VOLUME 7 □ NUMBER 3 IEEE spectrum MARCH 1970

IEEE spectrum

features

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

Three new reasons to specify

DIGITEC

Model 262

Digital Multimeter

(AC, DC, Ohms)

A .1% Multimeter for everyday usage. A necessity for design and development, production, quality control, or anywhere that DC volts and current, AC volts, and ohms are measured. Battery pack available.

$375

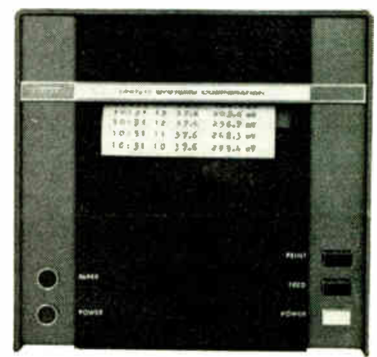
Model 311

Precision Calibrator

(Voltage & current source)

A .01% precision voltage and current Calibrator that serves as a working standard. In addition, the high current capability may be used as a lab source for developing those critical circuits so essential in electronics today.

$650



Model 691

21 Column Printer

(3 lines per second)

A Drum Printer which is expandable from 4 to 21 columns to satisfy your specific requirements. This versatile printer accepts all standard BCD inputs and provides 38 symbols along with "floating" decimal point.

starting at **$770**

See us at booth 2E 44-46

IEEE 70 MARCH 23-26

DIGITEC® by UNITED SYSTEMS CORPORATION

918 Woodley Road • Dayton, Ohio 45403 • (513) 254-6251
Representatives Throughout the world

For complete specifications request new catalog D69A

Circle No. 4 on Reader Service Card.

departments

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 Policy on the Presentation of Socio-Technical Material

The IEEE is a voluntary worldwide organization of individuals who have joined together because of a common interest in electrical and electronics engineering. The IEEE furthers these interests chiefly by providing a forum for the presentation and discussion of papers in its publications and at its meetings. In the past, these forums have served mainly to disseminate information of an exclusively technical nature. The airing of controversy, whenever it arose, has been a traditional and important part of this dissemination process, serving to speed the generation of new technical information to resolve divergent views.

The discussion of technical matters will continue to be the primary function of the forum provided by the IEEE. However, today, with engineering developments having a profound impact on society and social conditions shaping the course of technical developments, it is essential that discussions of the social as well as the purely technical aspects of electrical and electronics engineering work be included. Since this is a new area in which factual data are somewhat limited, it is expected that controversy will be generated. It is also expected that, as in the past, the airing of such controversy will be beneficial and will speed the resolution of critical socio-technical problems throughout the world.

Therefore it is important to the aims of the Institute and the professional needs of its members that controversial papers on the impact of technology and society on one another be permitted in IEEE publications and meetings. To enable the presentation of such material in a manner appropriate to the needs of the membership, and recognizing that the IEEE membership is widely divergent politically, the following guidelines are to be implemented by the Publications Board and followed by editors and meeting organizers:

1. The subject matter should be relevant to the field of electrical and electronics engineering and to its relationship to the needs of society. If the relevance or appropriateness is not self-evident from the author's presentation, it should be made clear by the addition of a suitable introductory statement. The discussion of pertinent interrelated social, economic, and technical aspects may lead to political conclusions on the part of the author. Since political conclusions ordinarily have particular relevance to a specific national environment, such limitations on the author or speaker's conclusions should be made clear if they are not self-evident.

2. Every reasonable effort should be made to provide for adequate and timely presentations of differing viewpoints. This may be accomplished either by planned simultaneous presentations or by presentation of one side of an issue, clearly identified as such, with provision for prompt, subsequent presentation of representative audience discussion and rebuttal.

3. It should be made evident to the audience that the opinions expressed are those of the author, and no endorsement by the Institute, its officials, or its members is implied.

Policy on social material

In my open letter to members of the IEEE, which appeared on page 6 of the September 1969 issue of *Spectrum,* I invited comments on a Proposed IEEE Policy Statement on the Presentation of Controversial Material. The proposed statement was developed as a result of reader reaction to material that had appeared earlier in the year in *Spectrum*. It may be of some interest to members to know of the response to my letter and the final action taken by the Board of Directors.

A total of 296 members responded to the open letter. Of these, 201 respondents, or 68 percent, expressed themselves in basic agreement with the proposed policy statement and in favor of including relevant social and political material in *Spectrum*. The remaining 95, comprising 32 percent, opposed the proposed statement, feeling that *Spectrum* should avoid social and political material and restrict itself to purely technical articles. Suggestions for minor revisions were included by 28 respondents.

The responses were reported to the Board of Directors and considered by them in formulating a final statement of policy. The statement finally and unanimously approved by the Board at its January meeting, entitled "IEEE Policy on the Presentation of Socio-Technical Material," accompanies this letter.

On behalf of the Board of Directors as well as myself, I would like to express my sincere thanks to all those who gave us the benefit of their viewpoints on this important question. Your valuable suggestions were appreciated.

F. K. Willenbrock
President IEEE, 1969

Bibliographical correction

The paper by Harmuth in November *Spectrum* is a valuable indicator of the increasing interest being taken in Walsh functions, with their natural affinity for binary computers. The pity is that the prime reference in the field, Walsh's original paper, is once again misreferenced: In 1949, Fine[1] placed it in volume 55 instead of 45 in *Trans. Am. Math. Soc.*; Pratt et *al.*[2] copied the error in *IEEE Proceedings* in 1969, and now Harmuth[3] does the same. It will probably only be stopped now if a more than usually emphatic correction is made: the correct reference is Am. J. Math., vol. 45, pp. 5–24, 1923.

J. Rodney M. Vaughan
Litton Industries
San Carlos, Calif.

1. Fine, N. J., "On the Walsh functions," *Trans. Am. Math. Soc.*, vol. 65, pp. 372–414, 1949.
2. Pratt, W. K., Kane, J., and Andrews, H. C., "Hadamard transform image coding," *Proc. IEEE*, vol. 57, pp. 58–68, Jan. 1969.
3. Harmuth, H. F., "Applications of Walsh functions in communications," *IEEE Spectrum*, vol. 6, pp. 82–91, Nov. 1969.

An author responds

In his letter in the December issue of *IEEE Spectrum*, elicited by my article "Space Exploration—Wisdom or Folly?" in the October issue, Mr. Flores questions my reasoning and uses some arguments that I wish to challenge. He states that we should establish well-studied criterions for the merit of investigations with respect to the benefit of society before the government allocates funds. He thinks that the supersonic transport project would have benefited society more.

I agree with Mr. Flores that possible spin-offs from the space program alone do not justify the enormous expenditures. In my article, I purposely did not mention secondary spin-offs such as developing new materials, which Mr. Flores wants to explore through the SST project. I think, too, that the possible impact of new technologies on society should be carefully evaluated, as I said in my article. All social and environmental problems that plague this country, such as poverty, urban decay, and pollution, should be attacked with the same determination, planning, and funding as the space program. However, the question should not just lead to the alternative: either funds for solving the social problems or funds for the space program. Rather, the question should be how space technology can and will be utilized for helping to solve these social and environmental problems. Many of these problems have become continent-wide and global problems—such as overpopulation, energy generation, pollution, food production, mineral resources, and weather. And global tools, such as satellites to monitor weather, earth resources, crop disease, and fish detection, are necessary to measure and control them. These problems will increase with growing population and with the demand for a higher living standard, for more food, material goods, and energy. In order to be equipped to meet the growing problems, scientific knowledge must expand and new technologies be developed.

Space exploration opens a new frontier for science and technology far greater than the SST project can. The new space environment has the potential for certain discoveries in physics, biology, and medicine that cannot be expected on earth. Knowing the seeming barrenness of other planets, we will appreciate the earth more, with its great variety of life forms, organic materials, and beauty. We will then ask how it came about that the earth did develop to such a rich and diverse place. We might find the clue in the geology of the moon and other planets. Perhaps the earth was once as barren as the moon. From understanding the processes that occurred on the moon, we might be able to deduce what processes occurred on earth in its ancient history. This could help us to find ways to preserve the natural life on earth, and might even enable us to create an earth-like environment on the moon or other planets. Perhaps it is man's final destination to carry life to other planets like a seedling rather than to expect life there.

May I be allowed to point to a more loosely related question that always comes to my mind when discussing the space subject. As I am for assessing the value of space exploration before large amounts of funds are committed, I feel that other large expenditures also should be examined, expenditures that do not appear as a nicely lumped sum of tax money such as that for the space program funds. As one example out of many, one should ask if it is wise to spend billions on exaggerated advertising. During the last five years, television commercials increased by 40 percent. The cities and towns have become ever-expanding forests of huge expensive billboards and signs. I throw more than half of my mail, i.e., useless advertisements, immediately into the wastebasket. To produce this excess of paper, more forests have to be cut; additional papermills have to be built, which create excessive pollution; the transportation and mail systems are burdened; the garbage collection is clogged and the garbage decomposition again produces pollution. The consumer does not pay taxes for this waste but he pays for it in subtly increased prices and service fees.

Another point that we should not forget is the fact that man still needs adventures, i.e., explorations with uncertain results and risks. America could not have been discovered, and the socially underprivileged of Europe could not have immigrated and developed her into what is now the wealthiest nation on earth—without adventure and risk. The human mind needs challenge and new frontiers, otherwise it would deteriorate and the social problems would not decrease but increase.

H. Trauboth
NASA
Marshall Space Flight Center, Ala.

Tax exemption

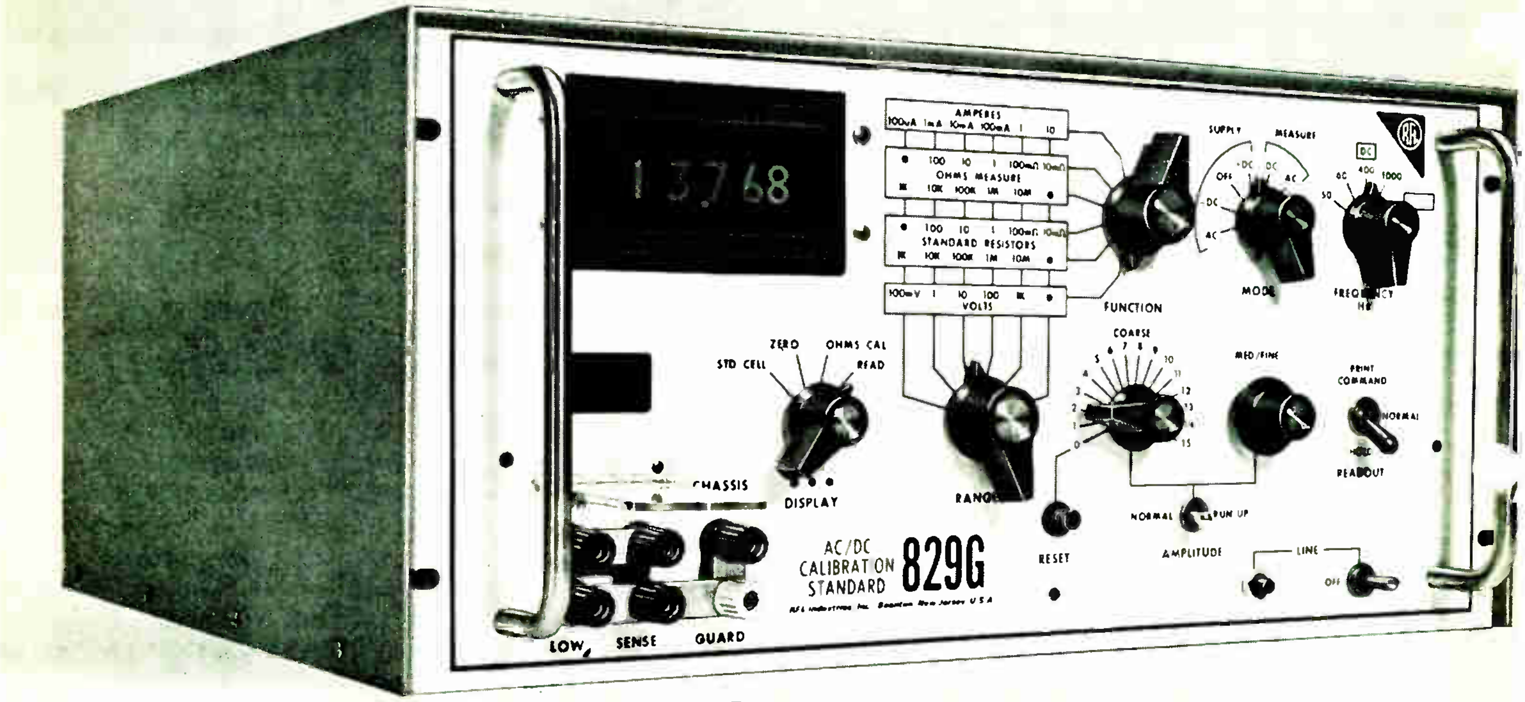
After listening to conversations and reading articles on professionalism, recognition, and related topics, I am making a motion as a member that IEEE members vote to change the IEEE tax exemption in accordance with that of the National Society of Professional Engineers. As described by Paul H. Robbins, P.E., executive director of the National Society of Professional Engineers, in the December 1969 issue of *Spectrum*, this would gain "no restrictions on activities to influence legislation."

Donald H. Ross
Torrance, Calif.

Portable pensions

The letter on "Portable Pensions" by Dennis Beech in the January 1970 issue of *Spectrum* really struck the nail on the head. The IEEE is the only logical organization to handle pensions, being far and away the most influential single center for us E.E.s. The IEEE does handle life insurance, health benefits, etc., and only logically should become involved in pension work. This is even more important than the other things, as there are external life insurance and health insurance companies. To my knowledge, there is none for us as far as pensions are concerned. Mr. Beech mentioned the usual difficulties in getting a pension due to changes because of con-

calibration and measurement in one instrument.



Model 829G
The best of both worlds

Now you can take advantage of the unique RFL Model 829G Universal Calibration Standard — a new, doubly useful instrument that provides a precision source of AC and DC volts, amps, ohms, — plus precision measurement of these parameters from external sources . . . with 5-digit read-out.

The new 829G replaces the famous 829 series, for 15 years the industry calibration standard. Take a look at just some of its capabilities:

AC or DC voltages in 5 ranges, 0.1V to 1000V. Currents in 6 ranges, 100uA to 10A. Ten precision resistance values, 0.01 ohm to 10 meg. AC frequencies: 50,60,400, 1000 Hz and EXT. AC Watts capability when locked to another AC source. Four-terminal sensing in both source and measurement modes, plus high accuracies, high resolution, and high output regulation. And all this priced at what you'd expect to pay for a calibrator alone. You can't afford not to have it—the cost $3100.00.
Write for complete data today.

RFL Industries, Inc. Instrumentation Div. Boonton, N. J. 07005. Tel.: (201) 334-3100 / TWX: 710-987-8352 / CABLE RADAIRCO, N. J.

RFL Industries, Inc.

VISIT BOOTH 2E02 AT IEEE

tract cancellations, etc. in the aerospace industry. Many of us know of considerable company hanky-panky that takes this form: Get rid of a man who's been with you a long time and is due to retire and thus save x dollars by not having to pay his full pension. The IEEE could really do a job in the pension area.

The university plans called TIAA-CREF could logically serve as the model initially. These plans are "portable pensions" for professors that hold no matter what schools they transfer to.

Hubert D. Goldman
Bayside, N.Y.

Transnationalism

I congratulate Dr. McCue on his "Spectral lines" of January 1970, "Thoughts on Being Nonnational." The subject is very pertinent to discussions now current in Canada on the merits of setting up a purely national electrical/electronics society. Some very illustrious IEEE members are sponsors of this idea. I oppose it, and am aware of numbers of colleagues who share my opposition view.

I consider that IEEE policy statement no. 22 (representation to national agencies of a technical nature in any country) provides the mechanism whereby the existing IEEE organization can serve the membership in various countries most adequately. In my view a duplicate society in Canada would fragment and frustrate the membership. It would most certainly cause "time-sharing" problems to individuals should the Canadian society set up duplicative activities.

Experience with certain types of IEEE organizations has shown me that, left undisturbed, the majority group may gradually phase over to the philosophy that they are discussing matters that impact only on the United States. If a foreign national is present, it takes only one reminder to bring the group back to its "nonnational" or "transnational" character. If we, non-U.S. nationals, have complaints that some IEEE bodies tend to be too U.S.-national, it is up to us to press for international representation on these bodies. I would urge that Headquarters encourage this approach. Region 7 (Canada) is the largest Region outside the U.S.A. Perhaps Canadian members should show leadership in aiding the nonnational character of the IEEE.

A similar perspective, shown in the last-published Secretary's report, is that the proportion of higher-grade members is greater in the U.S. Regions than it is in the non-U.S. Regions. This gives evidence of a chain reaction, particularly at the Fellow grade. It is quite easy for a nominator to locate five Fellows to sponsor his candidate if the candidate is working in a large U.S. industrial complex. Grave difficulties exist in locating five Fellows who are cognizant of the work of, for example, a deserving candidate in Trois Rivières, Quebec.

To further the climate of transnationalism in the Institute, I would urge the Board of Directors to consider factors that would aid the recognition of merit outside the U.S.A.

J. A. M. Lynch
Canadian Forces Headquarters
Ottawa, Ont., Canada

Sic transit gloria mundi!

Come, Sancho, let us mount our steeds and ride forth once more to slay the dragon of ignorance. Tell me not that my dream is impossible, nor that my dragons are windmills!

For, lo, the dragon of ignorance is a loathsome, rioting beast that stalks our land in hordes today, causing educated but cowardly men to tremble and run for cover. Once hunted nearly to extinction, he cleverly donned many disguises and sequestered himself in the very halls of his enemy, our public schools, where he has multiplied his kind, and now attacks these institutions that fed and sheltered him.

Oh woe! He has even insinuated himself (perhaps in great numbers) into that once great and learned society, the IEEE, where he corrupts our once sacred editorial boards. When it would be just as easy and far more noble for these boards to banish him by holding high the standards of excellence for those poor benighted members already in his grip, he saps their strength and convinces them that their only salvation lies in defining ignorance as virtue.

I see you doubt me. Then read for yourself the insidious editorial in the November 1969 *Spectrum* and the equally insidious article on page 19 of the December issue, where his pawns argue that we must now renounce our linguistic heritage and accept barbaric plurals for Latin words in order to preserve their singular forms. What diabolic cant is this?

What has happened to all the brave men who once rode with us? Do they no longer see the devils and monsters that threaten us? Once we were millions; now only you and I remain. But as long as I draw breath, I must fight on. Quick, Sancho, my sword and spear!

B. M. Oliver
Hewlett-Packard Co.
Palo Alto, Calif.

Dr. Oliver's quixotic foray imperils the compilers of all recent dictionaries of American English.

The Editor

Pollution

The December 1969 issue of *Spectrum* contained a letter on pollution from Mr. Herr. I wish to add my endorsement of his suggestion that IEEE do some constructive antipollution work.

After years of being a conservation nut, I suddenly find that I'm on the side of the politicians. But I am afraid that only a surface treatment will come of the current antipollution kick.

What is needed is solid technical work on every level of engineering to apply all that we know to designs that will be functional and still preserve the ecology. This is where the IEEE comes in.

Why doesn't the Institute form Mr. Herr's committee, sponsor ecology symposiums, offer prizes for outstanding designs that minimize or prevent pollution, and sponsor research?

If the engineering students at CalTech and the University of California at Berkeley can sponsor antismog research, certainly a group of mature engineers can really make a dent in this problem.

R. N. Buland
Richmond, Calif.

Pollution and licensure

I would like to answer Alexander J. Kelly, whose outburst against P.E. licensure, on pages 10 and 11 of the January issue, I find interesting. I happen to be a P.E., and I had to take all the exams. Although we simple

The BUSS HTA fuseholder measures only 1-25/32 inches in overall length and extends behind the face of the panel only one inch.

The holder features the popular bayonet type knob. A strong coil spring inside the knob assures good contact when the fuse is inserted into the holder. If a test hole in the knob is needed, a breakaway hole can be punched out to allow use of a test probe.

Rugged in construction to withstand vibration and shock, the HTA fuseholder can also be furnished with a special washer to make it drip-proof from the front of the panel. And the best feature of the HTA fuseholder is that it has *famous built-in BUSS quality*. You can't get it anywhere else.

For more information on the HTA fuseholder, or anything else in the complete line of BUSS small dimension fuses, fuseblocks, and fuseholders, write for BUSS Bulletin SFB.

Circle No. 8 on Reader Service Card.

Bussmann Mfg. Division
McGraw-Edison Co.
University at Jefferson
St. Louis, Mo. 63107

SUPPLIED THE ECONOMICAL WAY ... THRU DISTRIBUTORS
BUSS QUALITY
FUSES

people, as a group, may not be as cognizant of the dangers of atomic radiation and other nuclear phenomenons, we try to be familiar with everyday things like a section modulus or a wiring diagram for an across-the-line starter. I sometimes have to interview engineering help. The average P.E. will show some knowledge in everyday technical problems, but many non-P.E.s seem to be untouched, *tabula rasa*, by those simple problems that confront us every day, although they may have considerable experience in space travel.

As to the responsibility for air pollution and many other social ills, for which Mr. Kelly makes us as a group responsible, that may very well be so, but only as citizens, not as P.E.s. Mr. Kelly seems gracefully to forget that appropriations for all those nice improvements have to be signed by somebody and this some*body* is very seldom a P.E. Although many of us are not too concerned about the danger of electromagnetic radiation (except the unlucky ones who wear heart pacers), all of us are aware of the dangers of ungrounded electric power systems, say in the 480-volt range. According to published reports, approximately five persons are electrocuted every day in the United States. Many of these catastrophes can be traced directly to the ungrounded power system.

Yet many existing 480-volt power systems are ungrounded, and when this writer tries to convince the plant management about the necessity of installing grounded systems, he is invariably greeted by the question, "Will this increase sales?" Similar questions are asked about air pollution prevention equipment, and since the NEC is only a recommendation, not a law, many other codes are the same. The P.E., no matter how sincere his intentions, is helpless. This does not absolve us from the guilt for many of the ills Mr. Kelly is enumerating, but these are the facts.

Eric Hauser
Rego Park, N.Y.

A divergence of views

If David Ginsberg's letter ("Broadening the Curriculum," January 1970) was meant as a model of the dispassionate engineer employing his beloved scientific method through strict attention to empirical experience, elementary logic, and the quantitative evaluation of factors, it fails miserably on each and every count. Mr. Ginsberg stands accused in his own words of employing a "propaganda club by a rigidily regimented self-righteous" (in this case) anti-intellectual. Quick to identify the scapegoats of those he would criticize, he creates his own straw man, whose ten years (his numbers) of tinkering with our social system has resulted in all of our current social problems, and then some.

Empirical evidence? Elementary logic? Quantitative evaluation of all the factors? All we are presented with are the biases and prejudices of a man who exposes his ignorance of the subject.

I would observe to Mr. Ginsberg that ranting is ranting, whether from the mouths of wild-eyed revolutionaries from the liberal arts schools he deplores or from tough-talking members of the Establishment who can see no further than their status quo.

Arthur R. Rifkin
Great Neck, N.Y.

The letter by David Ginsberg in the January *IEEE Spectrum* is a masterpiece. Congratulations are in order to Mr. Ginsberg for such a scholarly and effective letter. These doublethinking liberal sociologists, psychologists, political scientists, and others of warped intellect are indeed responsible for the massive harvest in degeneracy sweeping America. It is time for straight-thinking engineers to stand up and be counted on the side of proven principles.

John K. Carter
Ojai, Calif.

Research needed?

Referring to "Focal points" in the November issue of *IEEE Spectrum*, I would like to relate to you some interesting information.

Five years ago, I visited a large broadcasting transmitter located on the top of a hill, and in that circumstance, I was very surprised to find that none of the personnel have had a male child, but that all have had only female children.

I think it would be very interesting to carry on a statistical investigation on the families of the people working in strong microwave fields.

Luigi Totaro
Dallas, Tex.

Don't just READ about come and SEE

here are now over forty UK companies exhibiting at the 1970 IEEE lectronics Show in New York—many of them new to the US market.

dd a note to your check list and look them up. Your visit could just reveal hat elusive product you've been trying so hard to track down!

udge for yourself the quality of the products on show. Revel in the vorkmanship and the comprehensive pecifications. Admire the performance and ou'll begin to warm to the idea of doing usiness with us. Check our prices and elivery schedules and you'll hate yourself or not having done so before.

n case you missed the comprehensive istings on British products and exhibitors ublished in the January '70 issue of Spectrum' as a special supplement—a heck list is included in this ad. Fuller details of our products can be found by cross-eference to the listings in this issue or in he Show Guide.

Further details or a copy of the special 'Spectrum' supplement are available from the Information Office, **The Electronic Engineering Association,** Berkeley Square House, Berkeley Square, London, WIX. 6JU. England.

SPECIAL SUPPLEMENT JAN '70 SPECTRUM

Spectral lines

Engineers and revolution. Engineers seldom look upon themselves as inciters and fomenters of revolution. But a revolution need not be bloody; it is any drastic transformation of social institutions. For many years now, one of the principal sources of pressure on existing social institutions has been engineering. Recently, the pressures have mounted so high that we are living in a revolution.

As an example in miniature, consider the effects of biomedical engineering—starting, say, with the clinical use of X rays a little before 1900. It has utterly transformed medicine, by facilitating fundamental research and by multiplying manyfold the practicing physician's knowledge of his patient's condition. No longer does the doctor come to the home, austerely equipped with a little black valise containing a stethoscope, a thermometer, and some tongue depressors; now the sick person goes to a hospital, less comforting than the old-time doctor, but far better equipped for finding out what has gone wrong. Loss of personal patient–doctor acquaintance is offset by the fact that now the practice of medicine really works.

As long as the chances of being helped by a doctor were less than 50:50, which in the U.S. was true in the large until 1913, medical care was something that many people could do without. Now it is important, and there is a widespread feeling that everybody has a right to it; the increased demand makes possible a more fine-grained specialization and a larger corps of auxiliary personnel. Medicine, not long ago an art, has become an industry.

Clearly evident in this example are three normal effects of technological change. One is that the system (medicine) is better able to produce the desired result; another is that the system's inner organization is revolutionized. A third and characteristic one is that producing the desired result has created, outside the system, a new problem—in this case, a great increase in numbers of the aged and helpless.

Absorbing the effects of biomedical engineering has required tradeoffs, but the improvement is manifest and the penalties relatively light. With certain other engineering developments—most notoriously, the internal combustion engine—the penalties are turning out to be so overwhelming as to have generated the feeling "We cannot go on like this." It is easier to say that than to say what to do. Or even, with hindsight, what our predecessors should have done. Academic people, and the young whose conceptions they have shaped, are inclined to put the blame on profit-oriented businessmen. But what social planner, if in control of affairs, could have foreseen that the automobile and the tractor would turn so many cities into near-insolvent ghettos?

What the revolution is about, it seems to me, is that engineering is now being practiced on such a large scale that the tradeoffs and side effects have, so far as possible, to be taken into account. Because this has seldom been done, people who wring their hands over "the human condition" usually regard technology as hostile to the good life.

