer to the question of how we envision our future, both as a nation and as human beings. Beer[42] and Mesthene,[43,44] among many others, have made a searching analysis of this basic problem. To quote Beer:

"The risk which faces us today is the probability that society will yet refuse to study the systemic generators of human doom and will disregard the cybernetic capability which already exists, competent to bring these many but interrelated forms of crisis under governance . . .

"At present, the most obtrusive outcome of the system we have is a gross instability of institutional relationships and of the economy. This cannot last. The society we have known will either collapse or it will be overthrown. In either case a new society will emerge, with new modes of control; and the risk is that it will be a society which no one actually chose, and which we probably will not like. I shall argue that we must use our science to detect the latent outcomes which will one day characterize the future of mankind. And let us so engineer our systems that their latent outcomes suit our social purpose. It is true that the outcomes cannot be fully determined, because there is noise (or shall we call it free will?) in the system. But a systemic design, taking due account of cybernetic laws, may be expected to produce behaviour which is predictable in terms of the overriding social need for stability."

The tools for a rational approach to the restructuring of society are only now beginning to be hammered out. A new interdiscipline has come into being under the name of policy sciences, and a journal by that name serves as its forum. The main features of policy sciences have been summarized by Dror,[45] in an announcement of the AAAS symposium on that subject, as follows:

1. Policy sciences are an interdiscipline, focusing on public policy-making.

2. Policy sciences are based on behavioral sciences and analytical approaches, relying also on decision theory, general systems theory, management sciences, conflict theory, strategic analysis, systems engineering, and similar modern areas of study. Physical and life sciences are also relied upon insofar as they are relevant.

3. Fusing pure and applied research, policy sciences are mainly concerned with improving policy-making on the basis of systematic knowledge and structural rationality.

4. Policy sciences, as with all applied scientific knowledge, are, in principle, instrumental-normative in the sense of being concerned with means and intermediate goals rather than absolute values. But policy sciences are sensitive to the difficulties of achieving "value-free sciences" and try to contribute to value choice by exploring value implications, value consistencies, value costs, and the behavioral foundations of value commitments.

5. Policy sciences emphasize metapolicies (that is, policies on policies), including modes of policy-making, policy analysis, policy-making systems, and policy strategies. While the main test of policy sciences is better achievement of considered goals through more effective and efficient policies, policy sciences as such do not deal with discrete policy problems, but do provide improved methods and knowledge for doing so.

Dror, who is a senior staff member of the Rand Corporation, presented a paper at that conference bearing the title "Prolegomenon to Policy Sciences: from Muddling Through to Meta-Policymaking." In a similar vein, Jantsch[46] dealt with the subject "From Forecasting and Planning to Policy Sciences." These titles by themselves suggest the importance of policy sciences in relation to the topics that have been discussed in this article, for which "Managing Modern Complexity" might well have been chosen as an alternative title.

REFERENCES AND NOTES

1. Allen, F. R., "Technology and social change: Current status and outlook," *Technology and Culture*, vol. 1, pp. 48–59, Winter 1959.

2. *Technology Assessment*. Hearings before the Subcommittee on Science, Research, and Development of the Committee on Science and Astronautics, Nov. 18, 24; Dec. 2–4, 8, 12, 1969. U.S. Government Printing Office, Washington, D.C., 1970. Appendix O contains the Proceedings of the Engineering Foundation Research

Conference on "Technology Assessment" held at Proctor Academy, Andover, N.H., Aug. 4–8, 1969 from which the following contributions are cited: R. A. Carpenter, "Technology assessment—a new responsibility," pp. 361–368; M. Kranzberg, "Historical aspects of technology assessment," pp. 380–388; D. M. Richman, Post-conference correspondence, pp. 489–491.

3. "The environment: a national mission for the seventies," *Fortune* (special issue), Feb. 1970.

4. Bright, J. R., ed., *Technological Forecasting for Industry and Government*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

5. Cetron, N. J., *Technological Forecasting*. New York: Gordon and Breach, 1969.

6. Gruber, W. H., and Marquis, D. G., *Factors in the Transfer of Technology*. Cambridge, Mass.: M.I.T. Press, 1969.

7. Ayres, R. U., *Technological Forecasting and Long-Range Planning*. New York: McGraw-Hill, 1969.

8. Rose, J., ed., *The Effect of Technological Advances on Environment, Life, and Society*. New York: Gordon and Breach, 1969.

9. Gordon, J., and Ament, R. H., *Forecasts of Some Technological and Scientific Developments and Their Societal Consequences*, Rept. R-6, The Institute for the Future, Middletown, Conn., Sept. 1969.