Taking the tradeoffs and side effects into account now will involve a shift in values, but a shift is probably appropriate. In advanced countries, maximizing production is no longer a necessity; we can afford to look at the tradeoffs.

Another way of describing the revolution is to say that the piecewise approach to large engineering developments is being replaced by a systems approach, and the system is more than just the industry that is undertaking the innovation; it may be the whole nation. As a result, engineering problems have become political problems. Examples at the moment are cable TV, the ABM system, and—on a smaller scale—the location and design of power generating stations.

Much of my mail comes from men who think that in this time of engineering-induced revolution, the proper stance for an engineer is to thrust his head firmly into the sand and keep it there. Others say that an engineer should be aware of social problems, but that giving consideration to them will destroy the Institute. Others speak warningly, and vaguely, of Pandora's box. These letters, along with no small number expressing approbation, come in because the Publications Board has encouraged IEEE SPECTRUM to give some attention to social and political questions that have particular significance to electrical engineers.

It now appears that my most alarmed correspondents have not spoken for the majority. As Dr. Willenbrock reports in this month's Forum, the response to his open letter is favorable to the new editorial policy by more than 2 to 1.

In attempting to foresee the social effects of engineering action, and to make a rational choice of tradeoffs, complete success is not to be expected. Perhaps there can be no success at all. But in the present climate of opinion, the effort will assuredly be made. It is heartening that so many engineers are unwilling to leave study of the problems entirely to the sociologists and the politicians.

J. J. G. McCue

Ballistic-missile defense radars

The practical defense of the United States against the menacing specter of devastating high-speed weaponry has touched off lofty debates of epic proportions, but few are aware of what such a system really involves

Charles M. Johnson U.S. Army Safeguard System Office

Ballistic-missile defense has been continuously researched in the United States for the past 14 years. The Safeguard system itself is an outgrowth of both the Nike–Zeus system and later ballistic-missile components developed by various government agencies. The classic result of this early work was clear—an ICBM could be intercepted by another missile. Hence, the real controversy concerning the workability of such systems centered around an entirely different problem—the "high-traffic environment"—or how to make tens to hundreds of near-simultaneous intercepts while tracking thousands of targets. This article describes the radars that have led to a solution of this problem, and how such radars operate within the framework of the world's most sophisticated defense system.

Although the Safeguard Ballistic-Missile Defense (BMD) system, its objectives, its major components, and their functions have been described in various news articles during the U.S. Congressional debate over deployment of a BMD system, let me briefly review some of the general features before addressing the specific subject of ballistic-missile defense radars.

First of all, the objectives of the Safeguard system as announced by President Nixon in March 1969 are

- To defend our Minuteman missiles against a possible first strike by Soviet Union intercontinental ballistic missiles (ICBMs).
- To defend our SAC bombers against a possible Soviet Union submarine-launched ballistic missile (SLBM) threat.
- To defend our population against the possible Chinese ICBM threat of the next ten years.
- To protect our population from an accidental launch by any nuclear power.

The major components of the Safeguard system are the following (see Fig. 1 for Spartan and Sprint data):

- PAR (perimeter acquisition radar)—UHF phased-array radar for long-range search, acquisition, and track of incoming enemy ballistic missiles.
- MSR (missile site radar)—the battle management radar. The MSR is primarily a precision target-tracking and missile-guidance radar. It is a high-traffic-capacity phased array that operates in the S-band region. It has less range capability than the PAR but higher resolution.
- Spartan—a long-range exoatmospheric interceptor capable of making intercepts at ranges of several hundred kilometers from the launch battery. It carries a high-yield warhead that can kill enemy reentry vehicles (RVs) at a range of several kilometers.
- Sprint—a short-range high-performance terminal interceptor. This interceptor has an operating range of several tens of kilometers against enemy RVs. Greater radar tracking precision is required for Sprint intercepts than for Spartan intercepts due to the much smaller Sprint nuclear warhead.
- Data-processing system (DPS). Although a data-processing subsystem is an integral part of each of the Safeguard phased-array radars, it is of sufficient fundamental importance to system operation that it is considered a major component. Functions performed include radar control, incoming target selection and tracking, interceptor guidance, message formating, and data reduction for command and control displays and consoles. The Safeguard data processor is configured as a multiprocessor and is denoted as the central logic and control (CLC) system. Several processors with access to a common memory can operate in parallel on several different tracks or on different parts of the same track.

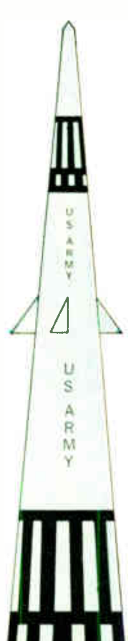
Fundamentally, the technical prescription for the Safeguard BMD system comprises the following:

- Intercept a missile with a missile.
- Kill a nuclear weapon with a nuclear weapon.

FIGURE 1. Safeguard components.

SPARTAN
Height: 16.7 meters
Range: 60 km
Speed: 13000 km/h
Warhead: ~ 1 megatonne

SPRINT
Height: 8.2 meters
Range: 40 km
Speed: Unreleased
Warhead: ~ 1 kilotonne



• Establish continuous surveillance of possible threat corridors with phased-array radars.

• Sense high-traffic engagement conditions with phased-array radars.

• Control radars and interceptors with high-capacity data-processing systems.

• Initiate or veto system response via a command chain headed by the President of the U.S.

In addition, the Safeguard system is designed to be

• Continuously alert and available for possible threats.

• Capable of being exercised at frequent intervals to full radar and data-processing capacity as an integrated multisite defense system.

• Operable while undergoing maintenance.

• Capable of being tested on-site for "glitches," troop readiness and proficiency, and operability.

Although achievement of some of these features can only be measured in full after the Safeguard system is deployed, most have been thoroughly researched, developed, and tested during the years of the Nike–Zeus to Safeguard evolutionary development program.

The development status of the Safeguard system at the present time is as follows:

• The Spartan missile is undergoing developmental test firings at Kwajalein Atoll in the Pacific.

• The Sprint missile is undergoing developmental test firings at White Sands, N.Mex.

• A prototype MSR has been completely constructed at Meck Island (Kwajalein) and is being used to track targets.

• A four-processor Safeguard data-processing system has been completely installed at the Meck Island MSR.

• A PAR brassboard model is being developed at General Electric.

• Tactical software development is under way at the Data Processing Center at Bell Telephone Laboratories, Whippany, N.J.

Under fully projected Safeguard deployment (Fig. 2), a deployment to be made only if the threat requires it, 12 batteries, each consisting of an MSR plus Spartan and Sprint missiles, might be deployed over the continental United States. In addition, a total of seven PARs will be deployed around the perimeter of the country, each PAR being located close by (within about 16 km) an MSR/interceptor battery. All of the MSRs and PARs are interconnected by communications lines and the entire network is controlled from a Ballistic-Missile Defense Center (BMDC) located at Colorado Springs, Colo.

Under phase 1 of the Safeguard deployment, the only deployment currently requested by the present Administration, only two defense batteries are authorized; both are located at Minuteman Wings, one at Grand Forks, N.Dak., and one at Malmstrom Air Force Base, Mont. Also included in the phase 1 deployment are two PARs located near the MSRs at the defense batteries. The phase 1 deployment will give some protection to our Minuteman force; however, the major purpose of this deployment is to gain experience and work out problems that develop in the manufacture, installation, test, and operation of a ballistic-missile defense entity.

Evolution of ballistic-missile defense

The Safeguard system is the result of both the Nike-Zeus system, proposed for deployment by the Army in 1962, and later ballistic-missile defense components resulting from research sponsored by the Advanced Research Projects Agency (ARPA) and by other Department of Defense and AEC research-and-development programs.

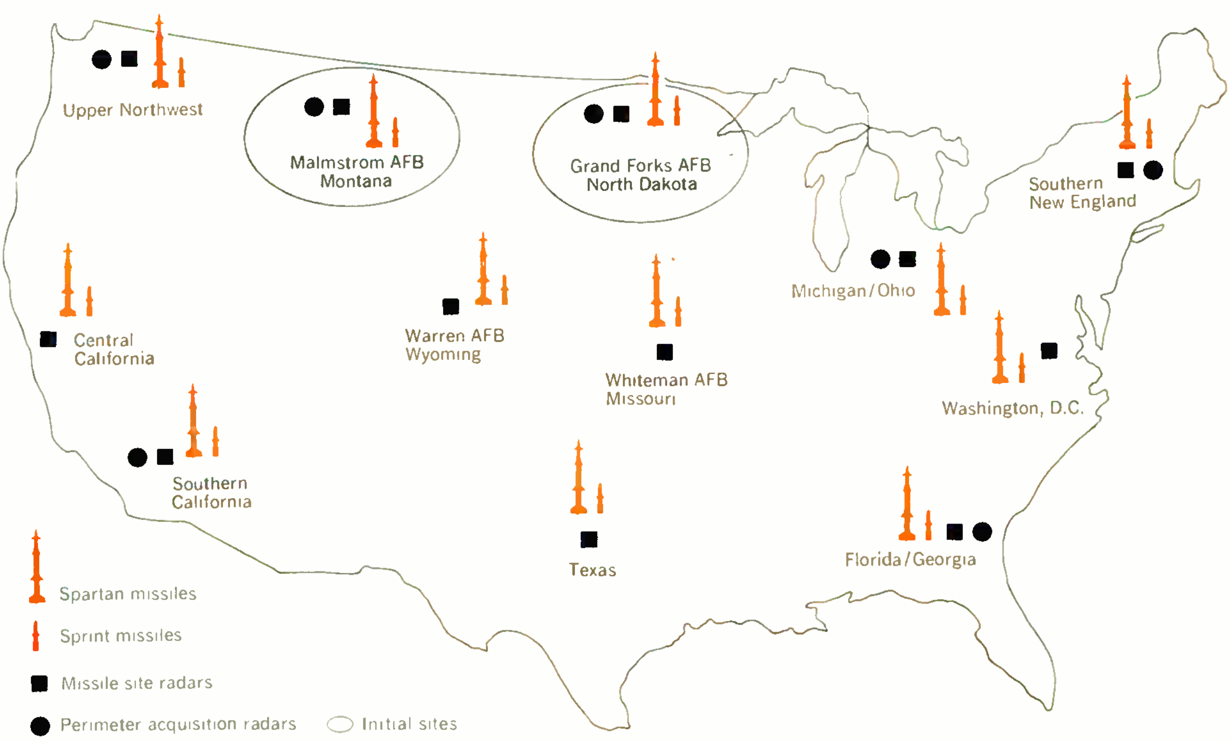
The outstanding result obtained from this eary Nike-Zeus work was proof that an ICBM could be intercepted by another missile. This capability, along with the fact that one nuclear weapon could be killed by another, is seldom questioned by anyone at this time. Most of the controversy as to whether or not a BMD system will work centers upon problems associated with a "high-traffic environment"—i.e., on problems associated with completing tens to hundreds of near-simultaneous intercepts while at the same time tracking hundreds or perhaps thousands of targets. These high-traffic engagements cannot be linearly extrapolated from what is known about a single engagement. There are many interacting problems such as crossing targets, which complicate radar tracking, and simultaneous intercepts, which cause queueing problems for the data processor. In addition, sizable numbers of nuclear bursts clutter the environment and, in many instances, attenuate the radar signals, thus putting an additional burden on both the sensors and the data-processor subsystems. It was not until phased-array radars were developed that one could realistically conceive of a BMD system coping with such a high-traffic environment.

Two basic defense concepts are involved in the Safeguard BMD system—area defense and terminal defense. Area defense involves the detection and tracking of incoming nuclear-armed reentry vehicles (RVs) with long-range radars (PARs) and the interception of these vehicles with long-range defense missiles (Spartans) while they are still high above the atmosphere. This concept is illustrated in Fig. 3.

Terminal defense involves the interception of enemy RVs with short-range high-acceleration interceptor missiles (Sprints) after the RVs have reentered the atmosphere and after they have been sorted out by the atmosphere from decoys, chaff, and confusion devices. Terminal defense is conceptually illustrated in Fig. 4.

Generally, these two types of defense complement each

FIGURE 2. Safeguard deployment in the United States.



other and provide an in-depth defense that places maximum stress on the offensive force.

Perimeter acquisition radar

The perimeter acquisition radar (PAR) is a UHF phased-array type and is used in the Safeguard system to search for, acquire, and track ICBM and SLBM targets at long range. Figure 5 shows a typical PAR site layout. The primary function that determines the PAR power and aperture characteristics is wide-angle search. The requirement to track a great number of objects simultaneously is a strong secondary determinant of these characteristics.

The principal factors that determined frequency-selection and modulation-waveform requirements for the PAR were (1) cost of power and aperture versus frequency, (2) radar cross section (RCS) of reentry vehicles versus frequency, and (3) propagation effects in both benign and nuclear environments versus frequency. Secondary factors considered in selecting frequency and modulation waveforms included radio frequency interference (RFI) and available bands for frequency assignment.

The performance of a *search* radar can be summed up by the following expression:

$$\frac{\sigma_\lambda PA}{T} \sim \frac{\Omega}{t} \quad (1)$$

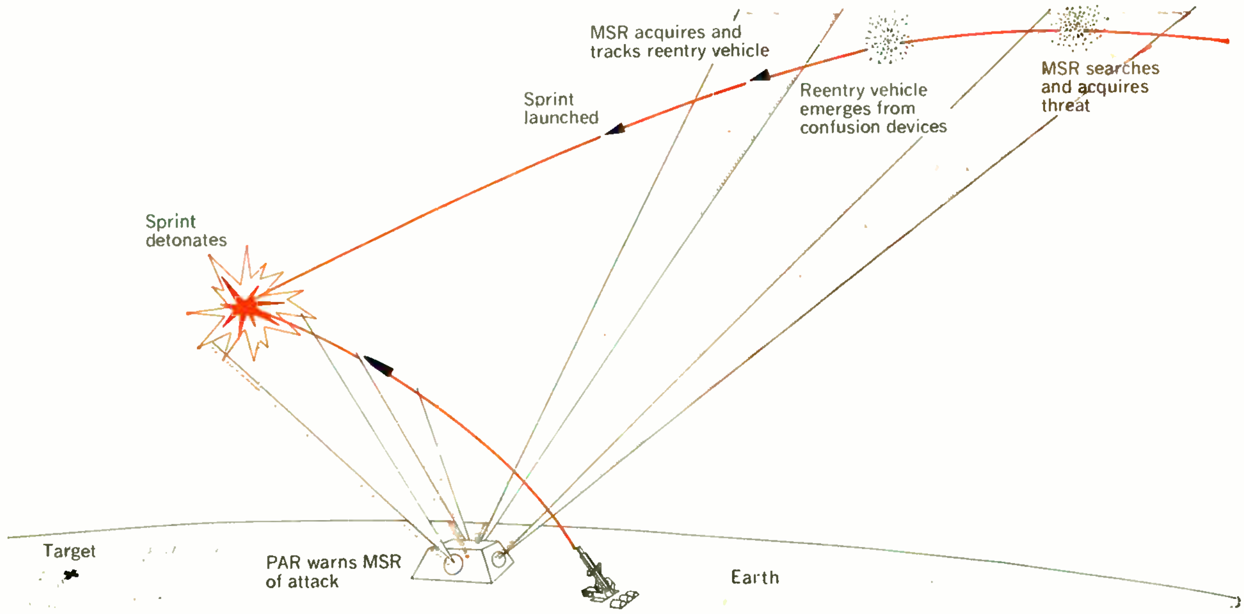
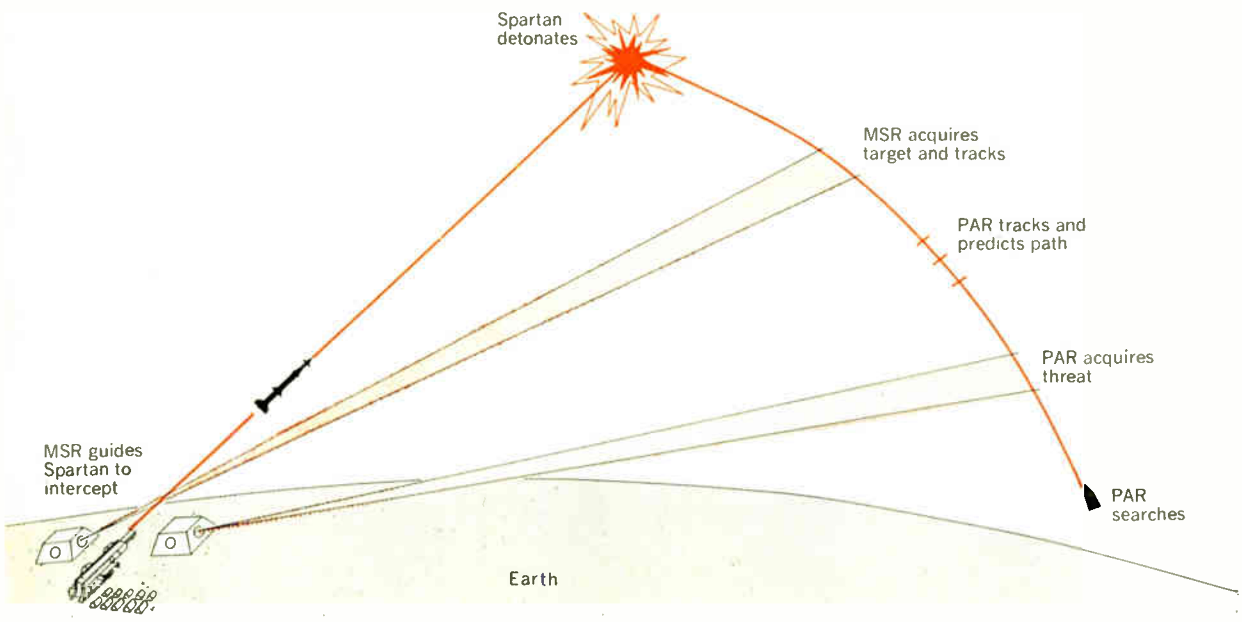


FIGURE 3. Concept of area defense.

FIGURE 4. Concept of terminal defense.



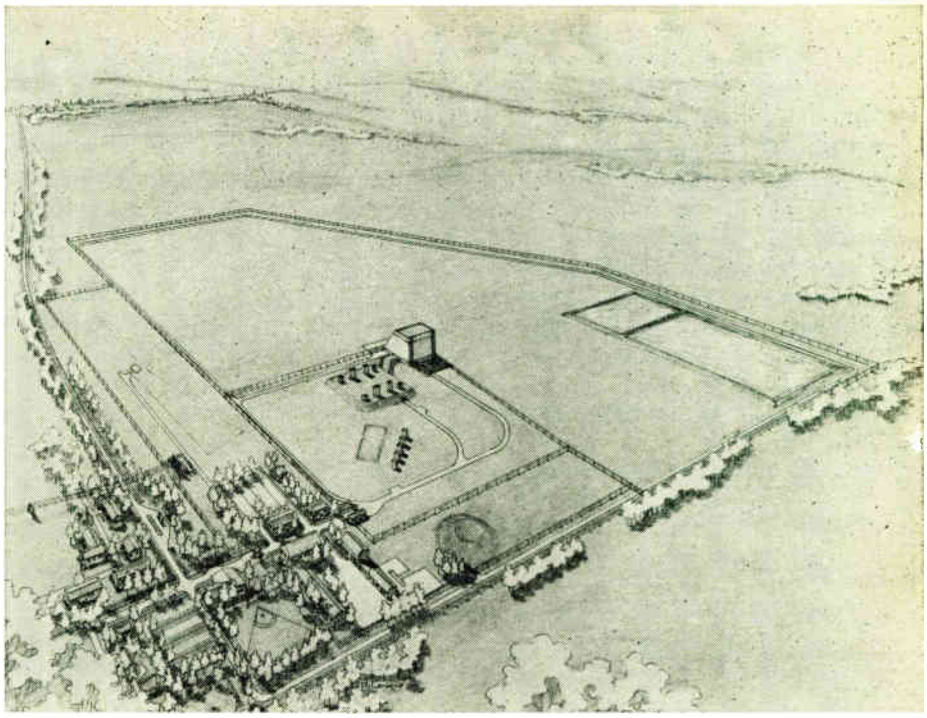
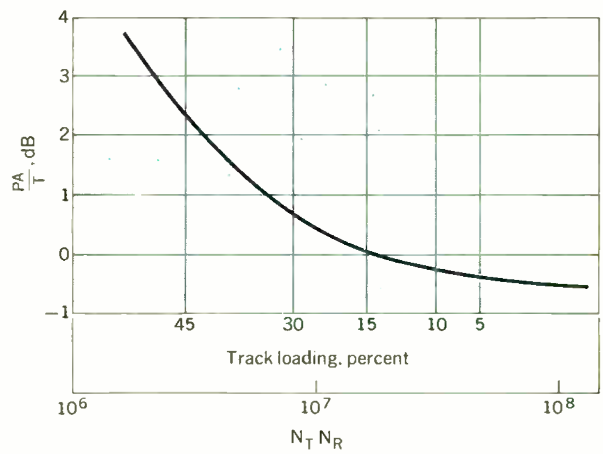


FIGURE 5. Typical Safeguard PARsite.

FIGURE 6. Power-aperture/temperature requirements for PAR.



where σ_λ is the target RCS, P the average power of the radar, A the effective receiver-aperture area, T the effective noise temperature, Ω the required solid-angle coverage, and t the time allowed for the radar to search Ω.

The basic point brought out by this expression is that average transmitter power and receiver aperture are important measures of a radar's search capability, but transmitter gain, or aperture size, is immaterial. Wavelength (or frequency) dependence only enters the relationship indirectly through σ_λ. For large spherical bodies, σ_λ is independent of wavelength, but for conical sections nose-on, $\sigma \sim \lambda^2$, where λ is the radar wavelength. From this relation, it can be seen that it is generally advantageous to operate a search radar at as low a frequency (long wavelength) as possible so long as the wavelength is similar to but less than target dimension.

If the radar must perform a sizable tracking role, then transmitter gain becomes important. The following expression describes the performance of a *tracking* radar:

$$\text{Track accuracy} \sim \frac{PA^2G\sigma_\lambda}{\lambda^2 T} \qquad (2)$$

where G is the transmit gain, λ is the radar wavelength, and σ_λ is again the RCS of the target. The wavelength dependence strongly favors short-wavelength (high-frequency) operation if the target RCS is independent of frequency. However, for conical sections where $\sigma \sim \lambda^2$, the wavelength dependence drops out of Eq. (2).

The PAR design was originally conceived as a VHF phased array with two apertures, one for transmit and one for receive. In this original design, the receive aperture was about 30.5 meters in diameter, but the transmit aperture was much smaller. The PAR is now being developed in the UHF region and only one aperture, common for both transmit and receive, is used.

Briefly, the reasons for these changes are as follows. The PAR searches for, and tracks, exoatmospheric ICBM targets. The radar beam must penetrate the ionosphere and, consequently, it is subject to the refractive and attenuative effects of both natural ionization phenomenons, such as aurora, and nuclear-induced ionization phenomenons, such as beta blackout. Generally speaking, the intensity of these high-altitude ionization effects decreases with ν^2, where ν is the radar operating frequency. Thus, the shift from VHF to UHF significantly reduces the effects of these ionization phenomenons on PAR performance.

The switch to a single aperture occurred because of the increased track capacity that could be achieved with a

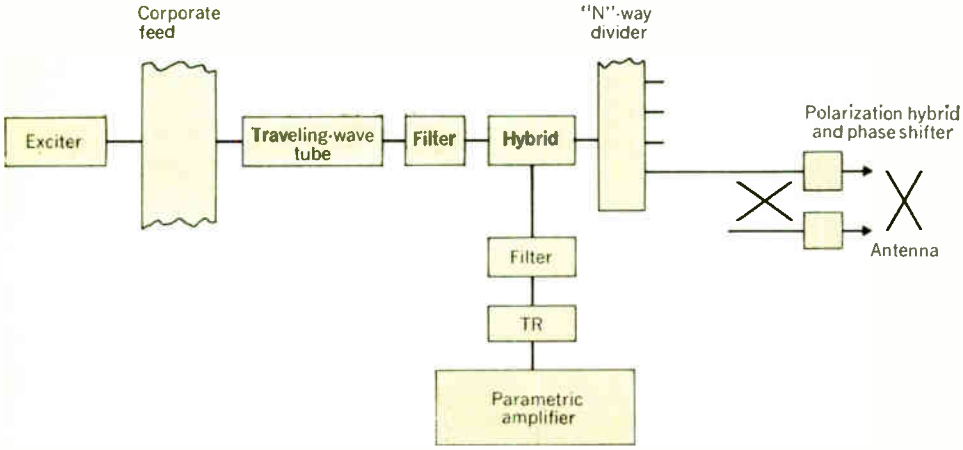


FIGURE 7. Schematic design of PAR.

larger transmit aperture, and the cost reduction that could be achieved from the smaller building resulting from combining both the transmit and receive functions into one aperture.

Figure 6 displays the relative cost in *PA/T* of providing for the PAR track load as a function of $N_T N_R$, where N_T is the number of transmit elements and N_R is the number of receive elements. Thus, the track requirement can easily dominate the design of PAR-like radar when the transmit aperture is small compared with the receive aperture.

With the foregoing general system considerations established for the PAR (and I must apologize and plead classification restraints for being so qualitative and skimpy in covering these system aspects), let's take a look at how we arrived at the detailed PAR design. First of all, at the time of the last major transient in the PAR design, namely, in early 1968, several phased-array radar approaches had been developed and some actually implemented, and these were available for us to consider.

The principal approaches included

1. *FPS-85 radar.* This system utilized low-level heterodyne steering for the transmitter elements and a high-level transmitter output tube at each radiating element. Separate transmit and receive apertures were used. A significant advantage of this scheme is the relatively low cost of the low-level transmitter phasing system. A disadvantage is the relatively high cost of using a complete transmitter chain per element (if full PAR bandwidth and power requirements were to be met by this design).

2. *Space-fed arrays.* In this case, a single or a few high-power transmitter tubes are used to feed a phased-array antenna from a focal point behind the array. Correspondingly, receive horns are used in this focal area behind the array to detect the return signal. The use of a single aperture forces the system to use high-power phase shifters for at least the portion of the array illuminated by the transmitter. The major advantages of this design are the simplicity of the antenna feed and the relatively low cost of a few ultrahigh-power transmitter tubes. However, these transmitter tubes must be designed for exceptionally high reliability in order to match the high inherent availability/reliability of the phased-array radar. Also, these ultrahigh-power transmitter tubes require ultrahigh-voltage power supplies and these supplies are generally more difficult to engineer and they occupy larger amounts of premium radar building space than is the case with lower-voltage supplies.

3. *Corporate-fed arrays.* In this category, a fairly broad range of designs is possible. The choices include single-aperture and double-aperture configurations, high-level or low-level transmitter steering, and a transmitter tube per radiating element or one transmitter tube for many radiating elements. Generally, in the phased-array category, endless arguments can be generated as to the optimum subarraying ratio, transmitter number and division, and receiver number and technology.

4. *Solid-state arrays.* This category is similar to approach 3 except for the substitution of a solid-state transmitter output stage for a tube transmitter output stage.