10. Bauer, R. A., *et al.*, *Second-Order Consequences*. Cambridge, Mass.: M.I.T. Press, 1969.

11. Traviss, Irene, and Burbank, Judith, "Technology and the polity," Harvard University Program on Technology and Society, Res. Rev. No. 4, 1969.

12. Traviss, Irene, and Silverman, Linda, "Technology and values," Harvard University Program on Technology and Society, Res. Rev. No. 3, 1969.

13. Jungk, R., *Technologie der Zukunft*. Berlin: Springer-Verlag, 1970.

14. Toffler, A., *Future Shock*. New York: Random House, 1970. (Contains extensive bibliography; see also listing of additional journals on p. 472.)

15. *Technological Forecasting*. New York: American Elsevier ($24/year).

16. *Futures*. New York; Iliffe-NTP, Inc. ($30/year).

17. "Selected papers from a seminar on the management of technology transfer, U.C.L.A., March 1968," *IEEE Trans. Engineering Management*, vol. EM-16, pp. 99–129, Aug. 1969.

18. "Selected papers by the Working Group on Research Management, 22nd Military Operations Research Symposium (MORS), Monterey, Calif., Dec. 1968," *IEEE Trans. Engineering Management*, vol. EM-16, pp. 137–248. Nov. 1969.

19. *Technology Assessment Seminar Proceedings*. Subcommitteee on Science, Research, and Development of the Committee on Science and Astronautics, House of Representatives, Sept. 21–22, 1967; U.S. Government Printing Office, Washington, D.C. 1967.

20. *A Study of Technology Assessment*. Report of the Committee on Public Engineering Policy, National Academy of Engineering, to the Committee on Science and Astronautics, U.S. House of Representatives, July 1969; U.S. Government Printing Office, Washington, D.C. 1969.

21. *Technology: Processes of Assessment and Choice*. Report of the National Academy of Sciences to the Committee on Science and Astronautics, U.S. House of Representatives, July 1969; U.S. Government Printing Office, Washington, D.C., 1969.

22. Brooks, H., and Bowers, R., "The assessment of technology," *Sci. Am.*, vol. 222, pp. 13–21, Feb. 1970.

23. *Technical Information for Congress*. Report to the Subcommittee on Science, Research, and Development of the Committee on Science and Astronautics, U.S. House of Representatives; prepared by the Science Policy Research Division, Legislative Reference Service, Library of Congress, Apr. 1969; U.S. Government Printing Office, Washington, D.C., 1969.

24. Jantsch, E., *Technological Forecasting in Perspective*. Organisation for Economic Co-operation and Development (OECD); OECD Publications, 2 rue André-Pascal, Paris, 1967. (Anyone interested in this field should not be without this volume.)

25. Cetron, M. J., and Johnson, J. N., "Technological forecasting in a dynamic environment," *IEEE Trans. Engineering Management*, vol. EM-16, pp. 190–222, Nov. 1969.

26. *Assessing Technology Transfer*. NASA SP-5067, Office of Technology Utilization, 1966; U.S. Government Printing Office, Washington, D.C.

27. Schon, D., *Technology and Change*. New York: Delacorte Press, 1967. (Schon suggests as a means of enhancing the "natural" process of technology transfer the rewarding of federal R & D contractors for reporting innovations by granting them a "royalty" payment when the innovations are used for governmental purposes, or by providing an industrial extension service.)

28. Michaelis, M., "Can we build the world we want?" *Bull. Atomic Scientists*, vol. 24, pp. 43–49, Jan. 1968.

29. Lear, J., "Predicting the consequences of technology." *Saturday Review*, pp. 44–46, Mar. 28, 1970. (This article describes the aims of the bill submitted by E. Daddario in which the creation of an Office of Technology Assessment is proposed; it also gives a summary of the historical background.)

30. Walsh, J., "Congress: Some progress toward putting house in order," *Science*, vol. 170, pp. 45–48, Oct. 2, 1970.

31. *Toward a National Materials Policy*. Report to the Committee on Public Works, U.S. Senate, by an *ad hoc* committee formed by the Legislative Reference Service, Library of Congress, at the request of Senator J. Caleb Boggs, Apr. 1969; U.S. Government Printing Office, Washington, D.C., 1969.

32. Kimberley, J.L., "Controversy lead in gasoline," *Lead*, vol. 33, no. 2, pp. 5–7, 1970. (The author is executive vice president of the Lead Industries Association, Inc.)

33. Wormuth, F. D., "Government and science," *Center Mag.*, vol. 3, pp. 41–46, Mar. 1970.

34. McClelland, D. C., "The role of achievement orientation in the transfer of technology." (Chapter 4 in Ref. 6.)

35. Doctors, S. I., *The Role of Federal Agencies in Technology Transfer*. Cambridge, Mass.: M.I.T. Press, 1969.

36. Hamilton, A., "State Technical Service: Congress swings the axe," *Science*, vol. 166, pp. 1606–1608, Dec. 25, 1969.

37. "Program evaluation of the Office of Technical Services," Report by Arthur D. Little, Inc., to the U.S. Department of Commerce, Office of Program Planning (Contract 9-35335), Oct. 1969.

38. Glaser, P. E., "Towards a federal technology transfer program," a presentation before workshop session II of the Eastern Regional Conference on Science and Technology for Public Programs, M.I.T., Cambridge, Mass., Apr. 2, 1970.

39. Schon, D. A., "Forecasting and technological forecasting," a condensed version of the "Working Papers of the Commission on the Year 2000," *Daedalus*, vol. 96, no. 3, pp. 759–770, Summer 1967.

40. North, H. Q., and Pyke, D. L., "Technological forecasting in planning for company growth," *IEEE Spectrum*, vol. 6, pp. 30–36, Jan. 1969.

41. Jantsch, E., "New organizational forms for forecasting," *Technology Forecasting*, vol. 1, pp. 151–161, 1969. (This article cites 18 additional references that should be of interest to those charting the future course of an organization.)

42. Beer, S., "Managing modern complexity," Panel on Science and Technology, Committee on Science and Astronautics, U.S. House of Representatives, Jan. 27, 1970 (p. 4); U.S. Government Printing Office, Washington, D.C. 1970.

43. Mesthene, E. G., "How technology will shape the future," *Science*, vol. 161, pp. 135–143, July 12, 1968; based on a chapter from the book *Environment and Change: The Next Fifty Years*, Indiana University Press, Bloomington, Ind., 1968.

44. Mesthene, E. G., *Technological Change: Its Impact on Man and Society*. Cambridge, Mass.: Harvard University Press, 1970.

45. Dror, Y., "Approaches to policy sciences," *Science*, vol. 166, pp. 272–273, Oct. 10, 1969.

46. Jantsch, E., "From forecasting and planning to policy sciences," *Policy Sci.*, vol. 1, pp. 31–47, 1970.

Walter H. Kohl (F) received the Dipl. Ing. and Dr. Ing. degrees in engineering physics from the Technical University in Dresden. In 1930 he left his native Germany to emigrate to Canada, where he joined Rogers Electronics Tubes, Ontario, ultimately becoming a company director. In 1935–40 he was also a special lecturer at the University of Toronto. From 1952 to 1958 he was Research Associate and lecturer at the Electronics Research Laboratory, Stanford University. He joined NASA's Electronics Research Center in Cambridge, Mass., in 1966, and later was named chief of its University Affairs Office. He remained there until the Center's dissolution in June 1970 and now is a consultant in Belmont, Mass. Dr. Kohl is the author of a series of books on materials technology for vacuum devices. In recent years he has focused his attention on the impact of science and technology on modern society.

New product applications

Semiconductor power-switching amplifiers are designed to actuate motors and solenoids

Especially adapted for servo systems with inductive actuators, the MCA and MCB units are a family of hybrid class-D power amplifiers. These amplifiers use switching techniques to implement the pulse-width-modulation operation. The MCA and MCB 1000 types employ a complementary output stage and planar output transistors. The circuits operate from a dual unregulated power supply (to ± 40 volts). The new MCA/MCB 2000 amplifiers with an external inhibit point (see Fig.3) will replace the existing series.

Continuous dc output current for the A units is 2.5 amperes and for the B units it is 5 amperes. Peak output currents, with a 25 percent duty cycle, are 5 and 10 amperes respectively. Operating temperature range for all units is −25° to 125°C.

Switching frequency for the A units is 20 kHz and for the B units, 40 kHz. Input hysteresis for minimum, typical, and maximum is, respectively, 175, 200, and 225 mV. Typical thermal resistance is 2°C/W and maximum is 3°C/W. Switching time for the A units: typical, 1 μs and maximum, 1.5 μs. Time for the B units: typical, 0.5 μs and maximum, 1μs.