Designs based on the foregoing techniques were generated that would meet the PAR system requirements. Each of these designs was costed and compared. The general cost trends were the following:

1. Required *PA/T* reduces as $N_T N_R$ increases.
2. RF power-generation costs tend to be lower as power per tube increases.
3. Losses increase as power per tube increases.
4. Prime power costs depend on efficiency.
5. Need for redundancy increases as the number of tubes decreases.

In addition to these directly related radar cost factors, hardened building costs, which amount to a major fraction of the total radar cost, are a function of the radar design. In the case of the PAR, building size is controlled primarily by aperture requirements.

It was concluded from this cost comparison that a corporate-fed common-aperture design with a few hundred high-level transmitters power-divided to the several thousand antenna elements was the most economical approach for realizing the PAR requirements; however, it should be stated that the total cost swing over the entire ensemble of designs amounted to a factor of less than 30 percent.

A schematic design of the PAR phasing network that

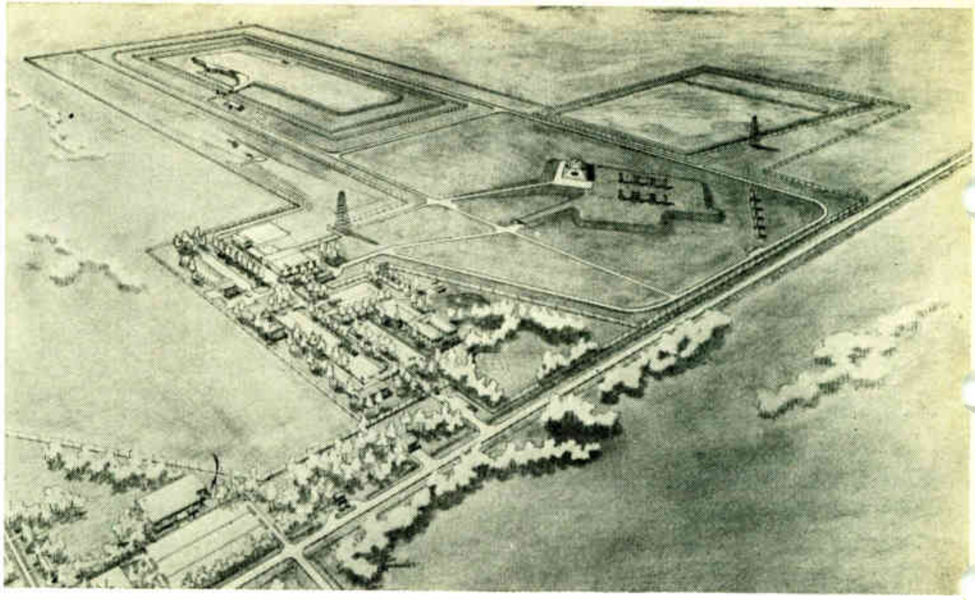


FIGURE 8. Typical Safeguard MSR site.

is currently under development is shown in Fig. 7. It is simply a conventional corporate-fed array with traveling-wave-tube output stages for the transmitter and parametric amplifiers for the receivers. The antenna array consists of a number of subarrays arranged to minimize perturbations caused by transmitter or receiver outages. The phase shifters are 4-bit diode type; orthogonal steering is used and a special-purpose beam-steering computer controls the array of phase shifters. The PAR central processor (CLC) furnishes the sine space coordinates to which the beam is steered.

The PAR utilizes a number of receivers matched to the various functions. The search receiver consists of multiple channels that permit simultaneous generation of multiple receiver beams. The track receivers utilize monopulse tracking to achieve accuracies that are a small fraction of a beam width.

The PAR, unlike some other Safeguard components, must operate continuously—monitoring all space objects that come within its field of view. By the time the radar is in place, this load will amount to several thousand objects per day. Consequently, the PAR transmitters and receivers and their supporting equipment must be sufficiently redundant to provide virtually 100 percent availability.

Missile site radar

Missile site radar is primarily a precision target-tracking and interceptor-guidance radar and is optimized to operate in a high-traffic environment (see Fig. 8). Secondarily, the MSR is also capable of autonomous search and target acquisition. It operates in the S-band region and has a range capability of several hundred kilometers against typical RVs. The MSR is of phased-array design with a 4.1-meter-diameter aperture. This single aperture serves for both transmitting and receiving. The MSR is generally configured as a four-faced radar to provide complete 360-degree azimuthal coverage.

The MSR building has about the same floor space area as the PAR but has only about 1/100 the antenna aperture area. Perhaps the only conclusion that can be drawn from this is that radar building size is not a rapid function of antenna aperture.

The basic design of the MSR was accomplished approximately five years ago and a prototype model is now being operated at Meck Island, a part of the Kwajalein test site.

Figure 9 depicts the MSR antenna array and feed. The antenna is a space-fed phased-array type with separate transmitting and receiving feed horns located at the rear of an anechoic chamber. The antenna thus acts as a focusing lens for the focal point feed. The phase shifters located in the antenna face serve both to collimate the microwave energy from the feed and to steer the resulting beam to the desired spatial direction. The diode phase-shifter cartridge that makes up the array is drawn in Fig. 10.

The MSR functional block diagram can be seen in Fig. 11. A special-purpose computer with appropriate drive amplifiers controls the array phase shifters. The MSR central data processor furnishes sine space coordinates to the beam-steering computer for the position to which the beam is to be steered. The beam-steering computer then solves the following orthogonal equation in order to determine the phase shift θ_{mn} that is required for each element:

$$\theta_{mn} = [m\alpha_1 + n\beta_1 + G_{mn} + (m\gamma_x^2 + n\gamma_y^2)](1 - \nu) \quad (3)$$

where G_{mn} equals the constants that take into account manufacturing tolerances, $(1 - \nu)$ is the frequency correction term, and

$$\alpha_1 = \frac{2\pi dx}{\lambda} \sin \alpha \qquad \beta_1 = \frac{2\pi dy}{\lambda} \sin \beta$$

where dx is the physical separation between the element columns, and dy is the physical separation between the element rows.

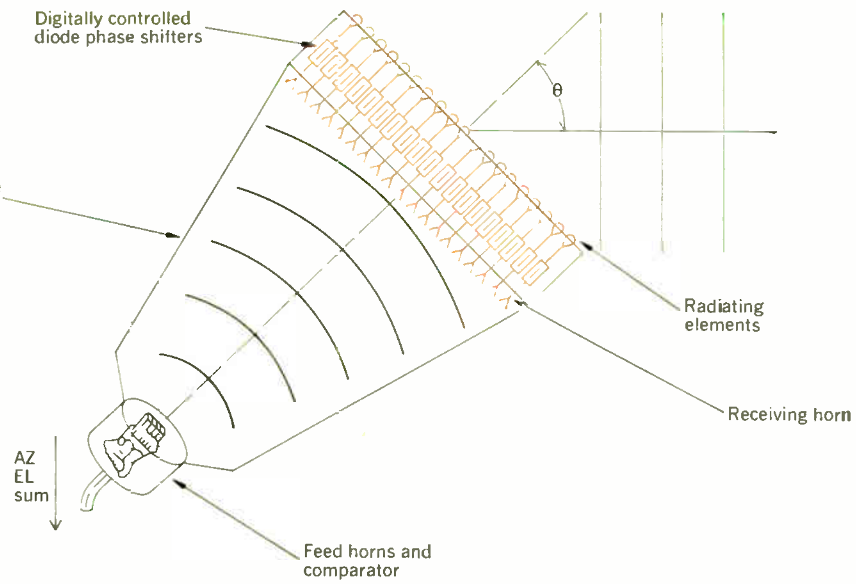


FIGURE 9. Missile-site radar antenna array.

The MSR transmitter consists of redundant, high-power klystrons capable of generating high peak power in a variety of pulse widths. The transmitter also produces exceptionally large amounts of average power, thus permitting the MSR to perform its multifunction roles. Because of its completely redundant design, it is possible to perform maintenance on the transmitter while it is "on-line."

The transmitter output tube is shown in Fig. 12. The MSR receiver uses redundant channels to achieve virtually 100 percent availability. There is a set of receivers matched to each of the possible transmitted waveforms. These receivers have a common RF section featuring a cooled parametric amplifier. The MSR has both search and track receivers. The track receiver utilizes amplitude monopulse to determine accurately the position of targets to within fractions of a beam width. All receivers can be switched to handle inputs from the different array faces.

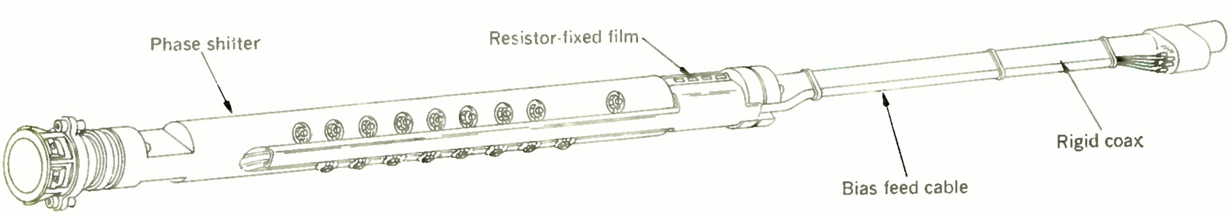
Data-processing system

The data-processing system (DPS) hardware for the PAR, MSR, and BMDC sites uses common hardware modules, or racks, interconnected as a multiprocessing system. The central logic and control unit (CLC), which is the heart of the system, uses at least two each of the following units: processor units, program storage units (each containing two 8K 64-bit word modules), operand storage units (each containing 16K 64-bit words), input/output controllers, and timing and status units. Additional processor units and storage units may be added as required. Multiprocessing, in addition to providing assurance of large throughputs by adding more processors, has the further advantages of modularity, permitting the most economical size of DPS installation at each site as well as providing, through redundant units, a high degree of fail-safe operation.

The following examples will give a general feel of how this data-processing hardware can handle the BMD problem:

1. When the PAR makes a track measurement on an object, it uses this measurement to update and refine the predicted trajectory of the object. Each such calculation requires very many operations in the data processor. However, the data-processor speed is such that a single processor unit can perform the entire calculation in about 1 or 2 milliseconds. Because the radars were designed to give high-accuracy measurements on each track pulse, only of the order of 100 pulse hits per target are needed, and these can be spread over several minutes of the target's trajectory. Therefore, the average time required of a processor is less than a millisecond per object per second. Thus, a single processor unit may perform trajectory predictions for over 1000 objects

FIGURE 10. Phase-shifter cartridge.



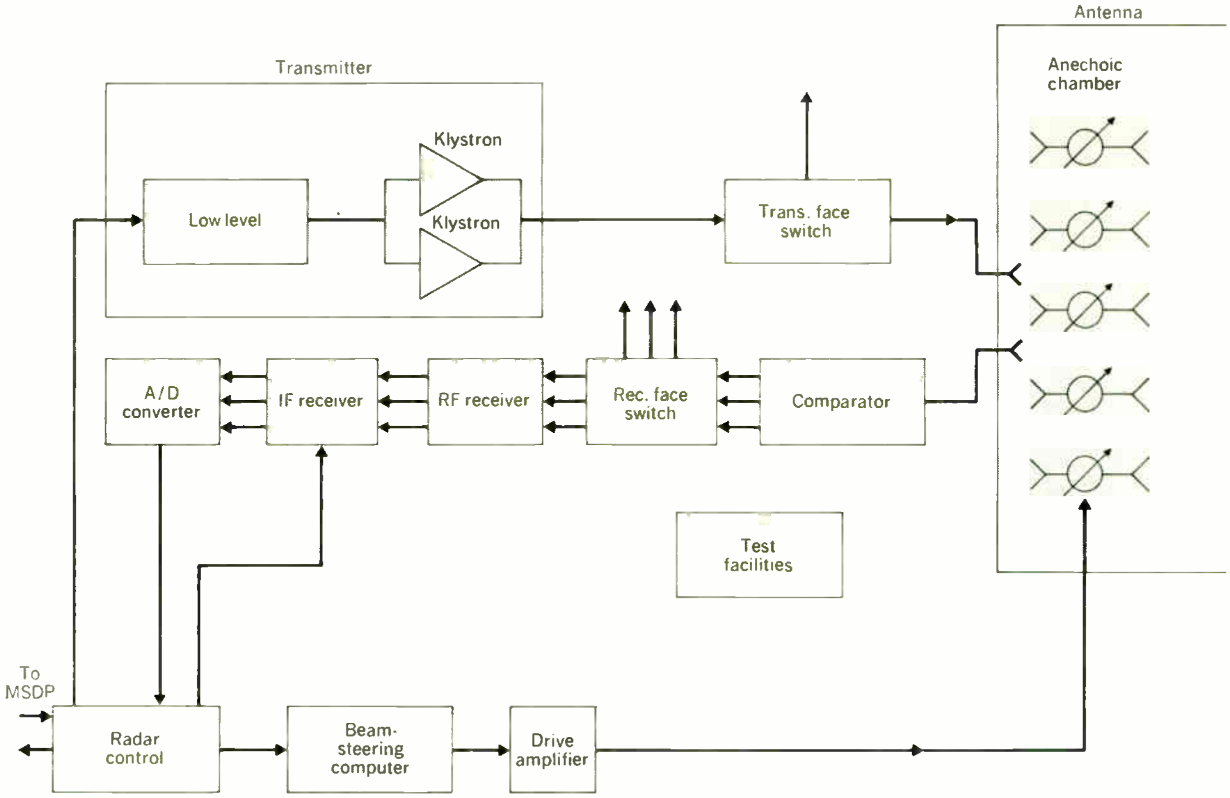


FIGURE 11. MSR functional block diagram.

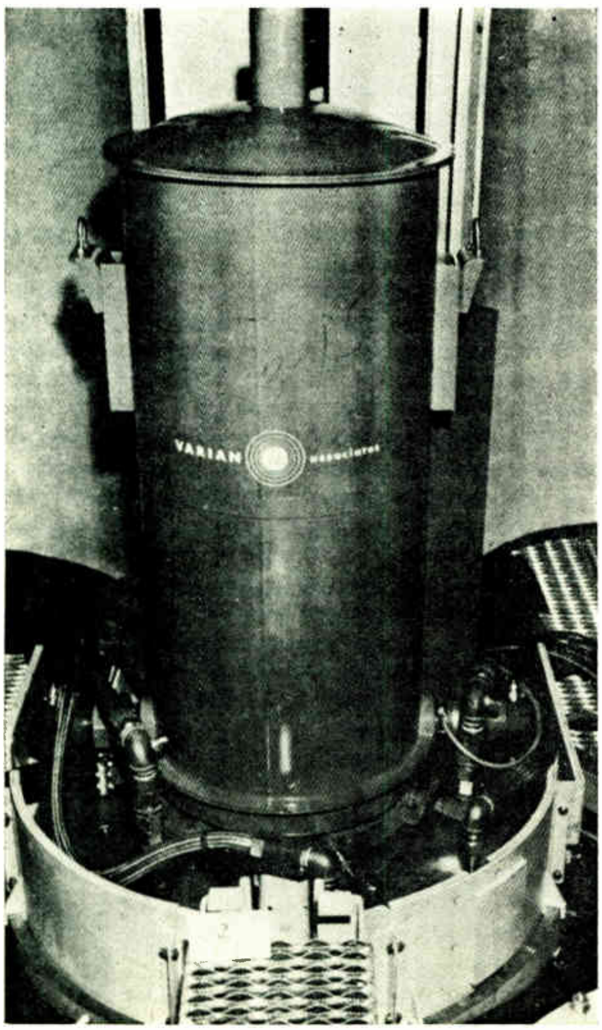


FIGURE 12. Klystron in solenoid of MSR transmitter.

simultaneously. Since a massive attack would last for minutes and not all objects would appear simultaneously, a processor unit would build up to this rate. As it finished with a particular track, it would drop the object and start on a new object. Thus if, just as an example, an attack lasted for 30 minutes with several hundred objects entering coverage per minute, a processor unit could handle several thousand objects over the duration of the attack.

2. Spartan and Sprint guidance calculations take longer time per calculation and are made at a higher average rate. However, the number of defensive missiles that must be in guidance simultaneously from one MSR is small compared with the total objects (reentry vehicle plus junk) a PAR must handle.

These two examples are discussed since, in the peak-load situation, they are the most demanding of data-processor time at the two different sites. The actual number of processor units required for each defense site is selected based on the postulated threat. Additional processor units are added to handle all of the other functions required to support these basic functions and to supply the necessary number of spare processors to meet availability requirements.

The software weapons process structure is composed of tactical routines plus radar and computer tests, as well as maintenance, diagnostic, and operating system routines. These run on a real-time basis with fixed operating deadlines tied to the radar major cycles.

I. Safeguard radar cost breakdown

Category	Percent of Investment Cost
Radar electronic components	25
Data processor	12
Installation and test	11
Spares	5
Other site equipment	5
Radar building	20
Power plant and power equipment	15
Support facilities, real estate, and design	7
	100

Software facilities based upon third-generation hardware architecture include base-register addressing, double-level indexing, indirect addressing, data editing, and registers for internal central-processor-unit (CPU) operations. The software structure is broken down into tasks, which are in turn controlled through a task-list table. Task precedence is controlled through conditional enablement tables, and run priority is controlled through position in an absolute enablement table.

Safeguard hardware and software are both representative of large-scale computer systems of today. Multiprocessors have been used in several data-processing systems as the logical way to provide large data-processing throughputs when uniprocessors with adequate capacity are not available.

The software for the Safeguard system is a complex real-time process but, again, these functions have been carried out before. Except for peak-tracking requirements, the Air Force FPS-85 phased-array radar at Eglin Air Force Base in Florida, which is controlled by a duplex IBM MOD-65 installation, performs tasks quite similar to those required of the PAR DPS.

Ballistic-missile radar improvements

Although several areas in phased-array technology remain where additional development could lead to significantly improved performance, phased arrays have now reached a state of reasonable technical maturity. Without doubt they constitute one of the major technical breakthroughs leading to effective ballistic-missile defense; but they are very expensive. The most important area to be addressed immediately is believed to be that of reducing the cost of complete phased-array radars. "Complete" means the building, power plant, installation, and testing as well as the electronic portions.

At the present time, our PARs and MSRs are estimated to cost between \$150 million and \$200 million each when all costs, from site activation to final check-out as an operational system, are included (investment cost). The cost for major parts of these radars is listed in Table I.

In addition to the costs that have been designated as investment costs, there is an operating and maintenance cost for these radars that amounts to about 5 percent of the initial investment per year.

Let me hasten to add that as expensive as these radars appear, they are still much cheaper than the corresponding number of dish radars required to handle comparable assumed traffic in a similar defense role.

The universal feature that comes to mind when one talks of future BMD radars, indeed, future phased arrays in general, is the desirability of all-solid-state design. Specifically, this development refers to the high-power transmitter output stages. Virtually all other parts of phased-array electronics are already solid state. Such widespread desire for this development is somewhat surprising when one considers the increased difficulty in radar operation caused by the characteristics of solid-state microwave sources, usually transistors. These difficulties include the following:

1. Complexities in radar signal and data processing are caused by the necessity to utilize the long-pulse, high-duty-cycle property of transistors if good efficiency is to be realized.

2. Special problems in prime-power conversion and distribution arise as a result of the requirement to furnish low voltage and high current to the transistor collectors.

3. Phased-array designs based on moderate power transistors tend toward use of larger numbers of radiating elements to achieve required system power levels. The resulting narrower beam, coupled with a high duty cycle, creates a need for more beam forming and greater steering ability than is usually required in a phased array using tubes. In addition, large apertures tend to make structural hardening more difficult.

Offsetting these technical difficulties and complexities is the solid-state promise of much simpler maintenance procedures and maintenance costs. If the long-life claims made by the semiconductor people are valid, then it is conceivable that phased-array radars can be built that require only periodic maintenance, or perhaps even no maintenance at all, for the life of the system. Such a radar could be unattended, and many such radars could be controlled from remote central sites. The availability of such components would certainly have a major impact on ballistic-missile defense concepts.

Essentially full text of a paper presented at the NEREM Conference, Boston, Mass., November 5–7, 1969.

Charles M. Johnson (SM) is the scientific and technical director of the U.S. Army Safeguard System Office. As such, he advises on all scientific matters, technical policies, and major technical decisions that pertain to developing and deploying the Safeguard system. A recipient of the B.E. degree in civil engineering from Vanderbilt University and the Ph.D. in physics from Duke University, he came to the Safeguard system in 1967 from IBM, where he was the manager of the Applied Research Department and was concerned with communications, solid-state devices, lasers, and radar systems. Before joining IBM in 1961, Dr. Johnson was director of the Advanced Systems and Microwave Departments at Emerson Research Laboratories, responsible for research and development programs in many areas. From 1956 to 1960, he was manager of the microwave and physics section at Electronic Communications, Inc. Other experience was obtained at the radiation laboratory of Johns Hopkins University, and as a consultant to various corporations. Dr. Johnson has published 27 technical papers on topics ranging from microwave spectroscopy and millimeter-wave techniques to ferrite devices and electronically scanned antennas. A member of the American Physical Society, he is listed in American Men of Science.

Underground transmission in the United States

Much more than cost factors has prevented the widespread installation of high-power underground systems. Long range testing programs not only must solve increased technological problems, but also should insure that such systems will deliver reliable service over prolonged periods of time

Charles F. Avila, Andrew F. Corry Boston Edison Company

The means for providing underground transmission of large quantities of electric power at a reduction in present-day cost by at least an order of magnitude ranks among the high imperatives of today's research efforts. Demand for the expansion of "undergrounding" is soundly based upon such factors as the rapid growth of cities, the doubling of power requirements with every passing decade, and esthetic objections to the appearance of overhead lines. Unfortunately, at the present time, such expansion is possible only at excessive cost. The authors present the problem as it exists within the United States, and offer solutions that are commensurate with the state of the art.

The problems and costs of putting high-voltage transmission underground are radically different from those associated with the low voltages of distribution. For one thing, the few centimeters of insulation that separate the high-voltage conductor from earth form a capacitor that absorbs much of the useful current. In a present-day 345 000-volt cable, practically all the current-carrying capability of the conductor will be taken by this capacitance in a distance of about 42 kilometers, so that no useful current can be transmitted beyond that distance.

Compensating equipment is available to overcome this effect, but is very expensive in itself and occupies large areas of costly real estate. The longest underground circuits now in use in the 230–345-kV range are about 32 km.

A Federal Power Commission study[1] shows that, on the average, at 345 kV the costs of underground cable are 15 to 16 times as much as overhead lines in suburban areas and 18 to 19 times as much in rural areas. (The cost of rights-of-way for overhead lines in rural areas is less than in the suburbs.) This explains why only about 0.6 of one percent of all transmission lines are underground, and almost all of that is in urban areas.

Most of the new generating capacity in the next 20 years is expected to come from 250 huge plants of two to three million kilowatts each, in place of the 3000 plants in existence today.[2] Economy and reliability will demand that transmission interconnections be longer and heavier. Yet popular demand and increasing population densities will require that an ever-increasing amount of this transmission be placed underground. It is, therefore, necessary to find much better means of underground transmission of far greater carrying capacity at far less cost.

The state of the art

The availability of air as an electric insulator and heat exchanger makes overhead transmission much less demanding of technology than underground transmission.

At 345 kV, the circuits of an overhead line simply require sufficient open spacing between bare conductors, a string of insulators about 3 meters long for attachment to the steel towers, a few subsidiary protective devices, and a proper respect for the thermal limitations of the wire under high current flow. There are a few complexities but, in the main, nature has been most cooperative.

Undergrounding is a different world! The conductor must be electrically insulated from earth within a few centimeters or so; the confined heat must somehow be disposed of; and the entire system must be protected from corrosion. Here nature has *not* been cooperative.

Both the electrical insulation and the soil around it are poor conductors of heat, and the warmer the soil gets, the less heat it will conduct, thus introducing a vicious circle that may well lead to failure of the insulation because of thermal degradation.

In addition, the cable, its insulation, and the earth make an extended capacitor. For a typical 345-kV cable, this capacitance will be about 11 000 kvar per kilometer. Such large charging currents result in serious operating constraints, particularly during light load periods.

With currently available insulation systems, the dielectric losses increase rapidly with increases in voltage levels. For a 345-kV cable on a typical duty cycle, dielectric loss can run as high as 26 watts per circuit meter! Because the total permissible loss on such a line is about 72 watts per meter, not much room is left (46 W/m) for I^2R losses in the conductor. The result is that the power-transmission capability of the 345-kV system is only 4.2 times that of the 69-kV system. In overhead transmission, a quintupling of voltage operation would increase capability about 25 times.

At 345 kV, the average line cost for a 48-km underground circuit in a suburban area alone would be around \$430 000 a km as against \$74 000 a km for an overhead circuit—but the overhead line would have a capability more than twice that of the underground line (1050 MW vs. 484 MW). In addition to the line cost, the underground circuit would require \$3 760 000 for compensation and terminal facilities not needed for overhead, so that the cost per megawatt-kilometer would be \$1050 for underground against \$70 for overhead.

The solution as of now

The early high-voltage cables, called self-contained cables, were designed to be installed in concrete or fiber ducts. The copper conductor was formed over a hollow core, was insulated by oil-impregnated paper, and was protected by a lead sheath, with a total diameter of 8.4 cm for 132-kV operation. The hollow core was filled with degasified oil under light pressure (one atmosphere) to prevent voids from forming within the insulation structure. If allowed to form, such voids could ionize and initiate electrical failure.

Later on, a high-pressure system, called pipe cable, was devised. In this system, the conductor is formed in segments without the central oil channel, insulated with paper, and shielded with a combination of metal and synthetic tapes, which also act as a barrier against casual moisture. Three of these cables are installed in a single buried steel pipe filled with oil (or, in a few cases, nitrogen gas) at high pressure (13–15 atmospheres). Voids are suppressed mechanically by the high pressure, thus assuring good electrical performance.

The first commercial installation of 345-kV pipe cable, 24 km long, was energized in New York City on May 1, 1964.

Pipe cable has now virtually superseded self-contained cable for underground transmission, as shown in Fig. 1 and Table I. Its predominance is due primarily to the economic advantages of longer workable pulling lengths, fewer expensive splices and manholes, and less critical limitations on the earth's profile. These advantages stem from its lighter weight and high-pressure operation.

Since each 115-kV manhole, with its associated splicing, costs about \$20 000, or about the installed cost of 70 meters of cable, there is strong economic incentive to use greater lengths than are possible with the heavier self-contained cable. Moreover, operation is simplified because the pressure facilities are all in one place and do not require periodic checking and adjustment. Also, maintenance costs are lower because the system is in-

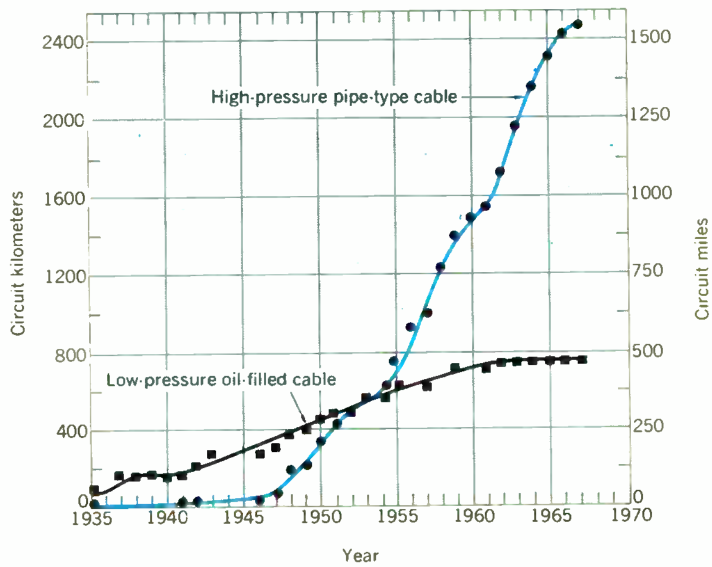


FIGURE 1. Underground transmission cable in the United States (69 kV to 345 kV).