A functional representation of the power amplifier is shown in Fig.1 where transistors Q_1 and Q_2 are 10-ampere devices with $f_T = 80$ MHz. This configuration has shown rise times of 500 ns with total delay time through the amplifier of 3 μs. When the amplifier is connected to an inductive load with appropriate current feedback there is a ripple current that is required even for the condition when no output current is required. This ripple current occurs at such a high frequency that the load cannot follow the instantaneous variations. The filter network (shown in the figure) comprising R_1C_1 provides a time delay to ensure that both transistors cannot be on at the same time, thus eliminating short-circuit currents between power supplies.

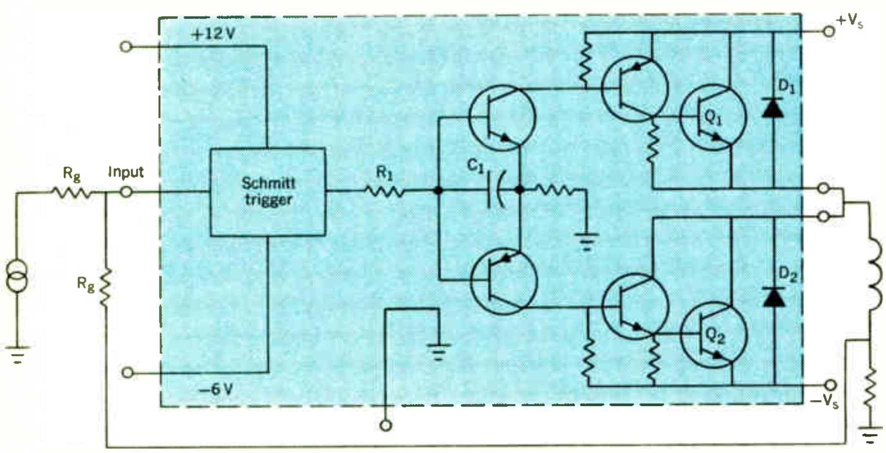
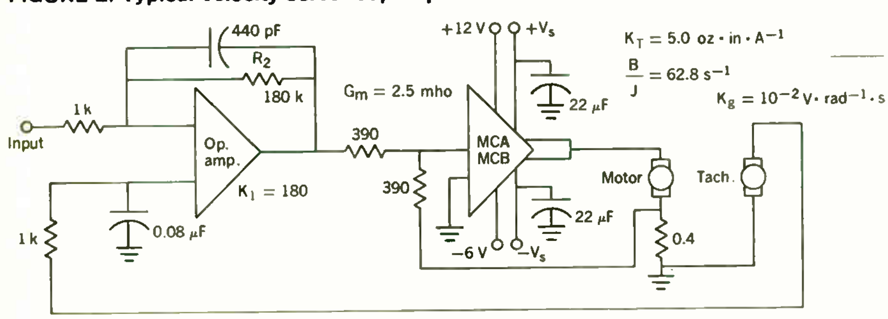


FIGURE 1. Power amplifier showing a typical application.

FIGURE 2. Typical velocity servo loop implementation.



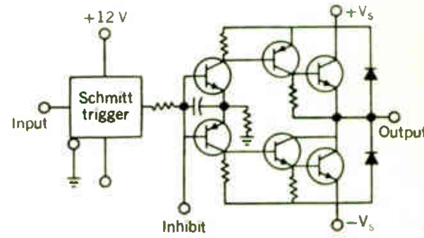


FIGURE 3. Added inhibit lead.

The basic oscillating frequency decreases with load current because of the effects of the *IR* drop in the load and any back EMF that may be generated. The current into the load rises as fast as the *L/R* time constant of the load will allow. The transfer function of the power amplifier is a transconductance when the current feedback mode is used. Normal amplifier operation is maintained for input frequencies of approximately one third the switching frequency.

Figure 2 shows an example of a closed-loop velocity system with tachometer feedback. Because the bandwidth of the power amplifier is 300 times the servo bandwidth, there is no frequency-dependent factor associated with the gain constant.

An important difference between the classical linear amplifier in a servo system and the pulse-width-modulated amplifier is the elimination of the motor electrical time constant in the loop equations. The reason is that the PWM system programs load current and is independent of the load impedance.

First step in the design of the servo system is to evaluate the overall transfer function of the system. Next, the overall system frequency response should be analyzed. After completing the design of the servo system, it is necessary to select the appropriate amplifier from either the MCA or MCB series. This choice depends on the three parameters: magnitude of supply voltage, required load current, and switching frequency.

This and other application examples are worked out in an information sheet available from TRW Semiconductors, Inc., 14520 Aviation Boulevard, Lawndale, Calif. 90260.

Circle No. 85 on Reader Service Card