I. Underground transmission cable in the United States by voltage class (through 1967)

	Voltage Capacity				
	69 kV	110–138 kV	230 kV	345 kV	Total
High-pressure pipe-type circuit, km	620	1696	56	97	2469
Low-pressure oil-filled circuit, km	384	364	2	—	750
Total	1004	2060	58	97	3219

herently rugged and less subject to mechanical change.

Installation requires meticulous attention to detail. The inner bore of the pipe must be as smooth as possible and free from any slivers or projections. It is customarily grit-blasted and coated with a rust-inhibiting enamel; and the outside is coated with an asphalt compound to protect it from corrosion. Lengths are welded together using a backing ring to produce adequate strength and a smooth bore, and to prevent welding icicles that might damage the cable.

Great care is taken to prevent moisture from entering the pipe and to remove any traces that might otherwise remain. At the end of each day's work, or at any other time it is not being worked on, the pipe is sealed and filled with dry gas.

Usually, the pipe is evacuated three different times to remove any moisture. The first time is upon completion of a manhole-to-manhole section of pipe, the second is after the cable has been installed in the section, and the third is after cable installation is complete throughout the line. Vacuums used are typically in the 100-micrometer range. In all cases, vacuums are held for a sufficient time to assure that moisture has been removed, and are then broken by dry gas, or, in the last instance, by the dry filling oil. While cable is being installed, dry gas is introduced at the feeding end to minimize the amount of moisture carried in by the atmosphere. What does enter is removed by the subsequent vacuum treatment.

Splicing cable is an exceedingly painstaking and meticulous procedure and requires "white room" techniques with strict temperature and humidity limits being maintained in the manhole.

To insure reliable continuous pressure control, two full-capacity pumps are used with controls so arranged that the loss of one pump has no effect on the system operation. High oil demand or low pressure can trigger alarms.

This care and attention to detail is reflected in the superior performance of the pipe system, as demonstrated in Fig. 2. Disregarding anomalous failures, the annual trouble rate per 100 km for pipe cable ranges from 0.0 to 0.5. This range compares with 0.25 to 2.1 for self-contained cable.

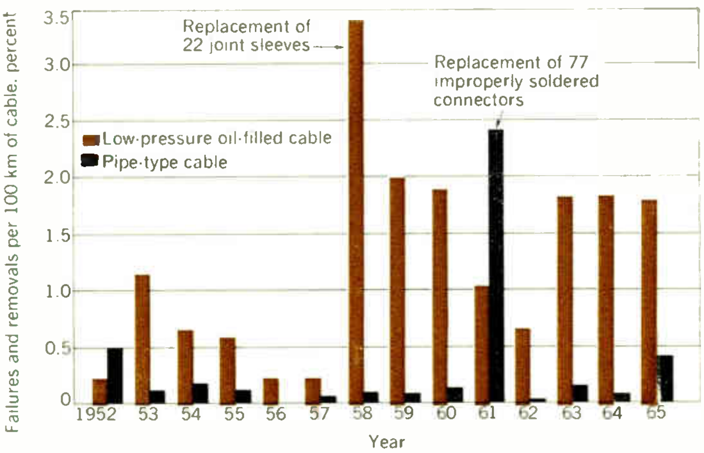
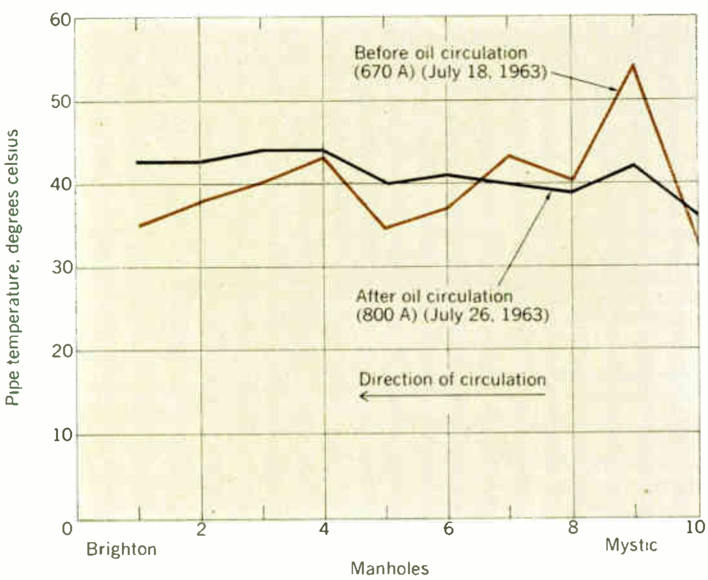


FIGURE 2. Underground transmission operating experience.

FIGURE 3. Effect of oil circulation on pipe temperature for line 329-510 (Mystic–Brighton) (115-kV cable).



The manufacturing techniques and precise quality control that contribute to this excellent record have been highly developed in the case of the 345-kV cable as a result of cooperative research at Cornell University by cable users, cable and accessory manufacturers, and university faculty members. The project demonstrated that 345-kV paper-insulated cable systems could be made to carry 500 MVA at up to 400 kV under operating conditions common for lower-voltage cables if a high degree of precision, uniformity, and compactness in tape application were realized and retained in all reeling operations in the dry state as well as during drying and impregnation.

To meet these objectives, it has been necessary in the fabrication of pipe and cable to use laboratory techniques rare in a large-scale manufacturing process. These include:

1. Applying 2.5 cm of paper to a conductor in a single pass within an enclosure held at 5 percent relative humidity.
2. Use of paper made with deionized water and dried to a moisture content of 1 to 1.5 percent.
3. Slitting and storing paper in a 5 percent humidity atmosphere.
4. Using a traversing take-up reel to eliminate bending in the horizontal plane.
5. Attaining vacuums of 1 to 10 μm in impregnating tanks large enough to accommodate more than 2 km of 345-kV cable with a 10-cm outside diameter.

Reels used to hold this cable weigh 10 000 kilograms (10 tonnes) when empty and 63 500 kg when loaded.

In field installation, the major development has been the economy of increasing section lengths without injury to the insulation. Typically, 115-kV and 138-kV cables run an average 800 to 1000 meters; and 345-kV cable, 600 meters.

Techniques have been developed for very long and very heavy pulls. Partially filling the pipe with oil reduces friction and gives buoyancy to the cable. The longest pull has covered 2500 meters of 1000-kcmil (1 million circular mils) 69-kV cable, from the California shore to the man-made islands used for drilling oil. The heaviest strain has been 2000 meters of 2000-kcmil 345-kV cable across the Narrows at New York harbor.

To analyze the tension sharing across the three conductors, a new apparatus can measure the force on each conductor at the connection to the pulling rope. This device, which fits into a 6-inch (15-cm) pipe, converts the force into an electronic signal that is continuously telemetered to the pulling location and recorded. Information gained by this device will make possible even longer pulls and greater savings. The mathematics of friction around bends and the resulting action of the three separate conductors has received considerable study.

Where there is adequate knowledge of the thermal environment, the maximum conductor temperatures for cables rated at 40–345 kV can be increased from 75°C to 85°C with consequent increases in capability. In research conducted at Princeton University from 1955 to 1961, the industry made important advances in the identification of the thermal parameters of soil, modes of measurement, and improved soil-compacting techniques. Individual companies have done further work showing that maximum soil conductivity is related to maximum dry density and that moisture retention is improved by minimizing the size of individual voids. These desirable characteristics are attained by careful gradation of particle size and insistence on thorough compacting in the field.

In operating an underground transmission system, a major concern is the maintenance of uniform and stable thermal conditions throughout the length of a line. Oil is circulated to level out hot spots and is cooled to reduce overall temperature. Oil circulation is presently used on about 420 out of 2470 km of cable; the longest loop is 45 km. Figure 3 displays typical results for a 14-km loop of 115-kV cable in Boston. Notice that the temperature profile with circulation is much flatter although the load is considerably greater. Figure 4 compares the variation at the hot spot and at the low-temperature manholes, thus demonstrating the averaging effect.

Benefits gained by forced cooling are shown in Fig. 5. New York has almost 48 km of pipe-type cable utilizing forced cooling, of which 13 km are 345-kV capacity. In Boston, the capacity of 21 km of 115-kV cable has been increased 26–33 percent by use of heat exchangers.

Higher-voltage cables promise even higher capabilities. Development of cable suitable for 750 kV rests upon the development of synthetic film or polymeric paper and oil combinations with lower dielectric constants and power factors than those currently obtained with oil-impregnated paper. At 500 kV both paper and synthetic types probably will be employed in practice and the particular choice in a given case will depend on such factors as rating requirement, circuit length, forced cooling developments, relative costs of cable, charging current compensation considerations, and relative design thickness, power factor, and dielectric constant parameters. Some estimates for the increased capabilities at higher voltages are listed in Table II.

Promising candidates

In the field of extruded dielectrics, conventional and cross-linked polyethylene insulation is being employed on a relatively extensive scale for voltages up to 35 kV. Some experimental and commercial circuits have been installed at voltages from 69 to 138 kV. Application to successively higher voltages is shown in Fig. 6. Success here has been achieved through improved dielectric strength and voltage life by use of chemical additives and meticulous manufacturing techniques.

This type of cable has an attractive simplicity of installation, particularly in cases of underground runs of moderate length and load capability at station outlets,

II. Capabilities of EHV cables

	500 kV		750 kV
	Paper	Synthetic	Synthetic
Dielectric constant	3.5	2.5	2.5
Power factor	0.0025	0.0010	0.0010
Insulation thickness	37 mm	30 mm	39 mm
Rating	600 MVA	800 MVA	1090 MVA

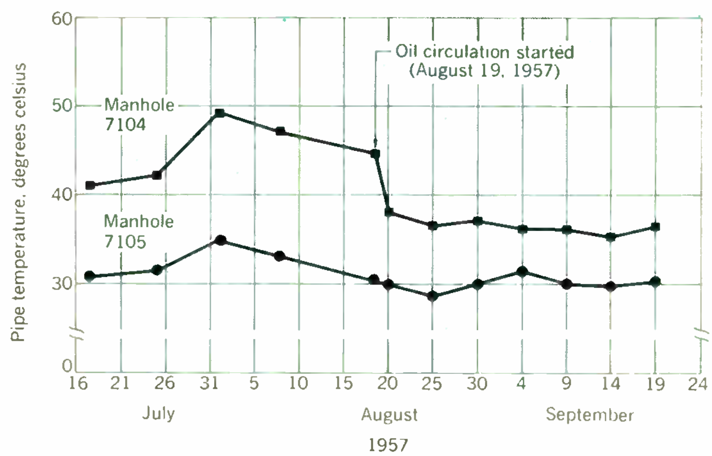
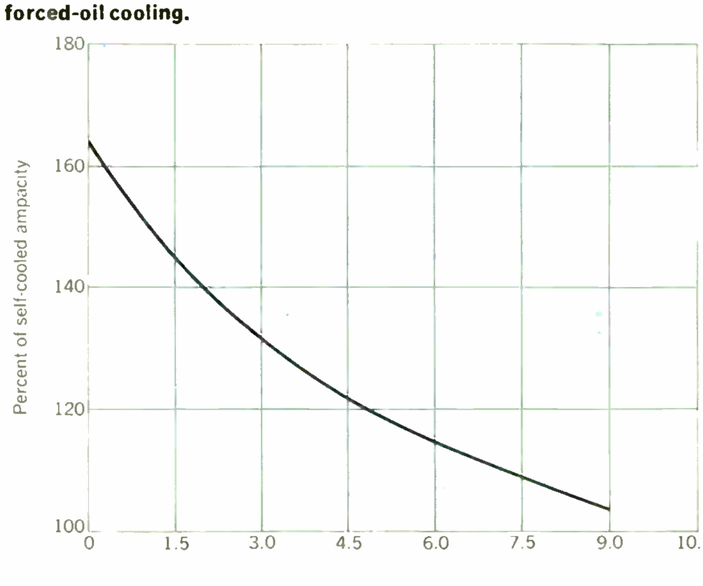


FIGURE 4. Comparative temperature variations at the hotspot and at the low-temperature manholes.

FIGURE 5. Typical capacity increase obtainable with forced-oil cooling.



road crossings, etc. Extension of this design to long, heavily loaded circuits at the higher voltages awaits developments in extrusion techniques as well as results of the Electric Research Council's research program, one objective of which will be to develop reliable procedures for accurately predicting long-term performance of extruded dielectrics for use at 138 and 230 kV. The Schering bridge, which in conjunction with temperature and voltage stress application has been used so successfully on laminate insulation, is of limited use in evaluating extruded structures. Although much work has been done during recent years in the development of discharge detection techniques, and very promising methods are now being used, the industry in general is not yet satisfied with the correlation between the results of these measurements and actual deterioration or breakdown.

Cross-linked polyethylene insulation is characterized by its thermosetting properties and by an extruded shield made from a semiconducting cross-linked copolymer. These characteristics give it a higher temperature capability (90°C) and the inherent stability that comes with a shield that is cross-link-bonded *void-free* to the cable insulation.

Both types of polyethylene enjoy the advantages of simple splicing and terminating materials and procedures. They are easier to handle, are not particularly sensitive to a wet environment, and have no expensive auxiliary systems. Further development work is aimed at improvement of the economic factors, extension to 230 kV, and possible use as dc cable.

Research

A five-year research project on synthetic insulation for extra-high-voltage cable, funded by the Edison Electric Institute, has been completed recently by the Illinois Institute of Technology Research Institute. Several interesting polymers out of a great many candidates were selected as being worthy for further development. Selection was made on the basis of electrical properties and on the compatibility of the materials with certain impregnants. The results have been forwarded to the Edison Electric Institute and a report to the industry is contemplated.

Edison Electric Institute itself is participating in an extensive research program on underground transmission, which is sponsored by the Electric Research Council. The Council consists of 12 representatives from the various segments of the electric utility industry—including investor-owned companies, federal, state, and local agencies, and cooperatives.

The Council's program is expected to cost $17 million over a five-year period. Its short-range goals are

1. To extend the voltage range of existing insulation systems, in particular for cable with oil-impregnated-paper insulation.
2. To improve the power-transfer capability of "state-of-the-art" systems.
3. To shorten the time and reduce the costs of splicing present-day cables.
4. To spur development of long-life, high-voltage cables with extruded insulation.
5. To develop an EHV cable with taped synthetic insulation.

Its long-range goals are

1. To bring into commercial use one or more systems that have the potential for carrying four to ten times the loads of present-day cables at substantial savings in investment. This category includes investigations into the properties of various gases as cable insulation, and into the operation of cables at the temperatures of liquid nitrogen or hydrogen or even down to the superconducting region of liquid helium.

2. To decrease installation costs by developing improved techniques and equipment.

3. To stimulate work in advanced concepts of underground transmission so that the industry may be in a position to take quick advantage of any breakthroughs.

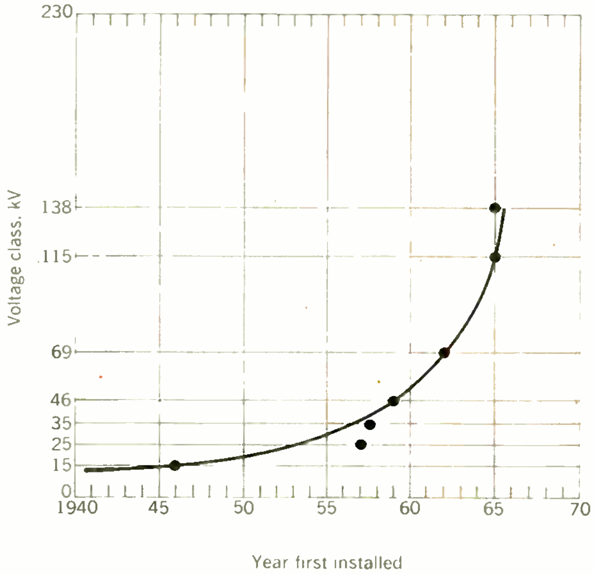
The most urgent objective is the production and testing of 500-kV paper-insulated cables and 138-kV extruded-dielectric cables. To this end, about $12.5 million will be spent to construct and operate a station to test sample cables in commercial lengths (300 meters). This facility has a voltage capability of 1100 kV and is the largest of its kind in the world. It was inaugurated in May 1969 and four 500-kV pipe-type cables have been under test at voltages up to 690 kV and temperatures up to 110°C. The test will cover a two-year period in order to simulate an expected service life of 40 to 50 years.

Samples of 138-kV extruded-dielectric cables were under test before the end of 1969, subjected to voltages up to 200 kV and temperatures up to 90°C. Other new concepts will be tested as they become available. These will include 500- and 750-kV synthetic cables as well as prototypes that result from other phases of the program.

For the next several years, the greater portion of expenditures for new underground transmission systems will be for those at 138 or 230 kV. Thus there is a greater near-term opportunity for savings at these levels than on extra-high-voltage systems. One such opportunity involves improving techniques for external forced cooling of present-day cables. A research project of this nature has recently been completed; it had three objectives:

1. To determine which forced-cooling techniques are readily applicable to state-of-the-art, high-pressure, oil-filled, pipe-type cable lines.

FIGURE 6. First installation dates by voltage class of polyethylene-insulated power cable.



2. To prepare a generic design of a typical cooling system for a specimen pipe-type cable line.
3. To establish the economic advantage of applying forced cooling to this type of cable.

To achieve these objectives, a system study was carried out to assess the cable-system thermal parameters that offer possible alternatives in the cooling system design. The cable chosen was of the pipe type with a conductor size of 2000 kcmil operating at 138 kV. The aim was to increase its rating of 245 MVA (at 90 percent load factor) to a forced-cooled rating of approximately twice this value.

Two possible cooling concepts were considered:

1. Chilling and circulation of the pressurizing oil.
2. Cooling the cable pipe by circulation of the refrigerant fluid in separate channels.

For each of these concepts, a thermal analysis was carried out in order to achieve an appropriate cooling-system design. Also, for each design, a present-value analysis on a dollar-per-MVA-mile basis was made, considering both capital and operating costs of system components, to assess the economic advantage of the forced-cooled system.

In general, the results of the system study indicate that forced-cooling ratings of 450 to 500 MVA can be achieved for a pipe-type cable with a conductor size of 2000 kcmil operating at 138 kV by either of the two cooling concepts. In fact, with low-viscosity oil, ratings up to 550 MVA can be achieved economically.

Cable-joint simplification

Present splicing techniques require critical environmental control and around-the-clock work schedules for long periods. Research will be directed toward simplification so as to cut both time and costs.

There are also other advantages to simpler jointing techniques besides these savings. A method of jointing that reduces the need for specialized work will ease the shortage of technicians. An EHV joint that can be completed quickly would likely eliminate the need for air-conditioning of manholes during the jointing operation. Finally, time saved in making joints will speed restoration after EHV cable faults, thus reducing outage time and decreasing emergency loading time on alternate facilities.

Polymeric paper insulation

The use of polymeric films introduces problems relating to complete saturation of the insulation by the dielectric fluids. Various means of embossing polymeric films have been devised to overcome this problem, but films permitting the porosity and ready liquid mobility that paper possesses are not yet available. Most means of embossing do not lend themselves to easy regulation and variability. Porosity of paper, on the other hand, can be controlled almost infinitely by means of its method of manufacture. Thus, the ideal of complete impregnation of an insulation with a dielectric fluid seems more nearly attainable with paper rather than film.

Research in this area will concentrate on methods of making paper from polymers with desirable electrical characteristics and comparing their impregnation and physical properties with those of cellulosic paper. The various lots of synthetic paper produced will be evaluated by the manufacturers of EHV cables.

Gas dielectrics

The concept of building a cable using only gas as the major insulation is developing rapidly. Industry-sponsored research into the fundamental behavior of compressed gases under simulated cable operating conditions is progressing satisfactorily.

The main effort has been devoted to investigating analytically and experimentally the effect of eliminating free-conducting particles from the interelectrode gap and studying the effect of coating highly stressed electrode surfaces with a strongly adherent layer of dielectric material. The work thus far has indicated that free-conducting particles are a major factor in reducing the insulating strength of an otherwise ideal gap and has demonstrated the benefits of electrostatically removing conducting and semiconducting particles from the active insulating portion of the system. Additionally, the work has shown that the electrical strength of an insulating medium is increased by having a dielectric coating over the central conductor.

There are currently a number of companies actively involved in the manufacture of prototypes and in trial installations of these systems. If progress continues at its present rate, some of these systems may well be commercially available in the relatively near future.

Resistive cryogenic cable systems

Cryogenic cable systems are attractive because electrical conductivity increases so markedly at low temperatures. For example, at liquid nitrogen temperatures, aluminum is ten times as conductive as it is at room temperature and, at liquid hydrogen temperatures, this factor increases to 500.

The disadvantages are the thermodynamic limits on heat removal at low temperatures, as well as the low achievable efficiencies of practical cryogenic refrigerators. In addition, such cables will be complex because of the special insulation, continuous refrigeration, and the necessity to maintain very high vacuums continuously. Because of this complexity, these cables will find use in extending the capability of underground transmission rather than in reducing losses.

The objective of a project now under way is to develop a conceptual design utilizing cryogenic cooling to operate at temperature levels of 20–100°C. Work completed to date indicates that the cryogenic cable may be a candidate for use in the range of 1000 to 5000 MVA. Other indications are that, although the liquid hydrogen cables appear to have a slight cost advantage over liquid nitrogen cables, they appear to be more difficult to develop and require more precautions in handling. For these reasons, liquid nitrogen cables will be preferred, at least initially. A demonstration-size unit is planned for 1971.

Superconducting cable systems

For two years, the Edison Electric Institute has had a project to determine whether superconductivity is feasible for ac power transmission. The first investigation delved into the ac losses in superconducting niobium of a size and shape suitable for practical power cables. Measurements have been made and the losses have been shown to be low enough that this parameter is not limiting.

Although it still has not been demonstrated that the dielectric behavior of compressed liquid helium will be

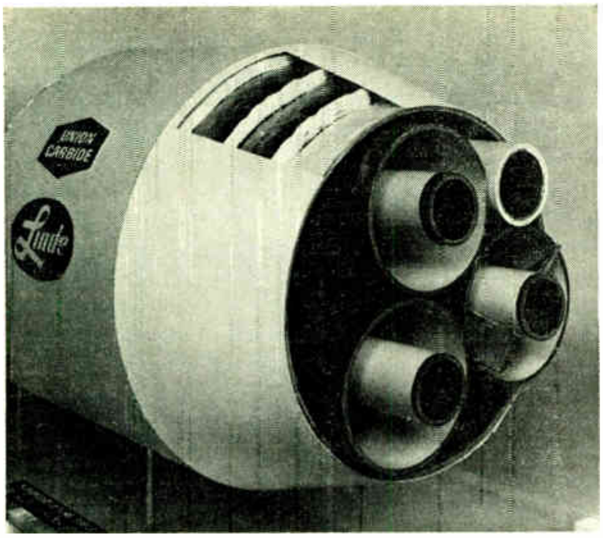


FIGURE 7. Model of the proposed configuration of 138-kV, 3-phase, ac superconducting cable rated at 1690 MVA with multishielded thermal insulation. The outer jacket has a 33-cm diameter; the phase conductors a 3.8-cm diameter. (Courtesy Union Carbide Corp., Linde Div.)

adequate, the preliminary indications were encouraging enough to make an economic analysis based on the information at hand. This analysis indicates that substantial savings over today's costs may be obtained for systems that require capabilities of 1000 to 10 000 MW (see Fig. 7).

It is interesting to note that at low power ratings the cryogenic enclosure and system installation represent approximately half of the total cost. At higher power levels, it is the shielded conductor and installation costs that dominate. Refrigeration costs were estimated to be less than 15 percent of total system cost for all cases studied. It is also estimated that an $8 million program over a period of 12 years will be necessary to bring the superconducting cable to the threshold of commercial operation.

DC cables

Research on EHV dc cables is being currently reviewed to determine if there is adequate industry demand to warrant earlier scheduling of the relatively high research expense that is required. Such a project would investigate all phenomenons associated with dc cable systems in the 100-kV to 600-kV range. Of particular interest is the effect of temperature and of polarity reversals on electric stress for both paper and synthetic insulations.

Conclusion

We have described the wide variety of work in the United States involving both the perfection of existing cable technology and the development of radically new means of conveying large blocks of electric energy underground. In the first category are the near-term projects that promise the quickest response to the urgent demand for expanded underground operation. In the second category are long-term projects that have the potential for greater returns. Both categories offer difficult and exacting research challenges.

The ERC research program is being continually reviewed. A number of subjects have been considered by the ERC Steering Committee that have been postponed for the present. In some cases, the state of the art is not sufficiently advanced to warrant ERC research at this time.

The day of simple, reliable, and economic transmission of great blocks of power for long distances underground is quite far in the future. It is our hope that every segment of the industry and the general public will support the research and development efforts that are necessary to bring this future nearer.

Updated version of a paper presented at the World Power Conference, Moscow, U.S.S.R., Aug. 22–24, 1968.

REFERENCES

1. "Underground power transmission," Federal Power Commission, Washington, D.C., Apr. 1966, p. C-3.

2. "Considerations affecting steam power plant site selection," Office of Science and Technology, Washington, D.C., Dec. 1968.

Charles F. Avila (F), born in Taunton, Mass., received the B.S. degree in electrical engineering and business administration from the Harvard Engineering and Business Schools in 1929 and the L.L.D. from the University of Massachusetts in 1963. Joining the Boston Edison Company in 1929, he rose through many key positions until becoming president and general manager from 1960 to 1967, whereupon he became chairman of the board and chief executive officer.

A director of over a dozen corporate, banking, insurance, and power companies and associations—including the Boston Edison Company, the Edison Electric Institute, and the Raytheon Company—Dr. Avila is a corporator of the Boston Five Cents Savings Bank and the Milton Savings Bank. He is a member of the executive, national, or coordinating committees of many professional organizations, as well as a trustee of Northeastern University and a member of the National Academy of Engineering. A registered professional engineer, he is also a noted designer of high-altitude military cameras, and was awarded the 1968 IEEE Edison Medal "for his early contributions to underground transmission, for his continuing guidance in the field of electrical research, and for his positive leadership in the development of the electrical utility industry."

Andrew F. Corry (SM), born in Lynn, Mass., received the B.S. degree in electrical engineering from the Massachusetts Institute of Technology at Cambridge in 1947. A member of the Boston Edison Company since that time, he has been active in overhead and underground transmission and distribution design, testing, and standardizing activities and research. Formerly responsible for system planning, he is currently assistant to the executive vice president of Boston Edison. Mr. Corry's major technical field has been design and application of high-voltage and extra-high-voltage underground cable, and he has published a number of papers involved with underground equipment application and design. A past Chairman of the Insulated Conductors Committee of the IEEE, he is at present Chairman of the Electric Research Council Steering Committee on Underground Transmission.

The low-power-drain microelectronic VHF amplifier

It is not always preferable to make miniaturized circuits by applying thin-film, thick-film, or monolithic techniques. This can be the case for many VHF micropower device applications—although more is involved than transistor design scale-down when applying discrete technology to miniature circuits

S. Cho, A. G. Thiele Motorola Inc.

Low-power-drain VHF amplifiers have some unique characteristics when they are miniaturized. And, if the design is special—requiring tight tolerances or intended for limited-quantity production—it is often desirable to make the RF micropower circuits from discrete components. Described in this article are the first-order device requirements and circuit considerations for such VHF amplifiers incorporating bipolar transistors operating at collector currents as low as 25 μA. Two specific designs are offered, along with information pertinent to their construction.

Recently, much attention has been focused upon devices and circuits capable of high-quality VHF performance at microwatt dc power-drain levels. A natural extension of this interest encompasses the realization of microelectronic, low-power-drain VHF circuit functions in a small-package volume.

Radio-frequency micropower devices possess unique electrical characteristics due to low operating currents and ultrasmall device geometries. At the outset, then, we will explain device characteristics and their influence on circuit performance (which must be understood by designers of stable low-power-drain VHF circuitry).

The discrete microcomponent approach is often to be preferred over thin-film, thick-film, or monolithic fabrication techniques—being, for example, more economical for limited-quantity circuits and more suitable for circuits requiring tight component tolerance.

Micropower design considerations

A basic performance limitation of a bipolar transistor in low-power-drain applications results from two effects: There is a decrease in device input and output conductances and there is a deterioration of transconductance g_m corresponding to the reduced emitter current (microamperes instead of milliamperes).

In developing first-order expressions relating device parameters to low-current VHF-amplifier performance, the simplified hybrid-π transistor model of Fig. 1 is employed. Model parameters g_m, r_π, and C_π relate, respectively, to emitter diffusion resistance r_e' (and emitter bias current I_E), low-frequency, common-emitter current-gain β_0, and common-emitter, current-gain–bandwidth product $\omega_T \equiv 2\pi f_T$. (See illustration.) Parameter r_b' represents the lumped-equivalent base-spreading resistance and C_μ is the collector–base junction capacitance.

General expressions for the small-signal, two-port parameters of a device—in terms of the hybrid-π parameters—are somewhat complex. However, for small-signal operation of a high-frequency, ultrasmall-geometry device at submilliampere emitter currents (say, 25 μA $< I_E <$ 500 μA), the following inequalities are usually satisfied: $r_\pi > 1$ kΩ, $r_b' < 50$ Ω, 1 mmho $< g_m <$ 20 mmho, $C_\pi < 2.5$ pF, $C_\mu < 0.5$ pF.

Under these conditions, and at VHF ($\omega < 2 \times 10^9$ rad/s), first-order expressions for the small-signal y-parameters of a device may be written as*

$$y_{11} \approx \frac{1}{r_\pi} + j\omega(C_\pi + C_\mu) \qquad (1)$$

$$y_{12} \approx -j\omega C_\mu \qquad (2)$$

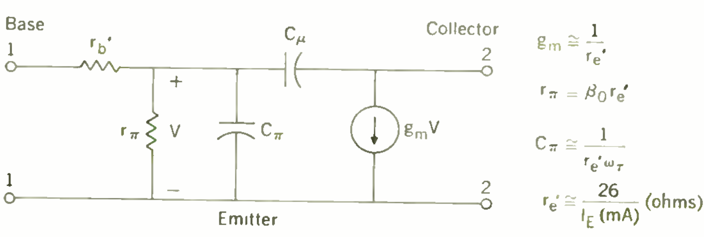
$$y_{21} \approx g_m - j\omega C_\mu \qquad (3)$$

$$y_{22} \approx r_b'\omega^2 C_\mu^2 + j\omega C_\mu(1 + r_b' g_m) \qquad (4)$$

Using Eqs. (1) to (4), first-order expressions for the matched, unilateral power gain[1] G_U, and Linvill's stability factor[2] c can be written as follows:

* According to convention, for a four-terminal network, the admittances are defined as:
y_{11} (= $g_{11} + jx_{11}$) is the admittance measured at terminals 1-1 with terminals 2-2 short-circuited.
y_{22} (= $g_{22} + jx_{22}$) is the admittance measured at terminals 2-2 with no load at the output and terminals 1-1 short-circuited.
y_{12} and y_{21} determine the short-circuit current at one pair of terminals when a unit voltage is applied at the other pair.

FIGURE 1. Simplified hybrid-π transistor model.



$$G_U \equiv \frac{|y_{21} - y_{12}|^2}{4(g_{11} + g_{12})(g_{22} + g_{12})} \approx \frac{g_m\beta_0}{4r_b'\omega^2C_\mu^2} \quad (5)$$

$$c \equiv \frac{|y_{12}y_{21}|}{2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21})} \approx \frac{g_m}{\omega C_\mu} \quad (6)$$

The value g_m is fixed for a given emitter bias current. An RF micropower device that has small r_b' and C_μ and a large β_0, by Eq. (5), appears to be preferred for a high-gain amplifier.

Intuitively, if the value of the feedback capacitance C_μ approaches zero, the stability of a device should improve; namely, Linvill's c factor should approach zero. On the contrary, according to Eq. (6), Linvill's c factor approaches infinity as C_μ approaches zero. This surprising behavior results from the fact that g_{22}, at low current levels, is roughly proportional to C_μ^2. This high, potential instability must be taken into account when designing a low-power-drain amplifier.

Reduced C_μ is desirable

Don't conclude, however, that the reduction of C_μ is not desirable in a high-performance, low-current RF device. In fact, the smaller the feedback capacitance C_μ is, the higher the available gain becomes for a given degree of stability. (This point will be illustrated in the following.)

Although there are various methods of representing amplifier stability, the alignability factor[3] k will be used in the present discussion.

$$k \equiv \frac{|y_{12}y_{21}|}{2(g_{11} + G_S)(g_{22} + G_L)} \quad (7)$$

where G_S and G_L, respectively, represent the source and load conductances seen by the device. In a well-designed multistage amplifier, the value for k should be 0.2 or less in order to permit a reasonable number of cascades.

Consider the single-stage amplifier of Fig. 2. Since the value of k for the device alone is, typically, very high at a low level of emitter bias current, loading conductances g_{11}' and g_{22}' are added (as shown) to reduce k to a reasonable value. Assuming that $g_{11}' \gg g_{11}$ and $g_{22}' \gg g_{22}$, and that the amplifier input and output ports are conjugately matched as shown in Fig. 2, first-order expressions for the alignability factor k and the amplifier transducer gain[4] G_T can then be written as

$$k \approx \frac{g_m\omega C_\mu}{8g_{11}'g_{22}'} \quad (8)$$

$$G_T \equiv \frac{4G_SG_L|y_{21}|^2}{|(y_{11} + Y_S)(y_{22} + Y_L) - y_{12}y_{21}|^2} \approx \frac{g_m^2}{4g_{11}'g_{22}'} \quad (9)$$

Equations (8) and (9) may then be combined to yield

$$G_T \approx \frac{2kg_m}{\omega C_\mu} \quad (10)$$

This expression clearly indicates why it is desirable to minimize the collector-base-junction capacitance C_μ in a device used in low-power-drain amplifier circuitry. In particular, Eq. (10) shows that reduction of C_μ is the only way to increase G_T for a specified alignability factor, emitter bias current, and frequency.

Mismatched output

Next, consider the single-stage amplifier of Fig. 3. Here, stabilization is achieved by output mismatching (with the input conjugately matched) without the use of loading conductances at the device terminals. For the situation shown in Fig. 3, assuming such a sufficient output mismatching ($G_L \gg g_{22}$) that k is again reduced to a reasonable value, first-order expressions for k and G_T are

$$k \approx \frac{g_m\omega C_\mu}{4g_{11}G_L} \quad (11)$$

$$G_T \approx \frac{g_m^2}{g_{11}G_L} \quad (12)$$

$$\approx \frac{4kg_m}{\omega C_\mu} \quad (13)$$

This last expression again indicates the desirability of minimizing C_μ, as previously discussed.

Note that Eqs. (10) and (13) imply that, for the low-current operating conditions and range of device-parameter values under consideration, the available transducer gain for a given device at a specified operating current, frequency, and alignability is independent of the device gain-bandwidth product f_T. This conclusion contradicts conventional design criterions for device application at higher current levels, but has been qualitatively verified by experiment.

It should also be noted that Eqs. (10) and (13) are identical except for the proportionality constant. This difference implies that the available gain with stabilization by mismatch at the output (Fig. 3) is 3 dB higher than that with stabilization by external loading conductances (Fig. 2)—for a given device, alignability factor, bias condition, and operating frequency.

Although the gain expressions given by Eqs. (10) and (13) are of first-order accuracy for simple, single-stage, RF micropower amplifiers, they can be used as a very convenient gain guide for a low-current RF device.

In an amplifier where neutralization is employed, a gain much higher than that predicted by Eq. (13) can be obtained for a given alignability factor. This condition exists because neutralization effectively reduces the feed-

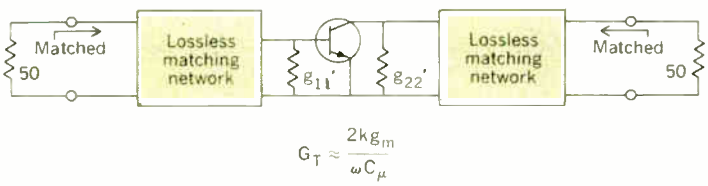
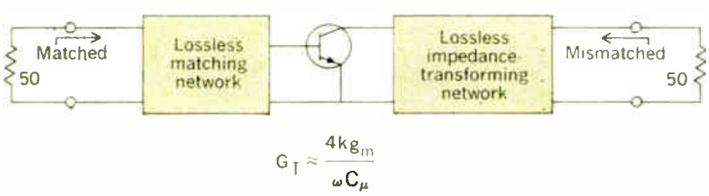


FIGURE 2. Single-stage amplifier with stabilization by conductance loading of both the device output and input ports. G_T is transducer gain.

FIGURE 3. Single-stage amplifier with stabilization by mismatching the device output port.



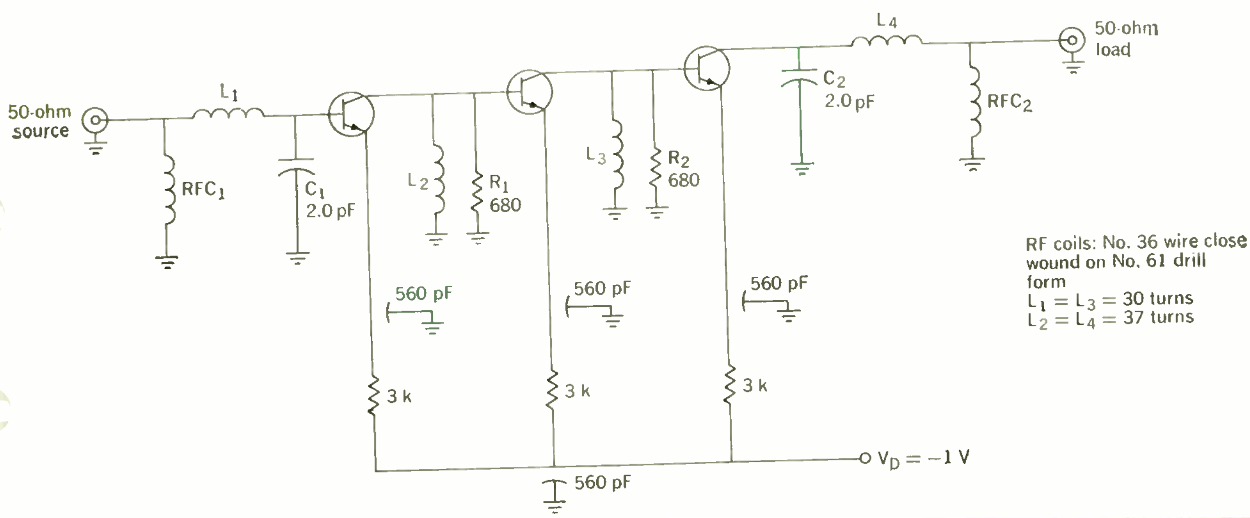


FIGURE 4. Schematic of a three-stage, wide-band, low-power-drain 150-MHz amplifier. Active devices are commercially available Motorola SMT807 ultrasmall-geometry n-p-n silicon transistors (each biased at $I_E \approx 100\ \mu A$ and $V_{CE} = V_{BE} \approx 0.7$ volt by the emitter biasing scheme shown). Interstage networks are direct-coupled. Source and load impedances are 50 ohms.

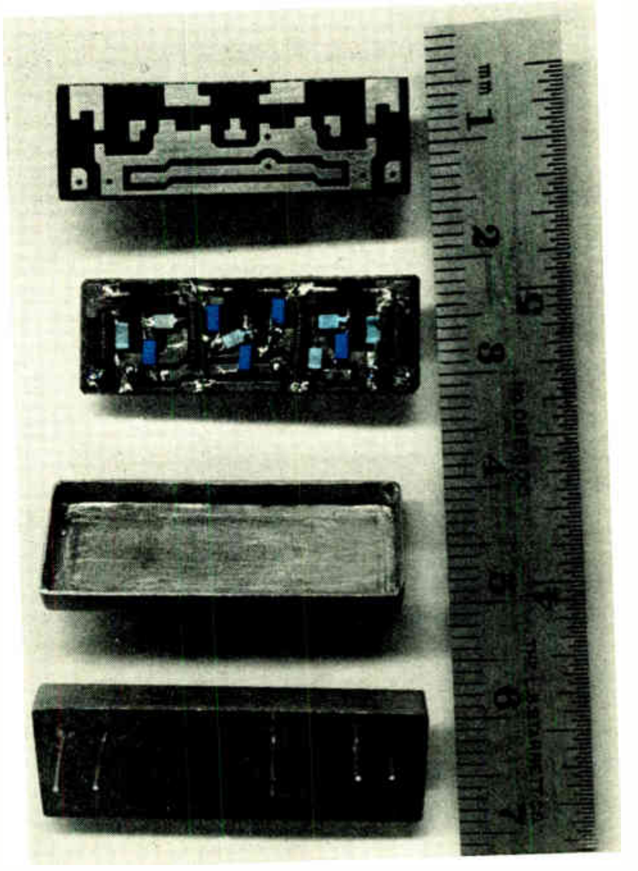


FIGURE 5. Assembly sequence and details for a microdiscrete three-stage wide-band amplifier. From top: baseboard, component assembly, housing, and completed amplifier. The Motorola SMT807 transistor dice (dark disks) are encapsulated in Micro-T miniature packages (2.16 mm diameter × 1.4 mm thick). Resistors (solid color) are Bourns, Inc., cermet microresistors (2.54 × 1.27 × 0.76 mm). Capacitors (tint) are Electro Materials Corp. ceramic chips (2.41 × 1.27 × 1.52 mm). Inductors are single-layer, air-core solenoids with a 1.03-mm ID. Circles are connection pins.

back capacitance between base and collector over a limited bandwidth. Thus, neutralization may be employed in trading off bandwidth for mid-band gain.

If an amplifier has not been neutralized, a higher gain can generally be obtained by mismatching than by conductance loading for a given alignability factor.

So far we have considered only the gain and the alignability factor of a micropower RF amplifier. It is also convenient to have some guide by which the bandwidth capability of an RF micropower device can be judged. This can be a special need in the design of a high-gain multistage RF low-power-drain amplifier that must have a relatively broad bandwidth. In a high-gain low-current amplifier, interstage resistance levels are, per se, very high.

First-order expressions for the input and output capacitances of a device are useful in estimating amplifier-bandwidth capability. In particular, for a device that is not neutralized, in low-current application (assuming a sufficiently low value for alignability factor k such that interaction between device input and output is negligible), suitable approximations to the device input and output capacitances are, from Eqs. (1) and (4),

$$C_{in} \approx C_\mu + C_\pi \tag{14}$$

$$C_{out} \approx C_\mu(1 + r_b'/r_e') \tag{15}$$

In low-current application, C_π is usually the dominant bandwidth-limiting capacitance and since

$$C_\pi \approx 1/2\pi f_T r_e' \tag{16}$$

the desirability of high-gain–bandwidth product f_T for broadband-micropower-amplifier application is evident.

Microelectronic applications

We will illustrate the utility and demonstrate the performance of the microdiscrete (discrete microcomponent) approach to the fabrication of linear microcircuitry in a small package volume by two specific examples.

Wide-band VHF amplifier. The three-stage amplifier shown schematically in Fig. 4 was designed to provide a 3-dB bandwidth in excess of 30 MHz at a nominal 150-MHz center frequency with a total power drain of 300 μW (300-μA drain from a single 1-volt supply). Mid-band transducer gain was to be the maximum attainable (consistent with the bandwidth and power-drain constraints), and the amplifier was to be operable over $-40°$ to 55°C.

From Fig. 4, the amplifier is seen to comprise five resistors, six capacitors, six inductors, and three active devices. Alignability and broadbanding are provided by (1) interstage loading resistors R_1 and R_2, and (2) by deliberate mismatching at the amplifier input and output ports. Radio-frequency chokes RFC_1 and RFC_2 provide ground returns for transistor bias currents and may be eliminated with no appreciable effect on amplifier performance if the 50-ohm source and load terminations are direct-coupled.

Fig. 5 shows assembly details for a small-volume microdiscrete realization of the three-stage wide-band amplifier of Fig. 4. The relative sizes and positions of the active and passive components are apparent. The assembly baseboard is 10.3 × 30.5 × 0.79 mm and is made of double, copper-clad epoxy laminate. All components are soft-soldered in place on the baseboard (after gold-plating of the etched interconnect pattern), although a combination of split-electrode welding and preform soldering is probably preferable for moderate production.

Final package volume is 1.15 cm^3 (0.07 in^3). Some reduction in package volume would be possible if thin-film components were substituted for the chip resistors and capacitors; however, any significant reduction in package length would compromise amplifier bandwidth and alignability, due to coupling among the inductors. Note that the two optional RF chokes (Fig. 4) have been deleted from the prototype assemblies shown.

Typical measured parameters of the Micro-T-packaged SMT807 devices used in the wide-band VHF amplifier are $C_\mu \approx 0.4$ pF, $r_b' \approx 40$ ohms, and $f_T = 460$ MHz—at the $I_E \approx 100$-μA, $V_{CE} \approx 0.7$-volt bias condition employed. Hence, $r_e' \approx 250$ ohms, $g_m \approx 4$ millimhos, and $C_\pi \approx 1.4$ pF. Each stage of the wide-band VHF amplifier was designed (by suitable choice of device terminations) to have an alignability factor $k \approx 0.2$. Stabilization is effected by input mismatch and output loading in the first stage ($G_{T_1} \approx 4kg_m/\omega C_\mu$), by input and output loading in the second stage ($G_{T_2} \approx 2kg_m/\omega C_\mu$), and by input loading and output mismatch in the third stage ($G_{T_3} \approx 4kg_m/\omega C_\mu$). Hence, the first-order prediction of the mid-band transducer gain of the three-stage 150-MHz amplifier should be

$$G_T \approx 32 \left(\frac{kg_m}{\omega C_\mu}\right)^3 \approx 25 \text{ dB}$$

From Eqs. (14) and (15), $C_{\text{in}} \approx 1.8$ pF and $C_{\text{out}} \approx 0.5$ pF. Assuming 0.7-pF stray capacitance from each interstage node to ground (including distributed capacitance across interstage resistors and inductors), net interstage capacitance is approximately 3 pF. Since the transistor input and output conductances are negligible relative to that of the 680-ohm interstage loading resistor, interstage bandwidth is, therefore, approximately 78 MHz. Moreover, the amplifier is synchronously tuned, with all four tuned circuits having approximately equal bandwidths, producing an appropriate bandwidth shrinkage factor of 0.435.[5] Hence, a first-order prediction of the 3-dB bandwidth of the three-stage, 150-MHz amplifier should be approximately 0.435 × 78 MHz ≈ 34 MHz.

Figure 6 is a graph of measured transducer gain vs. frequency for a typical microdiscrete prototype assembly of the three-stage wide-band 150-MHz amplifier. Mid-band transducer gain is 23.5 dB and 3-dB bandwidth is 31.8 MHz—in surprisingly close agreement with the preceding first-order predictions. The dc power drain is typically 300 μW from a single -1.0-volt supply. Typical

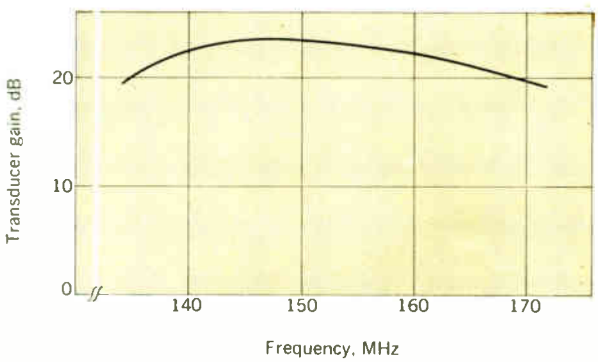
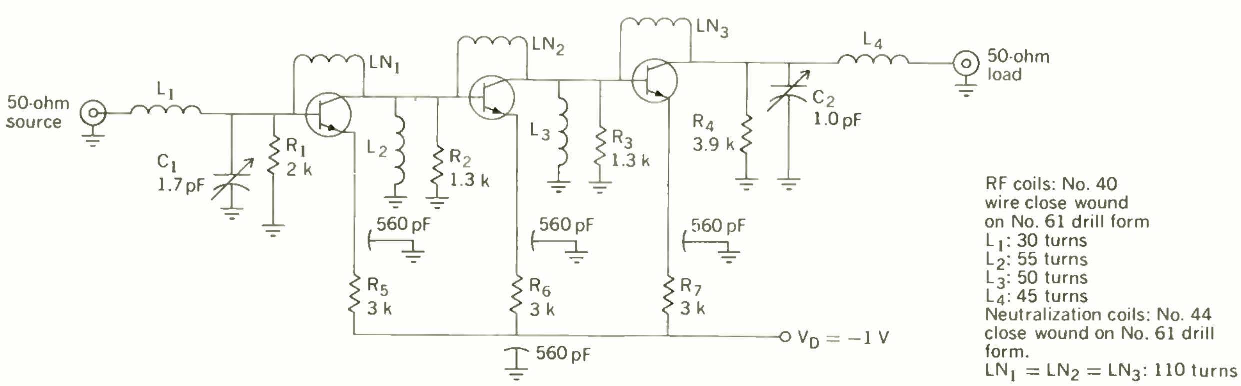


FIGURE 6. Gain vs. frequency for typical microdiscrete three-stage wide-band amplifier shown in Figs. 4 and 5. Parameters are $G_{T(max)}$ = 23.5 dB, bandwidth = 31.8 MHz, V_D = −1 volt, and P_D = 300 μW.

FIGURE 7. Schematic of a three-stage neutralized low-power-drain 150-MHz amplifier. As for the wide-band amplifier shown in Fig. 4, interstage networks are direct-coupled; emitter biasing scheme is also as in Fig. 4.



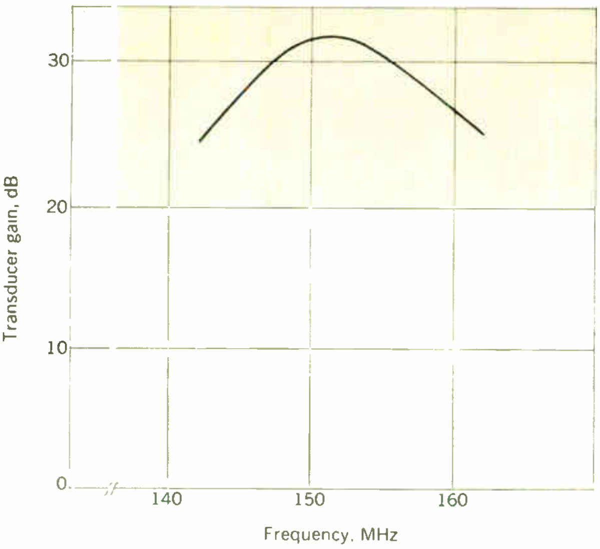


FIGURE 8. Gain vs. frequency for typical microdiscrete three-stage neutralized amplifier. Parameters are $G_{T(max)} = 31.8$ dB, bandwidth = 11.4 MHz, VSWR over bandwidth < 2.5:1, $V_D = -1$ volt, $P_D = 300\ \mu W$.

total variation of center frequency, mid-band gain, and 3-dB bandwidth over a −40° to 55°C operating temperature range are 6.1 MHz, 4.5 dB, and 0.8 MHz, respectively. Typical dynamic range (from noise level to 3-dB gain-compression point) is 49 dB.

Neutralized VHF amplifier. The three-stage amplifier shown schematically in Fig. 7 was designed to provide a mid-band transducer gain in excess of 30 dB at 150 MHz with the power requirements cited previously for the amplifier shown in Fig. 4.

The amplifier, Fig. 7, consists of seven resistors, six capacitors, seven inductors, and three active devices. Alignability is provided by neutralization coils L_{N1} through L_{N3}, which effectively provide narrowband unilateralization of the active devices; broadbanding is provided by loading resistors R_1 through R_4. The amplifier input and output ports are matched to 50-ohm source and load.

Assembly details for the small-volume microdiscrete realization of the three-stage neutralized amplifier are identical in essence to those for the three-stage wide-band amplifier described previously. Component types and sizes are identical, and the physical component layout is very similar to that shown in Fig. 5, with the notable addition of the three, 110-turn, single-layer neutralization coils between the base and collector leads of the three Micro-T-packaged transistors. Final package volume is approximately 25 percent greater than that for the wide-band amplifier, to minimize the effect of the metallic housing on amplifier alignment.

Figure 8 is a graph of measured transducer gain vs. frequency for a typical microdiscrete prototype assembly of the three-stage neutralized 150-MHz amplifier. Mid-band transducer gain is 31.8 dB and 3-dB bandwidth is 11.4 MHz, illustrating (relative to the previous three-stage wide-band amplifier) the effectiveness of neutralization in trading bandwidth for gain. Typical total variation of center frequency, mid-band gain, and 3-dB bandwidth over a −40° to 55°C operating temperature range are 2.3 MHz, 4.2 dB, and 0.6 MHz, respectively. Typical dynamic range (from noise level to 3-dB gain-compression point) is 46 dB, and input and output voltage standing-wave ratios are typically 1.5 to 1 at mid-band.

Conclusions

State-of-the-art ultrasmall-geometry bipolar transistors are capable of high-quality VHF performance at microwatt levels of dc power drain. However, conventional transistor amplifier design criterions are not generally suitable for relating active-device parameters to low-current VHF amplifier performance.

We believe that the microdiscrete approach offers the following advantages over thin-film, thick-film, and monolithic approaches:

1. Individual components may be selected prior to assembly, if and as dictated by tolerance requirements.
2. Circuits may be "repaired" (individual components replaced), if necessary, thus ensuring very high assembly yield.
3. Minor redesign may be readily and quickly effected.
4. Layout and small-quantity production costs are appreciably lower.

REFERENCES

1. Hejhall, R., "RF small signal design using admittance parameters," Applicatien Note AN-215, Motorola Semiconductor Products Inc., pp. 5–6.
2. *Ibid.*, p. 3.
3. Thornton, R. D., *et al.*, *Multistage Transistor Circuits.* New York: Wiley, 1965, pp. 232–233.
4. Ghausi, M. S., *Principles and Design of Linear Active Circuits.* New York: McGraw-Hill, 1965, pp. 62–65.
5. Thornton, R. D., *et al.*, *op. cit.*, p. 256.

Susumu Cho (M) received the B.S. degree in electrical engineering in 1963 from Tokyo Institute of Technology and the M.S. degree in electrical engineering in 1966 from the University of California, Berkeley. He has been employed by Motorola, Semiconductor Products Division, Phoenix, Ariz., since 1966, where he is currently a project engineer in the Electronics Section of the Central Research Laboratories. There, he is engaged in development of microwave and micropower integrated circuits and evaluation of developmental microwave power transistors.

Al Thiele (M) received the B.S. and the M.S. degrees in electrical engineering in 1961 and 1962, respectively, from the University of California at Berkeley. Since 1964, he has managed the Electronics Section of the Central Research Laboratories, at Motorola, Semiconductor Products Division, Phoenix, Ariz. His responsibilities include evaluation of developmental semiconductor devices, and development of microwave and micropower integrated circuits. Mr. Thiele has had three papers published, and is the author of the section on "Low-Frequency Amplifiers" in the Handbook of Semiconductor Electronics. He is a member of Phi Beta Kappa, Eta Kappa Nu, Tau Beta Pi, and Phi Eta Sigma.

FIGURE 3. A corner of the Faraday Room, showing a portion of the recent special exhibition entitled, "Man's Conquest of the Moon."

FIGURE 4 (below). The "Leonardo Room," containing many of the great pictorial folios of the master and a collection of 200 volumes of Vinciana. On the walls are two mural enlargements of some of his well-known drawings: a siege engine (left), and sketches of various structural elements and heavy fortifications.

an operating industrial organization engaged in the design and manufacture of electric equipment. The decentralized collection restricted group visits and sometimes made the use of books difficult for researchers.

Plans were made in 1961 for the design and construction of a building whose entire occupancy would be devoted to the library collection. The architectural firm of Sherwood, Mills & Smith, of Stamford, Conn., was retained to design the present structure and to provide all the necessary air-conditioning, humidity, and temperature controls to ensure the preservation of the Library's contents for the next century. The building, which harmoniously blends contemporary architectural design and structural materials with classical symmetry and the graceful lines of the Italian Renaissance arch fenestration, was completed early in 1964 and dedicated in May of that year (Fig. 1). The design won an architectural award in 1964.

The contents of the Library

General arrangement: reading room and mezzanine. The main reading room on the ground floor (Fig. 2) contains the collection of works in the general sciences, arranged alphabetically by subject—literally from "A" to "Z"—in separate glass-front cases, ranging from astronomy, botany, chemistry, geology, mathematics, medicine, physics, etc., down through zoology. Within each case the works are generally arranged alphabetically by author. There are presently about 22 000 volumes in the category of the general sciences.

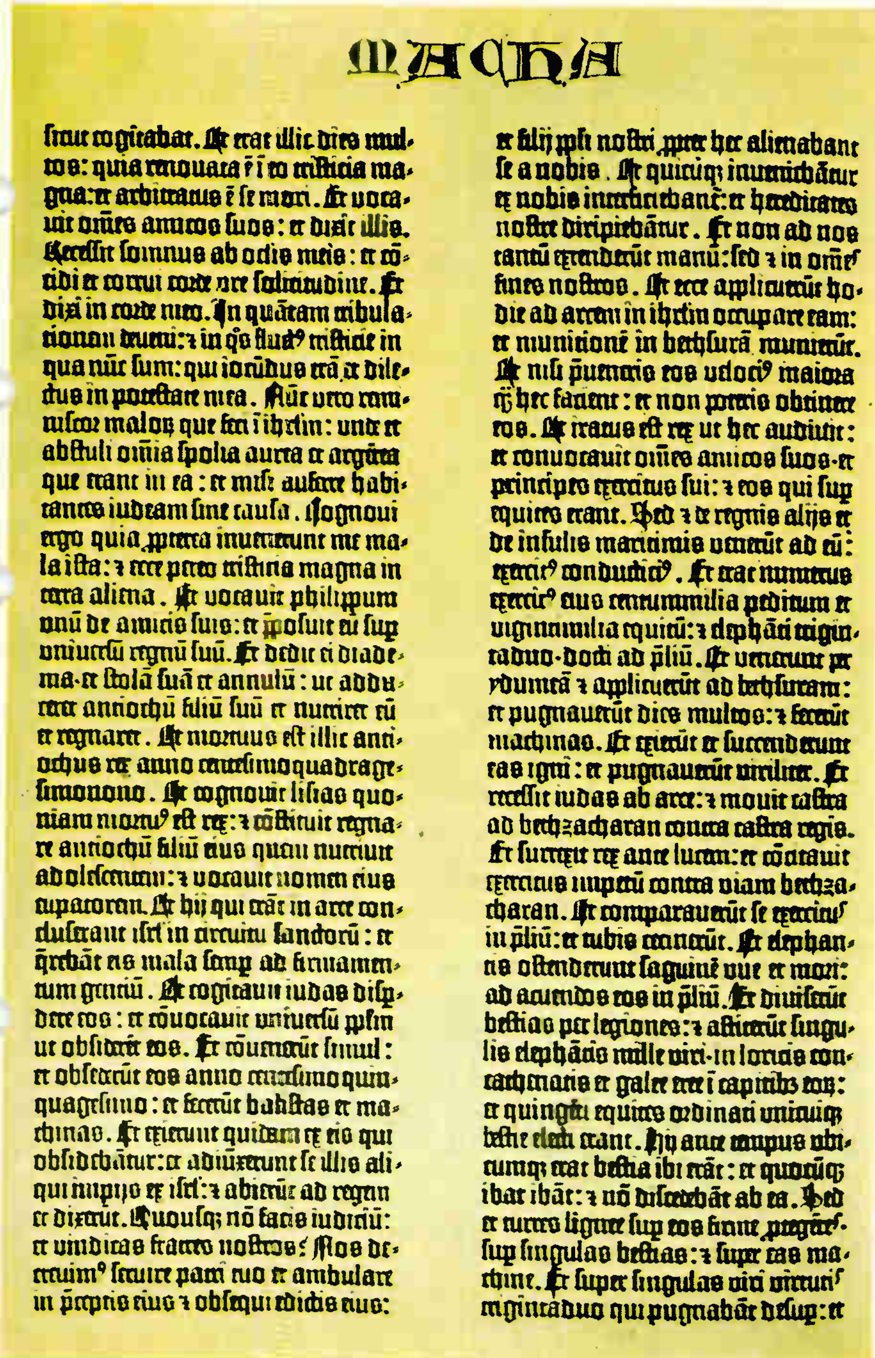


FIGURE 5. A historical example of the beginning of the "information explosion": a leaf of the Gütenberg Bible of 1455, marking the first use of movable type in printing (property of the Burndy Library). The text is in Latin.

The mezzanine space is devoted almost exclusively to some 8000 volumes on electricity and magnetism. Here, too, the books are arranged alphabetically by author. This constitutes the most comprehensive collection of books and papers in early electricity and magnetism.

The exceptions to these arrangements, in both the main reading room and on the mezzanine, are the special bookcases that are reserved for works on or by some of the greatest names in science: Newton, Einstein, Edison, Faraday, Franklin, Galvani, and Volta. More than 300 volumes from the library of Alessandro Volta (1745–1827) are in the special cases devoted to the works of this pioneer in electrical science.

The shelves at the south end of the main reading room are of unusual interest. Here may be found, for example, a complete issue of the *Philosophical Transactions of the Royal Society of Great Britain*, year by year, from 1667 (the time of Newton, Boyle, Hooke, and Wren) to 1939—almost three centuries of scientific papers published by that great institution.

Similarly, there is a run of the *Journal des Savants* of the Academie Française (the French analog of the Royal Society) from 1665 to 1792, when the Academie was dissolved by the Directorate of the French Republic following the overthrow of Louis XVI. These beautifully bound volumes are supplemented by a run of the *Histoire et Memoirs de l'Academie des Sciences* (1666–1790). In this same area of the reading room are stored 33 first-edition volumes of the 44 tomes of the *Histoire Naturelle*, written by the great French natural scientist the Comte de Buffon (whose son, with Lavoisier, was a tragic victim of the excesses of the French Revolution).

Rounding out the foreign runs of scientific works are collections of the Italian *Scelta di Opuscoli* (selected papers) from 1775 to 1807; *Memorie di Mathematica e Fisica* (mathematical and physical memoirs), 1782 to 1809; and *Archivio di Storia della Scienza* (archive of the history of science, published from 1919 to 1938).

One of the most interesting cases is devoted to Oriental science and technology, principally Japanese works of the 17th, 18th, and 19th centuries. This insight into an aspect of the culture and society of feudal Japan (going back more than 150 years before Matthew Perry's expedition to the Far East) reveals the influence of European science in those times. There are Japanese volumes in which numerous beautiful engravings on rice paper show Archimidean water screws, geared hoisting machinery, water wheels, etc., that were used in irrigation projects and in metallurgy. These devices must have been introduced to the Japanese by the Portugese traders and missionaries who were granted limited entry to Japan from 300 to 200 years ago. Among the valued museum exhibit pieces is a four-sectional chemical retort, or alembic, made of glazed terra cotta in Japan about 1770. There are also two manuscript painted scrolls, each nearly 12 meters long; they show scenes in the mining and smelting of gold and silver.

In addition to numerous publications of German scientific societies and institutions, one may find biographies of noted scientists published in English, French, German, Italian, and Latin. And one separate bookcase contains a representative collection of scientific and technical literature published in Russian.

First editions. The Library has an unusually large number of first editions of great works by famous scientists, such as Darwin's *On the Origin of Species* (1859); Newton's *Principia* (1687), *Opticks* (1704), and *Fluxions* (1711); Huygens' *Horologium Oscillatorium* (1673); Vesalius' *De Humani Corporis Fabrica* (1543); Harvey's *Exercitatio Anatomica de Motu Cordis* (1628); Copernicus' *De Revolutionibus Orbium Coelestium* (1543); Kepler's *Astronomia Nova* (1609); Tycho Brahe's *Epistolarum Astronomicarum* (1596, the author's copy); Hevelius' *Selenographia* (1647); Boyle's *The Sceptical Chymist* (1661); Volta's *Novus ac Simplicissimus Electricorum Tentaminum* (1771, Volta's copy), and *Del Modo di Render Sensibilissima la Più Debole Elettricità* (1782, Volta's copy); Galvani's *De Viribus Electricitatis in Motu Musculari* (1791); Faraday's *Experimental Researches in Electricity* (1832–1852, Tyndall's copy); Henry's *On the Application of the Principle of the Galvanic Multiplier to Electro-Magnetic Apparatus* (1831); Descartes' *Discours de la Methode* (1637); and Boyle's *De Coloribus* (1665, Lavoisier's copy). These are but a representative interdisciplinary sampling of first editions, which are not merely interesting to book lovers; they are essential equipment for scholarly study of the history of science, because later editions frequently reflect concepts and prejudices characteristic of the translator or editor, not of the author.

The 'incunabula.' The first great breakthrough in communication via the printed word came with the initial use of movable type by the German printer Johannes Gütenberg (c. 1455). All books printed prior to the year 1501 are arbitrarily referred to as *incunabula* (from the Latin "cradle" or "birth"). In this brief period between 1455 and 1500, the first "information explosion" occurred, when more than six million volumes were published by the then primitive processes of printing one leaf side at a time and stitching. The art of interrogating Nature by means of experiments, as practiced in modern science, was virtually unknown in this period of the Renaissance. Most of the valid scientific and technical knowledge of the time was based upon the records of the ancient Egyptian, Greek, and Roman astronomers, mathematicians, physicians, and engineers. The remainder was a heterogeneous mix of quasi-scientific thought, general philosophy, and erroneous assumptions predicated upon astrology, alchemy, religious dogma, and just plain superstition. Therefore, it is not surprising that scientific books made up a very small percentage of the total incunabula.

Nevertheless, the Burndy Library collection of scientific incunabula numbers more than 300 of these rare volumes, many of which were written by the greatest minds in medieval and Renaissance science: Regiomontanus (astronomy); Alhazen and Paccioli (mathematics); Mesue, Rhases, and Savonarola (medicine); Thomas Aquinas, Roger Bacon, Albertus Magnus, and Paludanus (philosophy); and Brunschwig (chemistry).

Lower level. On the lower floor are three rooms: a large hall (the Faraday Room) for special exhibitions (Fig. 3), the Leonardo Room, and the Sarton Room. In the Leonardo Room (Fig. 4), the Vincian collection of books, folios, and facsimile manuscripts of the master is kept in special cases. Here, mural enlargements of Leonardo's famous drawings of the geometric proportions of the human body, the art of cannon manufacture, fortifications and structures, a siege engine, and a botanical study are prominently displayed. This room is used as a lounge, a meeting place for small seminars, or a secluded area for scholars engaged in Vincian research.

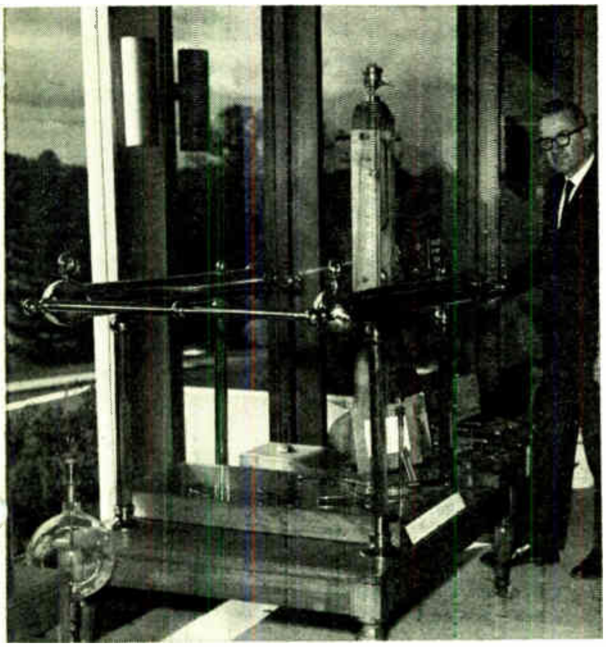
The Sarton Room, which is dedicated to the memory of the late Dr. George Sarton, the noted Harvard historian of science, is used for group meetings and lectures. Similarly, the main reading room upstairs is equipped with a concealed blackboard and built-in projection screen for visual displays. Portable chairs can be set up to accommodate almost 200 people at major technical meetings.

The Library as a museum

The Burndy Library contains several glass-topped exhibit cases in the main reading room and on the mezzanine, in which artifacts, antique instruments, memorabilia, rare books, and documents are kept on permanent and semipermanent display. In these cases one may find, for example, a printed leaf (Fig. 5) of the Gütenberg Bible of 1455; an Assyrian clay tablet, dating back to 1800 B.C., which bears a cuneiform inscription dealing with a consignment of wool; an incunable copy of the first printed work in science (Pliny's *Historia Naturalis*), published in Venice in 1469; a rare copy of Dr. William Harvey's *De Motu Cordis*, in English, published in 1653 (from the library of Dr. Harvey Cushing, the famous neurosurgeon of Yale University); Charles Steinmetz' working diary and log book of his early electrical engineering projects (c. 1888); one of ten surviving manuscript sheets of *Origin of Species* (the frugal Darwin used the manuscript of the book for household purposes after

FIGURE 6 (right). A mezzanine display case contains combination pocket compasses and sundials, and lodestones. These instruments and early magnets were made in England, France, Germany, and the Middle East from 1550 to the early 1800s.

FIGURE 7 (below). The "Granddaddy" of the Library's electrostatic generators—the de Saussure machine, built in Paris in 1805. Note the heavy brass busbars (which also serve as capacitors). The big machine is capable of discharging a spark up to 20 cm long.



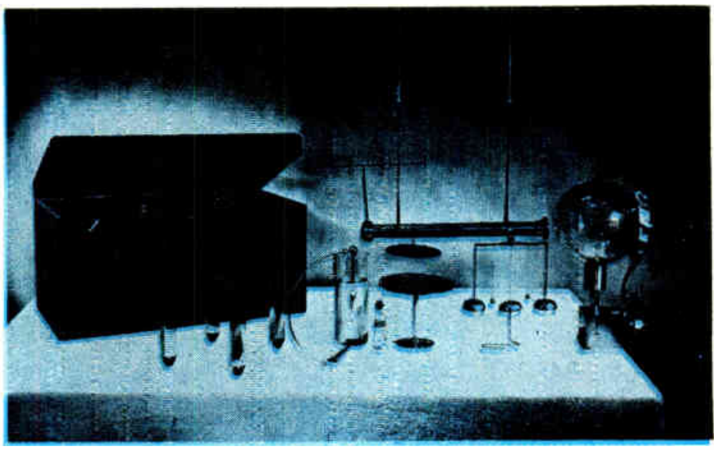


FIGURE 8. The Nairne Electrical Machine, with its traveling case and demonstration apparatus. Made in 1773, for showing the "electric phenomenon," this compact generator could be set up to make bells ring and cause vanes to rotate and sparks to jump across gaps—all for the amusement of 18th century spectators.

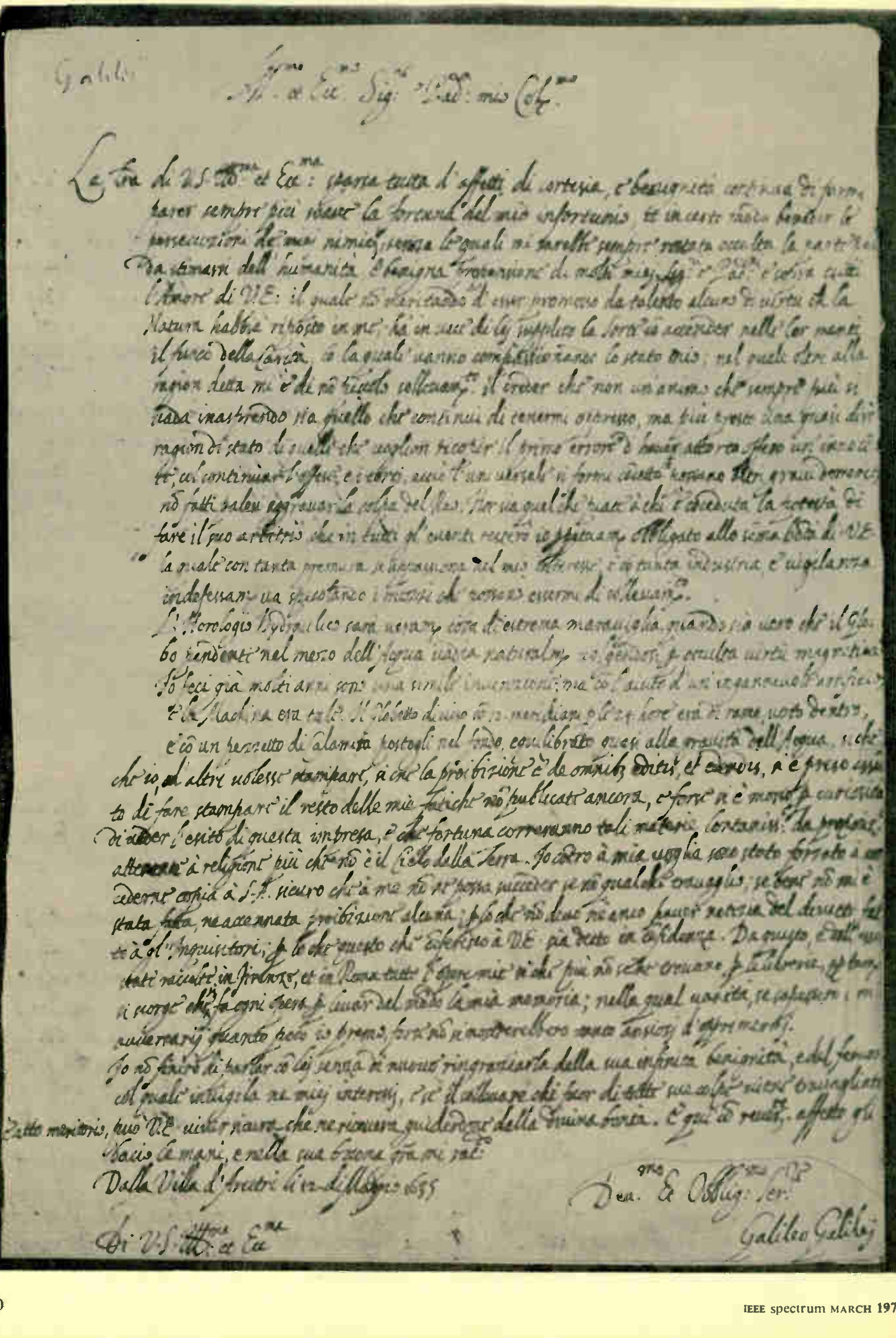
the type was set); Guglielmo Marconi's typed Nobel Prize oration of 1909 (with many drawings and autograph corrections); and a collection of lodestones, and of combination pocket compasses and sundials of European and Arabic origin, dating from 1550 to the early 1800s (Fig. 6).

Electrostatic generators. Several authentic electrostatic generators of the early "bottle and brush" and later disk types are kept on permanent display on the mezzanine. The oldest of these is a bottle-type generator made in Venice in 1740, and the largest is the De Saussure glass-disk machine (Fig. 7), manufactured in Paris in 1805. Most of these machines are in working order. Supplementing this collection are compact portable generators and their ancillary apparatus, such as the Nairne Electrical Machine shown in Fig. 8, which were used in traveling demonstrations of the "electric phenomenon" during the latter part of the 18th century. Excellent original examples of "Leyden jars" (the first practical capacitors) are also prominently on view.

Instruments, optical and otherwise. The Burndy Library is privileged to possess an impressive number of early telescopes, microscopes, surveyors' levels and transits, and sextants, some of which were built by famous instrument makers of the 18th and 19th centuries. Located in the Faraday Room and atop the bookcases on the mezzanine are examples of the earliest refracting telescopes to incorporate the Newtonian mirrors for increasing the light-resolving properties of the lens system. One of the telescopes was made by James Short in 1744; another by Addison Smith in 1750.

In 1959, the Library received a significant addition to its instrument holdings by the acquisition in Holland of a large portion of the Groenendijk Collection—some 150 devices and laboratory apparatus that a group of amateur scientists, called the "Society Felix Meritis," gathered from the time of the society's founding in 1777. In addition to Morse-type telegraph keys, relays, coils, etc., the collection contains many components of the first indicator telegraph systems developed in England by Cooke and Wheatstone.

Another remarkable display in the Faraday Room is the large Westinghouse–Hibben collection of early incandescent bulbs and vacuum tubes. Consisting of more



Galilei

Ill.mo et Ecc.mo Sig.re e Pad.: mio Col.mo

La lettera di V. S. Ill.ma et Ecc.ma [illegible] tutta d'affetti di cortesia, e benignità continua di [illegible] parer sempre più [illegible] la [illegible] del mio infortunio, et in certo [illegible] le persecuzioni de' miei nimici, senza le quali mi sarebbe sempre restata occulta la [illegible] da stimarsi dell'humanità e benigna [illegible] di molti miei Sig.ri e [illegible] e sopra tutti l'Amore di V. E.: il quale [illegible] d'esser promosso da talento alcuno di virtù che la Natura habbia riposto in me, ha in vece di lei supplito la forza [illegible] nelle lor menti il fuoco della carità, [illegible] le quali vanno compatendo [illegible] lo stato mio; nel quale oltre alla ragion detta mi è di non piccolo sollevam.to il trovar che non un'animo che sempre più si vada inasprendo sia quello che continui di tenermi oppresso, ma più tosto una grave [illegible] ragion di stato di quelli che vogliono ricoprir il primo errore d'havere altre volte offeso un'[illegible], col continuar l'offesa, e [illegible] universale si forma [illegible] non fatti valevoli aggravar la colpa del [illegible]. Però sia qualche [illegible] a chi è concessa la potestà di fare il suo arbitrio che in tutti gl'eventi resterò [illegible] obbligato alla somma bontà di V. E. la quale con tanta premura si [illegible] nel mio interesse, e con tanta industria, e vigilanza indefessa va [illegible] i mezzi che possono essermi di sollevam.to

L'Horologio hydraulico sarà veram.te cosa d'estrema maraviglia, quando sia vero che il Globo [illegible] nel mezzo dell'Acqua [illegible] naturalm.te [illegible] per occulta virtù magnetica. Io feci già molti anni sono una simile invenzione; ma co'l aiuto d'un [illegible] artifizio e la Machina era tale. Il [illegible] d'un [illegible] di rame voto dentro, e con un pezzetto di calamita postogli nel fondo, equilibrato quasi alla gravità dell'Acqua, si che [illegible]

che io, od altri volesse stampare, sì che la proibizione è de omnib.s editis, et edendis, et è preso [illegible] di fare stampare il resto delle mie fatiche non publicate ancora, e forse vi è morto per curiosità di veder l'esito di questa impresa, e che fortuna correranno tali materie [illegible] alteranno à religioni più che non è il Globo della Terra. Io contro à mia voglia sono stato forzato à [illegible] cederne copia à S. A. sicuro che à me non se possa succedere se non qualche travaglio; se bene non mi è stata fatta, ne accennata proibizione alcuna; perchè non deve ne anco haver notizia del decreto [illegible] à gl'Inquisitori; per lo che questo che [illegible] à V. E. sia detto in confidenza. Da [illegible] in Firenze, et in Roma tutte l'opere mie [illegible] più non se [illegible] trovare per le librerie, [illegible] si scorge che si fa ogni sforzo per [illegible] del mondo la mia memoria; nella qual [illegible] avversarij quanto poco io premo, forse non si mostrerebbero tanto ansiosi d'opprimermi.

Io non finirò di parlar con lei senza di nuovo ringraziarla della sua infinita benignità, e del fervore col quale [illegible] ne miei interessi, e se il [illegible] che fuor di tutte sue [illegible] travagliato [illegible], viva V. E. sicura che ne riceverà guiderdone dalla Divina bontà. E qui [illegible] affetto gli bacio le mani, e nella sua grazia [illegible]

Dalla Villa d'Arcetri li 12 di Maggio 1635

Di V. S. Ill.ma et Ecc.ma

Dev.mo Et Obblig.mo Ser.re

Galileo Galilei

FIGURE 9 (left). Galileo's letter to Peiresc. Written in Italian, the letter begins: "Most distinguished and excellent Sir, my esteemed patron—Your Excellency's letter, filled throughout with feelings of courtesy and goodwill, continues to make the fortune of my misfortune appear sweeter to me, and in a certain way to bless the persecution of mine enemies. . ." Excerpts from the second paragraph (sixteenth line of the text) begin: "The water-clock will truly be a thing of great marvel if it is true that the globe suspended in the middle of the water goes naturally turning by an occult magnetic force. Many years ago I made a similar invention. . .The little globe with 12 meridians for the 24 hours was of copper, hollow within, with a little piece of magnet placed at the bottom, and almost in balance with the density of water. . ." The letter ends: "From the Villa of Arcetri, the 12th of May, 1635—from your Illustrious and Excellent Lordship's Most devoted and obligated servant, Galileo Galilei." (Note: Because of space limitations, the reproduction of this letter has been considerably condensed).

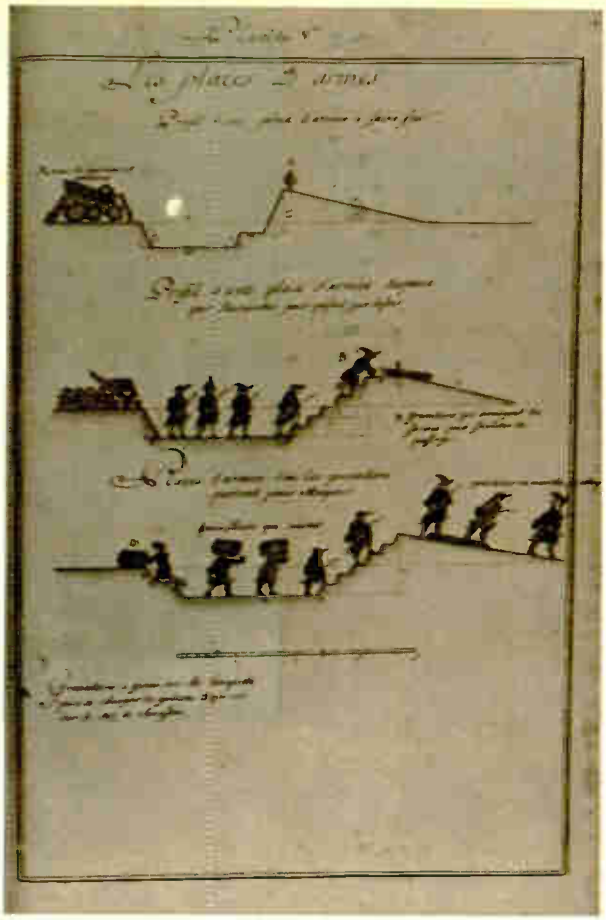


FIGURE 10 (right). Vauban's pictorial plate, from top to center: (A) shows the profile of a firing trench, overtopped with a wooden parapet, as used by besieging infantry; (B) trench with built-in firing steps and fascine parapet employed by attacking forces. Note in top and center profiles the "gabions" and "fascines"—portable wicker baskets (filled with earth when in position) and tied bundles of sticks, used as shields against defensive artillery and small-arms fire while attackers were "digging in" after consolidating an advance. Finally, the bottom profile shows grenadiers going "over the top" in a general advance, followed by engineer support troops carrying the empty gabions prior to setting up a gabionade in a more forward siege position.

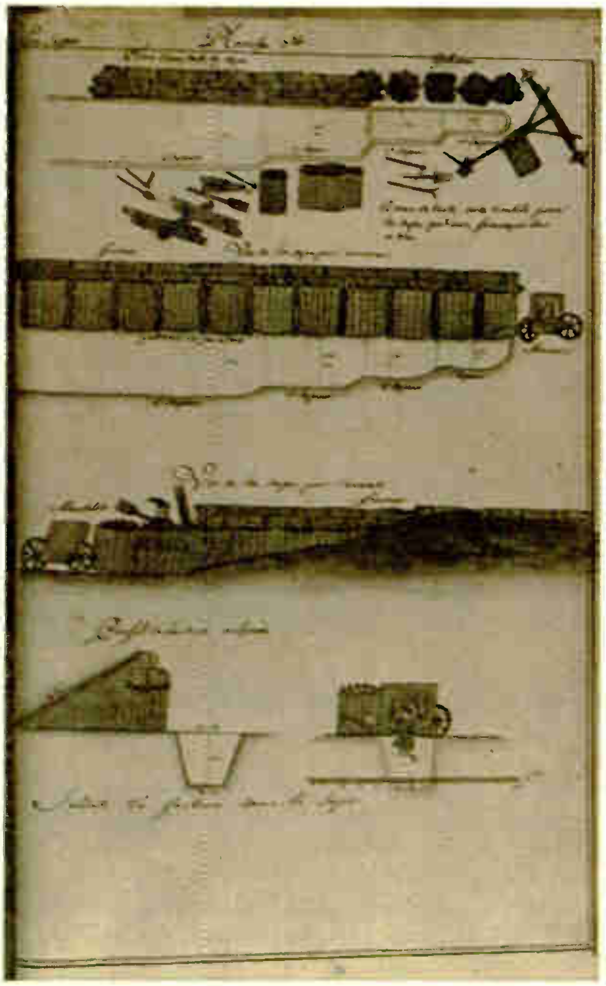


FIGURE 11 (bottom, right). Plan and elevation views of a gabionade, overtopped by fascines in a front-line attack position. Here, sappers are further protecting the attackers' exposed front by entrenching behind the barrier and using the excavated earth to form a glacis slope in front of the gabionade.

than 400 such bulbs, dating from about 1890 to 1915, the collection displays a broad range of unique electric lamp shapes, filament configurations, and socket bases.

The special collections

Manuscripts. The Burndy Library preserves an excellent collection of scientific manuscripts, drawn from sources throughout the world, which range from calendar notes by the astronomer Regiomontanus in 1475 to Einstein's 4½-page summary of "The General Theory of Relativity." Within this general time frame, the Library is the repository for

1. More than a dozen manuscripts by Sir Isaac Newton, written in several languages and including a 66-page autograph manuscript of his notes on alchemy and chemistry.
2. A remarkable letter written by Galileo to Nicolas Claude de Peiresc, the French astronomer, on May 12, 1635. In a strong and legible hand (Fig. 9), Galileo tells of his mental suffering under the persecution of the Inquisition, but also gives his scientific views on the proposed design of a magnetic clock and describes his own previous efforts in designing a similar device.
3. Correspondence (99 letters), written from 1742 to 1770, between Abbé Jean Nollet, the noted French

FIGURE 12. The Burndy Library publications include science monographs, biographical sketches of history's foremost scientists and technologists, and narratives of epochal events in the history of science and technology. At the right is the circular plaque of Michael Faraday cast in Hadfield manganese steel, the alloy developed by the versatile 19th century scientist.

physicist and electrical experimenter, and Étienne Dutour, theologian and natural philosopher, pertaining to the electrical and scientific discoveries of this period.

4. The Weiner Collection—gathered prior to World War II by Dr. Arnim Weiner of Prague; its more than 1000 items are devoted to the manuscript correspondence of the world's foremost scientists over a time span of 4½ centuries. The collection includes important manuscripts written by Born, Boyle, Euler, Fraunhofer, Kepler, Mach, Pasteur, Planck, and Priestley. It also contains abstracts of each letter or manuscript, and portraits of many famous scientists.

5. Rare notes, letters, and a manuscript by Wilhelm Röntgen, discoverer of X rays. (Just prior to his death, Röntgen had destroyed the records of his scientific studies and requested his friends who possessed any of his works to do the same.)

6. Forty of Michael Faraday's letters, including the one in which he outlines the formal announcement of his momentous discovery of the principle of electromagnetic induction (later expanded into the famous papers that were published in the *Transactions of the Royal Society*).

7. The *Manuscript on Fortifications* by the Comte de Vauban (Louis XIV's great general and engineer, the father of modern military engineering). The manuscript, signed by Vauban, is written on 228 pages, and is supplemented by 43 drawing plates (Figs. 10 and 11), in which the author delineates his methods of building various fortifications, the tools and engines required, the cannons and mortars essential for siege and defense, the plans for defense of and attack against bastions (including fields of fire), the effects of gunpowder and mines on defensive positions, etc. Vauban fortified many of the cities of France, including Verdun—which the Germans failed to capture during a 10-month siege in 1916.

8. A very beautiful and well-preserved manuscript, prepared in 1461, of the famous poem in "sesta rima" (rhymed sextets) by Cecco d'Ascoli (1257–1327), which contains the noted astronomer's views on the natural history of the world, including many of his very progressive ideas. The manuscript consists of four books: the first, on astronomy and meteorology; the second, on stellar influences (astrology); the third, on minerals and animals (here one of the earliest references is made to the properties of the magnet); the fourth, on a number of moral and physical problems.

Private libraries. During World War II, most of the library of John Tyndall, the noted 19th century English physicist, was sent to the United States. The Burndy Library was fortunate in obtaining about 250 volumes and a number of pamphlets (most of which are inscribed to Tyndall).

Toward the end of the Second World War, when Dr. Dibner was serving with the U.S. Army Air Force in the European Theater of Operations, he acquired the largest single collection of the library of Alessandro Volta, consisting of more than 300 items: books, manuscripts, original documents, correspondence, laboratory notes on the experiments that led to the development of the Voltaic pile (the first battery), essays on machines related to the operation of oil- and gas-fired lighthouse beacons, and many other priceless memorabilia.

The Burndy Library publications

Since 1942, the Library has been engaged in its own active publications (Fig. 12) program. The continuing policy is to publish at least one work per year, in which the themes alternate annually between subjects in general science and electricity or magnetism. The publications range from monographs and biographical treatises to annotated scientific bibliographies and narratives of epochal events in the history of science.

The first opus in the series of 24 works published to date was Prof. I. Bernard Cohen's *Roemer and the First Determination of the Velocity of Light* (first edition, 1942; second edition, 1944). Among the works authored by Dr. Dibner are *Leonardo da Vinci, Military Engineer* (1946); *Doctor William Gilbert* (1947); *Moving the Obelisks* (1950); *Ten Founding Fathers of the Electrical Science* (1954); *Heralds of Science* (1955); *Early Electrical Machines* (1957); *Agricola on Metals* (1958); *The Atlantic Cable* (1959); *Darwin of the Beagle* (1960); *The Victoria and the Triton* (1962); *Alessandro Volta and the Electric Battery* (1964); and *A Letter from Galileo* (coauthored with Stillman Drake, 1957).

The latest publication (1969) is *Leonardo da Vinci—Technologist*, by Ladislao Reti and Bern Dibner.

Special exhibitions

In the 35-year existence of the Library, there have been many special exhibitions of particular historical significance. The first exhibition off-premises was on the subject of medical history, and was held in 1948 at the Field Library, Peekskill, N.Y. Among the Burndy Library displays were the first printed book on medicine by Celsus, 1478; the illustrated incunable on surgery by Ketham, 1495; and Harvey's *De Motu Cordis* of 1628. Also shown was the account of Dr. Morton's first use of chloroform as an anesthetic in surgery, in 1846.

Another important exhibit, in 1952, celebrated the 500th anniversary of the birth of Leonardo da Vinci with the showing of Vincian printed and graphic material at the Silvermine (Conn.) Guild of Artists. Many of the earliest books describing his writings and drawings were displayed together with prints of his sketches and paintings. Models of Leonardo's mechanical devices were lent by Thomas J. Watson, Sr., of IBM.

In 1957, an exhibition of 20 paintings and prints under the title of "Art in Science" was arranged for the Phillips Academy at Andover, Mass. And in the same year, an exhibition of early electrical machines in book and print was held in New York in connection with a meeting of the AIEE.

Exhibits at the Burndy Library began in 1954, with a collection of books and prints relating to the "Martyrs of Science"—from Servetus, who was burned at the stake in Calvinist Geneva for heresy in 1553, to Lavoisier and Buffon (executed during the French Revolution) and the great explorers (Franklin, Cook, Ross, Scott, Amundsen, etc.), who lost their lives in the early polar expeditions or during the epochal voyages of discovery.

The passing of Albert Einstein in 1955 was observed by a memorial exhibition that included the great physicist's first publication in 1901, plus books, letters, manuscripts, and other memorabilia.

In 1956, the 250th anniversary of Benjamin Franklin's birth was appropriately noted by an exhibition of the original editions of his books on electricity together with some Franklin manuscripts. Also on view were the unique bronze bust of Franklin by Houdon and the famous engraving, *Apotheosis of Franklin*.

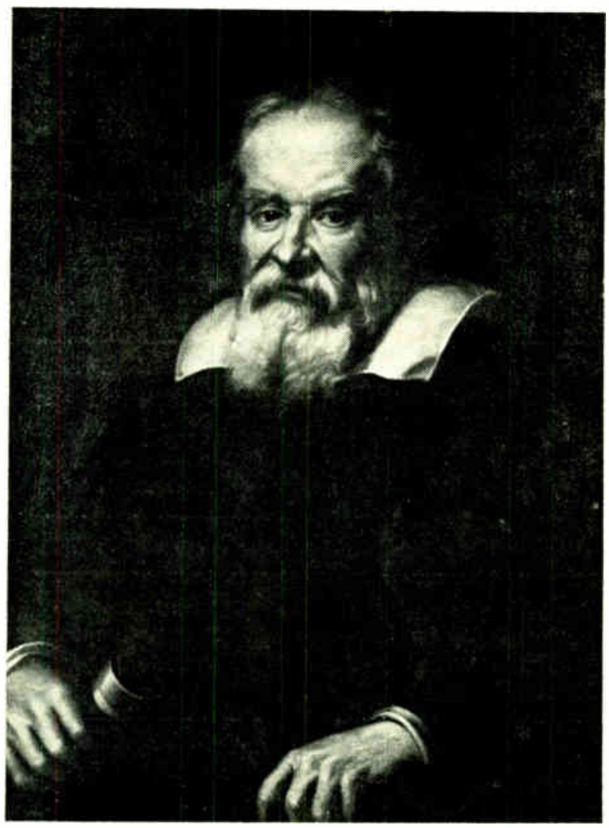
Late in 1968, the Library mounted a special showing, cosponsored by the American Institute of Physics and IEEE, of the selected papers, correspondence, laboratory notes, books, pictures, and memorabilia of Michael Faraday. Through the courtesy of the AIP, several unusual display items, shown at the 1967 exhibit at the Rockefeller University in commemoration of the centenary of Faraday's death, were made available on loan.

Last June, the Library's staff began to assemble one of its largest and most timely shows (Fig. 3). Appropriately titled "Man's Conquest of the Moon," the exhibit placed the three-century time span—from Galileo and Hevelius to Apollo 11—in a historical perspective through displays of the earliest selenographic charts (Galileo, 1610; Hevelius, 1647), made with the aid of primitive telescopes, to the fantastically beautiful color photos (courtesy of NASA) taken during the Apollo 8, 9, 10, and 11 missions. To enhance the pictorial exhibits, a scale model of the Lunar Module (LM) was donated to the Library by the Grumman Aircraft Engineering Corporation, designers and builders of the Apollo moon-landing vehicles.

Art and iconography

To demonstrate the pleasing interactive effect of art and technology through a merger of these disciplines in an esthetic setting, one has only to visit the Burndy Library where the drama of science and engineering is illustrated by appropriate art and iconography. On the north and south walls of the main reading room are hung portraits in oil of Charles Goodyear; the 19th century U.S. inventor, Sir Goldsworthy Gurney; Samuel F. B. Morse; Robert Stephenson, the civil engineer who built the first railway in Great Britain; Michael Faraday; John Ericsson, the Swedish-American marine engineer who designed the U.S.S. *Monitor*; Josiah Willard Gibbs;

FIGURE 13. The Burndy Library's portrait of Galileo. Evidence indicates this portrait, in the style of Sustermans, was probably painted in 1641, a year before his death, when the aged astronomer was already blind.



and Galileo Galilei (probably painted from life in 1641). The Galileo portrait is shown in Fig. 13.

In 1966 and 1967, the Library was pleased to receive from IBM a gift of 21 oil paintings of great scientists, medical researchers, engineers, and inventors (including six of those just mentioned). Portraits of Isaac Newton, Benjamin Franklin, William C. Gorgas, and Lee de Forest are contained in the collection.

Among the other works of art gracing the walls are a portrait of Einstein, painted by Josef Oppenheimer in 1924; a crayon sketch of Max Planck, made from life in 1932; a fine likeness in oils of Oersted; and a miniature of Spinoza by Arthur Szyk.

In the sculpture category, the Library has a large marble bust of Newton that was carved in about 1800, and was owned at one time by Sir Robert Peel, a 19th century English prime minister. Also prominently displayed in the main reading room is a terra cotta study of Buffon made by Pajou in 1773 (the marble bust carved from this study is now in the Louvre).

On the mezzanine, there are a marble bust of Volta and a bronze portrait of Edison as a young man. Here, too, one may see a large bronze bust of Heinrich Hertz, the famed German physicist who discovered electromagnetic radiation in 1886. And on the south wall of the main floor, there is a circular plaque (Fig. 12) bearing Faraday's likeness in bas relief. The piece was cast in England about 15 years ago in Hadfield manganese steel, the alloy developed by the versatile Faraday more than a century ago. Nearby is a terra cotta half-figure of the French microbiologist Louis Pasteur in a pensive mood.

Flanking the main entrance foyer are two exquisitely made, full-length marble statuettes of Robert Stephenson and Isambard Kingdom Brunel (the engineer who designed the famous cable-laying steamship *Great Eastern*), and on the far wall of the foyer is a large mural that shows how Domenico Fontana relocated the Egyptian obelisk to its present site before St. Peter's Cathedral in Rome in 1586. But the most highly prized work of art in the Library is the unique bronze casting of the bust of Benjamin Franklin, sculpted by Houdon in 1778, when Franklin was the U.S. ambassador to France.

Gifts and acquisitions

Visitors often ask Dibner: "Where did you get all of these treasures?" The answer is too complex to be given in one sentence. First of all, the Burndy Library collection of rare books, manuscripts, correspondence, and museum exhibits was painstakingly acquired over a period of more than 35 years—and is still in process—from many sources and by many avenues. A portion of the valuable works was donated by bibliophiles and Friends of the Library; other rare items were either purchased from antiquarian bookdealers throughout the world, or by successful auction bids by the director.

There also have been interchanges of gifts between the Library and professional societies and universities. For example, the Illuminating Engineering Society donated 32 volumes on electrical illumination; 38 volumes in electrical history, and other personal memorabilia were received from the widow of the late Laird Goldsborough, one-time Fellow of the AIEE. And, although the Library was initiated by building up a fabulous collection of Vinciana, more than 1000 of these books, pamphlets, drawings, and commentaries were donated several years ago to the then new Brandeis University Library. Still, about 200 books, relating primarily to Leonardo's scientific and technical interests, remain available to reward the researcher and the scholar.

Roster of visitors

In 1954, the use of a formal guest book for recording the names and addresses of visitors was initiated. Dr. Ladislao Reti of Buenos Aires, industrialist and noted Leonardo scholar, was the first entry. Subsequent early registrants were Dr. Iago Galdston, secretary of the New York Academy of Medicine; Prof. I. Bernard Cohen, historian at Harvard University; Dr. Herbert M. Evans, the eminent biologist who discovered vitamin E; Elgin B. Robertson and L. F. Hickernell, former presidents of the AIEE; and Ernst Weber, president of the Polytechnic Institute of Brooklyn and a former president of the IEEE. In 1957, the growing list included Dr. Gaichi Kamo, president of Otaru University · Hokkaido; Vladimir Zworykin, noted inventor and vice president of RCA Laboratories; Profs. Stig Ekeloef of the Chalmers Institute of Technology in Gothenburg, Sweden, and Derek Price of Cambridge and Yale Universities. Nobel Laureates Isidor Rabi and Charles Townes are included.

More recent visitors of especial note were Dr. John F. Fulton, the eminent physiologist and bibliophile of Yale University; Herman W. Liebert, director of Yale's Beinecke Library; and Dr. William C. Gibson, investigator of neuroelectricity and the history of medicine at the University of British Columbia. Early in 1969, the Library was honored by a visit from Dr. Ladislaus Marton, the Hungarian-American physicist who was pioneer in the development of the electron microscope.

In perusing the guest books, one sees the names of scientists, engineers, researchers, and scholars from every European and South American country, plus Israel, India, Japan, Malaysia, Hong Kong, Thailand, Australia, Mexico, and Canada.

Epilogue

Because of spatial constraints, the writer has presented only a broad-brush overview of the highlights of the contents, activities, uses, and purposes of the Burndy Library. Much that would be of especial interest to individual historians, writers, and scholars has been omitted. But to make amends for such limitations, a cordial invitation is always extended by the director to the reader to visit the Library for any and all detailed informati on a complete interdisciplinary range of scientific subjects and on the scientists who laid the foundations of our present great technological competence.

This article opened with the famous quotation attributed to Sir Isaac Newton. Perhaps it is appropriate to close the piece by noting the coincidence by whi Nature apparently provided for a continuity of ra genius: Newton was born in 1642—the year that witnessed the passing of Galileo, who was born in the year in which Michelangelo died. They were among the foremost innovators who heralded the future of art, science, and technology.

Gordon D. Friedlander's biography appeared on page 43 of the February 1969 issue.

Infrared and microwave communication by moths

The possibility that insects communicate by means of infrared and microwave radiation has fascinated entomologists for many years. Recent experiments, motivated partly by the desire to invent a better moth trap, show that the possibility cannot be altogether excluded

Henry S. Hsiao, Charles Susskind University of California, Berkeley

Analytical and experimental studies seeking to confirm or refute the hypothesis that moths communicate with each other by coherent electromagnetic waves have been under way at Berkeley for several years. This article is the first report from the bioengineering group conducting these studies.

Speculations that insects may use some form of electromagnetic communication (in addition to vision) go back to before the beginnings of radio. As early as 1894, the great English-born entomologist Charles Valentine Riley, first chief of the U.S. Entomological Commission and author of the monumental *Insect Life* (1889–1894), suggested that insects might sense subtle vibrations to which we are blind[1]—a hypothesis that was brilliantly vindicated when it was shown that bees and other insects see ultraviolet light. Among those who followed him, the famed Jean Henri Fabre, *doyen* of French entomologists and author of the ten-volume *Souvenirs Entomologiques* (1897–1907), found it hard to believe that scent alone could be responsible for long-distance sexual attraction among moths.[2] Still other entomologists suspected that insects possessed an undiscovered sense that alerted mates from afar, and suggested that insects might transmit as well as receive electromagnetic waves.

However, all these speculations were experimentally unsubstantiated and remained a collection of intuitions and impressions rather than of fact at least until 1948, when Grant pointed out that certain structures on insect antennae are of a size comparable to infrared wavelengths and might be electromagnetic receiving elements.[3] Two years later Duane and Tyler hypothesized that the infrared radiation arising from the body heat of the female moth could be the signal that attracts the male.[4] They actually measured a definite radiation pattern resulting from a temperature rise in active females.

Current interest in the subject is largely the result of the efforts of a former U.S. Air Force communications officer, P. S. Callahan, now an entomologist in the research service of the U.S. Department of Agriculture. He is currently working at USDA's Insect Attractants, Behavior, and Basic Biology Research Laboratory at Gainesville, Fla., and is also professor at the University of Florida. In publication after detailed publication, Dr. Callahan has steadily championed the idea that insects—in particular the corn earworm moth, on which he is an acknowledged authority—very likely respond to electromagnetic radiation below the visible spectrum.[5–14]

Experimental and analytical work to test this assertion has been under way at the authors' laboratory since 1967. The present article is the first report on this work to the engineering public—in fact, the first comprehensive report anywhere, since only partial reports of our work have been published previously even in the entomological literature[15] and as laboratory reports.[16] Although the results obtained to date cannot be considered as final, they allow certain definite conclusions to be drawn.

The Callahan hypothesis

Callahan postulated that night-flying moths locate mates and host plants by using infrared or microwave radiation. For example, a temperature difference between a moth and its surroundings would make the moth stand out as an infrared radiation source to other moths some distance away. Also, even over short distances (of the order of a few meters), insects might detect chemical sex attractants by a mechanism that involves electromagnetic radiation,[17] which passes through a "window" in the atmospheric absorption spectrum (a portion of the spectrum in which water vapor, CO_2, and other constituents do not render the atmosphere opaque) and resembles a "homing" signal for flying moths.

Employing a bolometer, Callahan measured body radiation from several species of moths and concluded that the power output was sufficient for the moth to behave as a living transmitter. Moreover, the normally observed wing-beating behavior in "transmitting" moths might be nature's way of "chopping" the "carrier" radiation, thereby producing a modulated signal that is distinct from background radiation. Last year William Husbeth, a graduate entomology student at the University of Georgia working with Callahan, used a fast-response, helium-cooled bolometer to detect radiation from two moth species, the corn earworm and fall armyworm. The results suggest that the AM infrared signals are quite different for each species.[18]

Light-microscope examination of the many sensilla on the moth antennae confirmed that each of the sensilla

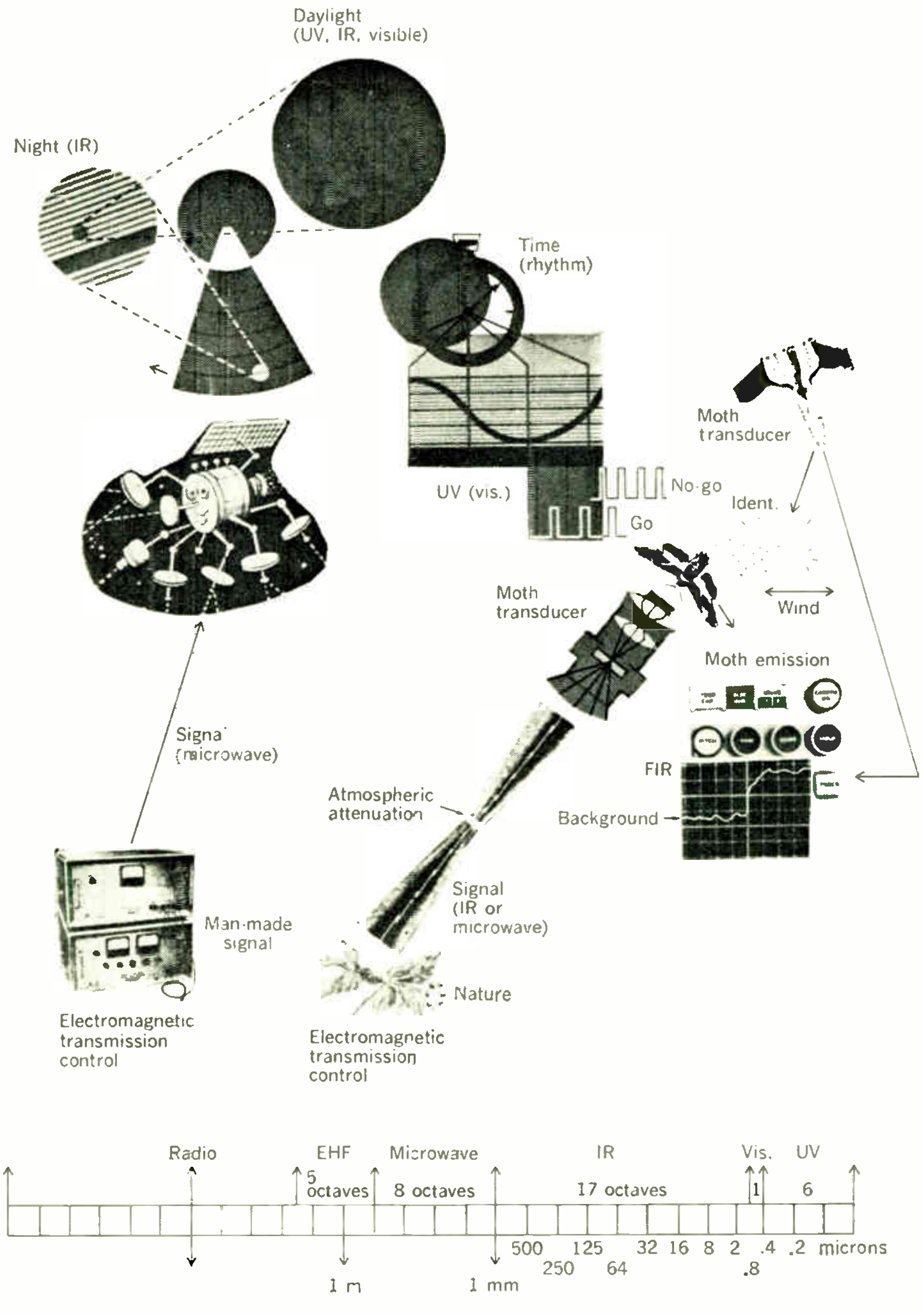


FIGURE 1. According to Callahan's original theory, daylight and night (top) program night-flying moths for "go" or "no-go" signal, depending on radiation present; "go" for IR (night), "no-go" for visible and UV (day); shorter UV wavelengths program the insect with regard to time, longer IR wavelengths direct and control his behavior. As "go" signal is received during night IR environment, "camera hold" is overridden and IR eye "camera on" takes over. Pitch, yaw, and gyro controls come into operation, temperature readout of background takes over, and airborne insect wanders on random and erratic course until IR or microwave emission from host plant or emitting mate locks in on moth transducer. Insect mate may be further tracked or identified by IR or microwave emission of scent molecules. Such emission would not be as strong as total plant or moth emission, but would be enhanced by movement of molecules toward insect.[10]

trichodea (hairlike sensors, also called "hair spines") is similar in shape to a dielectric-waveguide antenna and is of a size that would suggest optimum operation in the infrared range. The optically measured dielectric constant of the cuticle and the possibility that the layer of wax that surrounds the hair spine may act as a thermoelectret are both consistent with the hypothesis that the spine serves as a waveguide antenna. (In a thermoelectret, heat and electric field combine to produce a residual polarization; in the present case, that mechanism may serve as the IR transducer.) The moth's compound eyes might also receive IR radiation once they are dark-adapted.

Finally, in one of his recent experiments, Callahan reports that certain spines at the base of saturniid moth antennae yield neuroelectric responses when irradiated by a laser.[12] From this evidence he speculates that such an elegant communication scheme not only might account for long-distance assembly for mating and for foraging, but also might explain insect phototaxis—the Lorelei-like attraction artificial lights have for night-flying moths. The theory is schematically shown in Fig. 1.

Implications: pest control

The implications of electromagnetic communication by insects are awesome. Not only might it open a new chapter in entomological studies but it would directly influence the fields of neurophysiology, ecology, and electronics. But the greatest immediate market for this theory is in the economic control of agricultural pests. If insect populations could be effectively regulated through the design of competitively efficient traps that leave no contaminating residues, then the use of pesticides with harmful side effects could be curtailed. In addition, it is becoming evident that certain harmful insect species are becoming resistant or immune to many insecticides. The benefits of electromagnetic control are manifest and would provide sufficient motivation for further investigation even if the theory were not of inherent scientific interest.

The investigation being carried on at the University of California's Electronics Research Laboratory is proceeding along two separate but complementary approaches. One is a systems analysis to determine whether the precepts of such a theory are compatible with known laws of nature, by survey of the boundaries and limitations of its assertions. The other is a predominantly experimental attempt to observe such a method of communication in a behavioral response to infrared signals.

As in Callahan's work, the insect under study is the moth *Heliothis zea* (Boddie), Family Noctuidae, more commonly called the corn earworm or cotton bollworm. Callahan has made extensive measurements of body radiation and attraction to blackbody radiation sources for this moth. In addition, the morphology and behavior of this insect are comparatively well defined. The sensory structures, the dielectric properties of its chitin, and the life cycle are documented.[14] Furthermore, moths of this group, which are indigenous to North America, include the most notorious pests in the world. The larva of *H. zea* feeds on a large variety of fruits and vegetables, causing hundreds of millions of dollars worth of damage yearly in the United States alone. The clear and present danger of this pest provides all the motivation one could ask for probing all possible methods of control.

Systems analysis

The first analytical task is to formulate a scheme for electromagnetic communication from the available evidence. One might begin, for instance, by assuming that the signal transmissions originate from particular chemicals secreted by female moths. These chemicals or *pheromones* may radiate (or absorb) infrared frequencies. Entomological literature contains abundant evidence that chemical sex attractants are indispensable elements in the mate-finding behavior.[19] As long as ten years ago, Laithewaite showed that an empty box that had once contained virgin corn earworm female moths would continue to attract males, though not as strongly as would the female herself.[20] The use of scent is particularly well documented in Lepidoptera (butterflies and moths) and some Coleoptera (beetles). For some insects, for example the bark beetle *Ips confusus*, the pheromone has been not only identified but also synthesized and even shown to consist of three compounds, which act synergistically to induce attraction in nature. None of the three compounds is active by itself.[21] The pheromone for *Heliothis zea* has been bioassayed by Shorey and Gaston,[22] but no one has succeeded in synthesizing it yet.

To be effective over long distances, any electromagnetic communication scheme would just about have to be a high-directivity, modulated, narrow-band process. Moreover, it would have to be confined largely to one of the bands in which the atmosphere does not strongly absorb infrared radiation—in other words, one of the "windows" in the infrared spectrum of the atmosphere.

One may postulate that the insect antennae are the sites of the receptors of electromagnetic radiation. Experiments show that attraction of male moths to females is severely reduced when the antennae are impaired.[19, 23, 24] Callahan has shown that corn earworm moths have difficulty locating food when portions of the antennae are excised.[25] Our own tests have also shown that removing large portions of the antennae has a detrimental effect on the mating behavior of the same insect.[16] Clearly the antennae are the key structures for perceiving "calling" females.

Molecular transmitter. Research to date shows that many moth sex attractants resemble long-chain unsaturated alcohols or esters.[19] It is at least conceivable that such a molecule (or set of molecules) might be detectable electromagnetically. The stretching vibration of the pheromone could be the radiation source. Various pheromones would have different transmission (or absorption) spectra, which might explain why insect species rarely crossbreed.

The primary determinant of the effectiveness of a molecular or any other transmitter is whether sufficient energy is available. If we assume, quite conservatively, that communication can extend to a distance of 50 meters, approximately 2×10^{11} quanta per second must be transmitted. A quantum can be emitted after the molecule has been elevated in energy from the ground state. Such elevation may occur in three ways: (1) Raman scattering, where impinging light energy is reemitted; (2) chemical reaction, where the product molecule is in an elevated state; and (3) thermal activity of colliding molecules.

Calculations show that there is not enough energy from the first two sources, but that the third (which yields about 2×10^{14} quanta per second if we make the reasonable assumption of an emission rate of 10^7 molecules of pheromone emitted per second) is possible. But then we must also allow for interference from "transmitting" molecules such as water vapor, which would yield radiation that is 10^{16} times stronger than the radiation from pheromone molecules. To detect a signal in so much noise would require an unimaginably sophisticated receiving apparatus.

Long-distance olfactory communication. If we assume, following Griffith,[15] that a male moth requires a minimum of 50 stimulations per second from the calling female, the maximum radius of attraction due to diffusion in still air alone is a mere 1.8 meters. (To be sure, the results will be drastically affected by air movement.) At 50 meters the molecules are so thinly spread that they lead on the average to only one collision with the male moth antennae every 10 seconds. A purely olfactory scheme for *long*-range communication without wind is thus improbable on the basis of present evidence.

Receiving-antenna structures. The best reason for supposing in the first place that insects communicated electromagnetically was the startling similarity of the hair sensilla on the antennae to a radio antenna array made up of dielectric rods. But qualitative similarity in shape was in itself not enough. If this were indeed a form of infrared radio, the radio engineer would expect the antenna to be of a size comparable to infrared wavelengths. In a dielectric-rod antenna the upper cutoff wavelength is determined by internal high-frequency losses; the lower cutoff wavelength is fixed by decreasing efficiency as the wavelength approaches the length of the rod. Moreover, one would expect to find some sort of "feed" structure at the base of the dielectric rod to couple the received signal to the detector. Such a struc-

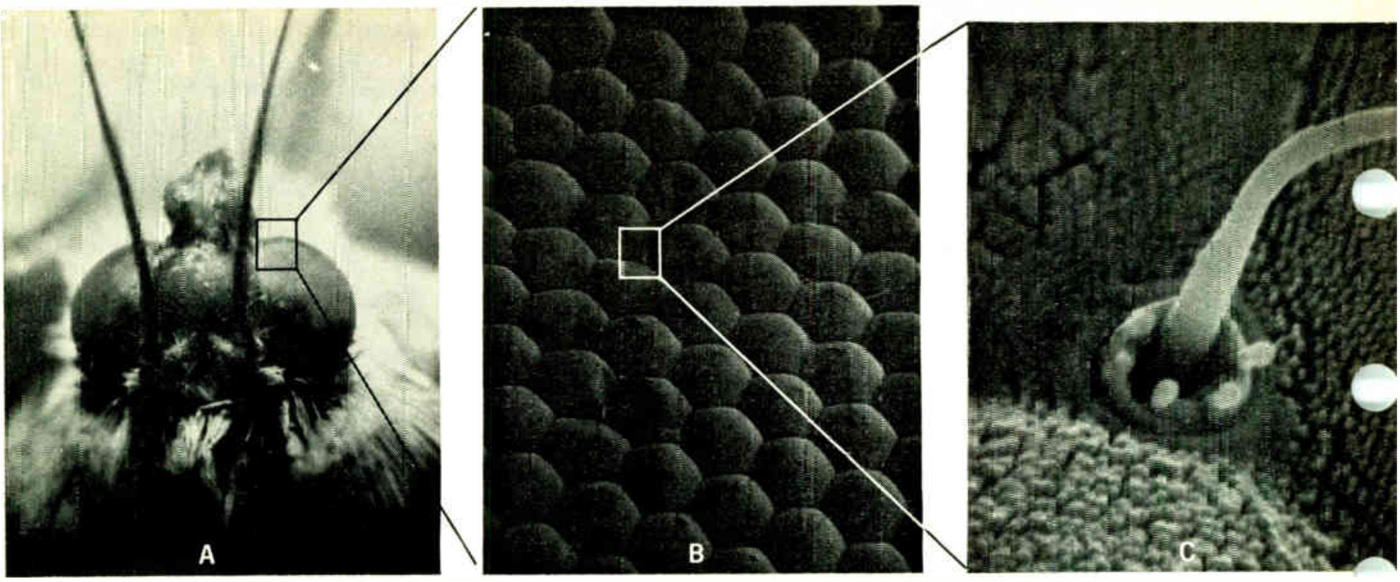


FIGURE 2. Successive enlargements of corn earworm antennae and eye. A—12× magnification obtained by optical microscope. B—Scanning electron microscope magnification (560×) of facets of compound eye; note structures occurring irregularly in interstices, but only in frontal portion of eye. C—Further magnification (11 300×) of one of the structures.

ture might resemble a grounded ring surrounding the base of the rod, with a detector, also at the base, transducing the radiation into a neuroelectric signal.

In determining the morphology of the antenna sensilla, the bioengineer has a powerful new tool at his disposal: the scanning electron microscope (SEM). This new photographic technique yields excellent detail with unusually great depth of field and, one may say, adds a new dimension to such studies. (That is not simply a metaphor; if stereoscopic-pair techniques are used, the results may be literally perceived as three-dimensional, with spatial relationships fully preserved.)

The SEM is so new that almost any investigation yields new results. For instance, in taking the series of photographs shown here as Fig. 2, we noticed some new detail in the crevices of the compound eye, occurring at irregular intervals. Further magnification yielded the astonishing structure shown as Fig. 2(C). The function of this structure is unknown at present and must be investigated by entomologists. (The incident serves to illustrate bioengineering at its best: the use of engineering tools and methodology in the discovery and elucidation of results of basic significance for the life sciences.) Another set of successive enlargements is shown in Fig. 3.

In an earlier example of this procedure, it was possible to correlate the structure of the observed sensilla with known sensory functions of the insect antennae. The results of this comparison were remarkable. The morphology of the sensilla trichodea revealed at magnifications up to 12 000× that each indeed resembled a dielectric-rod antenna element. Surrounding the base of the hairs was the very ringlike socket that had been postulated as the necessary "feed" structure. And supporting optical microscopy revealed perhaps the most suggestive evidence: a structure at the base of the spine that might serve as a detector of broadband IR energy.

The graduate student who conducted this part of the investigation, P. H. Griffith, remembered at this point that he was, after all, an electrical engineer, and proceeded to construct scaled models of several spines and to measure their operating bandwidths at the considerably longer (and thus more convenient) microwave wavelengths. The results of this study also supported the electromagnetic theory. Extrapolation back to the actual size of the tiny hairs showed that if each did indeed act as a transmitter or receiver element, the corresponding optimum wavelength would be 4.5 μm. And one of the atmospheric "windows" (Fig. 4) does fail at that wavelength! That is also the very wavelength to which certain mites appear to respond, as established in some hitherto unreported experiments.[17] However, some questions remain. First, no evidence was found for the existence of any structure that might perform the highly specific task of narrow-band detection that, under the assumptions of the present study, one might take to be essential to maintain the aforementioned lack of crossbreeding among different species. (However, a broadband detector would be just the thing for perceiving *temperature* changes.) Second, it is not altogether clear whether one may extrapolate dielectric-rod antenna design with impunity to frequencies corresponding to micrometer wavelengths. Most physicists would doubtless prefer to think of such structures in terms of quantum physics and not geometrical optics.

"Conventional" (i.e., transmission) electron microscopy (TEM) was next pressed into service and TEM micrographs were taken of cross sections of the antennae. Assistant Rita Aronow developed an original technique for staining, fixing, and sectioning these samples, which are extremely brittle and difficult to cut into the extrathin sections—of the order of 100 Å (0.01 μm) thick—needed in TEM investigations.[16] This set of micrographs held another surprise in store: the existence of pores approximately 500 Å in diameter on certain sensilla, with neurofibrils running right up to the pores and terminating near them, a configuration that would suggest that olfaction may after all be the primary function of these

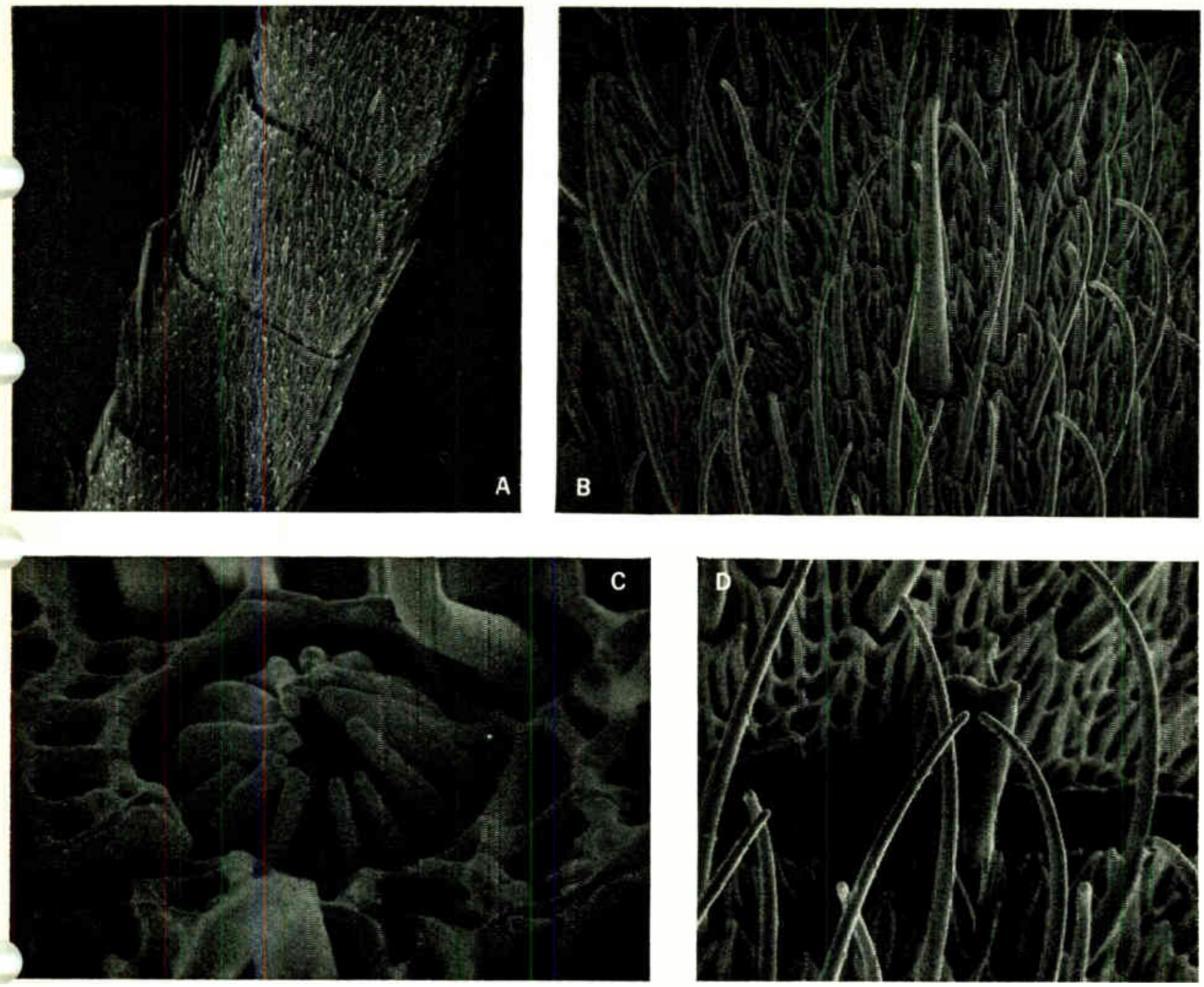


FIGURE 3. Scanning electron micrographs of antenna sensilla from corn earworm moth. A—Antenna segments (200×). B—Enlargement (1060×) of portion of one segment showing one of the thicker sensilla chaetica protruding from socket and surrounded by sensilla trichodea; flowerlike structures nearer surface are sensilla coeloconica. (Also seen, near bottom left corner: one of the stubby structures known as "shoehorn" sensilla.) C—Detail showing one of the sensilla coeloconica with peg in center (5400×). D—Detail showing still another antenna sensor, also seen on proboscis, called "taste rod" (2170×).

sensilla—even though the process of olfaction may in turn involve electric phenomenons.

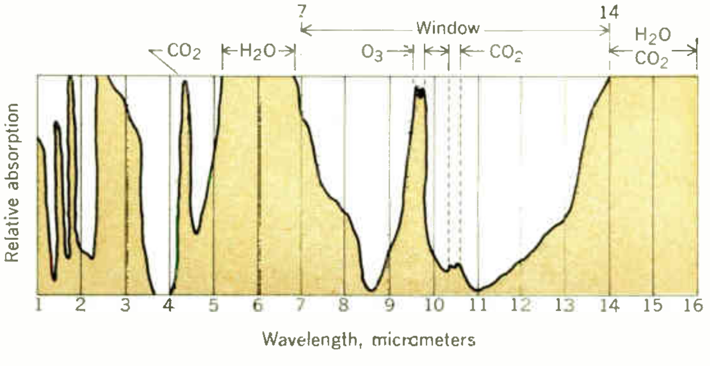
The overall conclusion drawn from this portion of the investigation was that one cannot exclude the possibility that moths do indeed detect electromagnetic radiation, but that they use this ability to monitor such environmental characteristics as humidity and temperature, rather than pheromone—which is more likely detected by olfaction.[15]

Experimental analysis

In our experimental analysis we are attempting to resolve the question of whether a behavioral response to infrared signals exists in *Heliothis zea*. It was first our intent to elicit a response, and then to specify which component of the stimulus and which sensory modality were responsible for the response.

The first behavioral response we examined was the attraction of male moths to females. Moths of both sexes were placed in 25- × 32- × 11-cm boxes and the number of matings was recorded as a function of the length of antennae remaining after amputation.[16] The results indicated that the most effective condition was neither with intact nor with totally excised antennae, but when small portions of the antennae were amputated (3 mm). These results are not supported by those obtained in similar

FIGURE 4. Absorption spectrum of atmosphere in infrared region. (Cited by Callahan.[7]) Detailed spectrum depends partially on atmospheric conditions.



experiments made by Callahan, who found a continuing decrease in the number of matings as antenna length was decreased.[18] In still other experiments, female moths were caged inside sealed containers with special "windows" transparent to infrared radiation. No response could be elicited from males outside the cages, not even when the females were irradiated with a broadband infrared source.

The second behavioral response we examined was the well-known attraction of *H. zea* to light. We began by releasing moths at various orientations with respect to a line between the initial position and a tungsten light source. The moths were tethered to tiny boats floating on an artificial pond in a manner that did not significantly impede their flight (Fig. 5), except to constrain it to the horizontal plane. A slip of paper placed under the moth is sufficient to keep it at rest until measurements can start; then removing the paper initiates flight. The percentage of moths making directly for the light source decreased symmetrically about the zero axis as the magnitude of the release angle was increased (Fig. 6). The curve became highly asymmetrical when one of the eyes was covered, regardless of whether the antennae were impaired or not. The conclusion of this experiment was that the eyes must be the only sensory detectors responsible for the attraction to light. In fact, we found that the antennae actually inhibit the phototaxis, since the attraction increases when the antennae are amputated! This increase in response also appears to depend on the amount of antennae excised.[15]

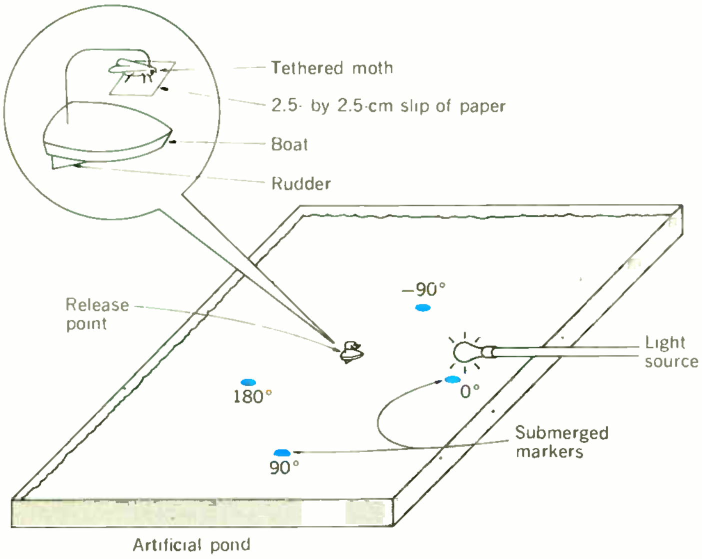
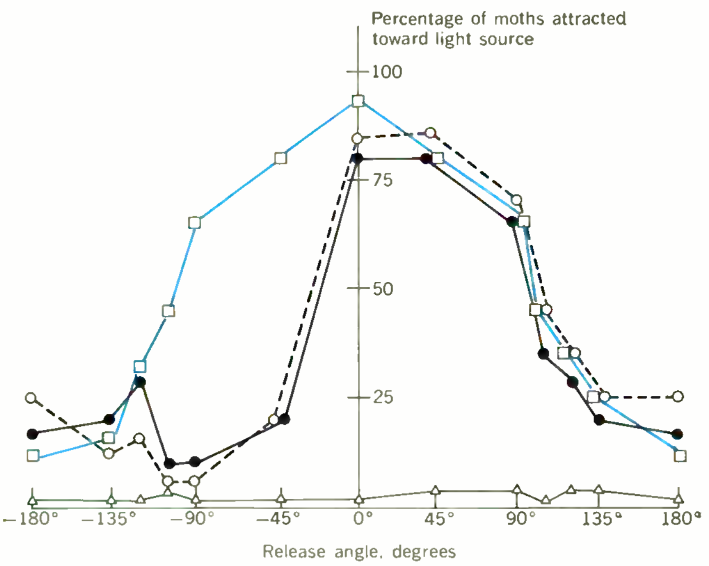


FIGURE 5. Apparatus used to measure attraction of corn earworm moth to light.[15]

FIGURE 6. Percentage of positive responses as function of release angles for normal and impaired moths.[15] (□ normal moth, ○ right eye painted over and both antennae excised, ● right eye painted only, △ both eyes painted.)



It is possible to quantify the attraction to light in another way, by measuring the turning response of the insect to a light source moving horizontally back and forth at various frequencies. The turning response is measured in terms of the relative air flow, as monitored by thermistors mounted behind each wing. By this method, first proposed by Roeder,[26] and using averaging techniques involving a "Lab 8" program on a PDP-8/I computer, we found that the moth tracked the light well at frequencies below 0.1 Hz but that the phase lag reached 90° at frequencies as low as 0.2 Hz. This would be another way to determine whether dynamic response to light is affected when parts of the antennae are excised.

To complete our analysis on the attraction of this insect to light we measured its spectral sensitivity. The moth was tethered on a long, lightweight wire suspended from a high ceiling and allowed to fly about in a confined area. A light source whose spectrum and intensity could be controlled was presented to the moth. The position of the moth was monitored by a tiny wire extension of the pendulum wire that dipped into an electrolytic tank containing the potential-divider arrangement shown in Fig. 7. (This arrangement also serves to confine the flight to what is essentially the horizontal plane.) The moth's position could then be read off as a pair of voltages and displayed on an *X-Y* recorder. Again the results could be quantified by chopping the recorder trace so that it consisted of a succession of dashes; the number of dashes in each quadrant of the trace area was then proportional to the time the moth spent in each during its sometimes erratic flight to the light source. (Later, automatic counting was introduced.)

This technique proved to be quite versatile. It permitted a determination of the spectral sensitivity of the insect by measurement of the response to variations in the spectrum of the light source and recording of the intensity needed to elicit equivalent responses. The sensitivity curves are shown in Fig. 8. There are two threshold levels. The higher of these is the level needed to initiate a response to light; the lower is that required for maintained response. The moth is more sensitive in the ultraviolet region than humans but no measurable response could be elicited in the near and intermediate IR range (0.7–20 μm). This is the first extension of insect sensitivity measurements to such long IR wavelengths.

By the same technique we also determined that the moth did not respond selectively to polarized light; nor did the spectral sensitivity curves change significantly for antennaless moths.

The results of this experiment not only appear to dispel the notion that infrared radiation from light sources attracts insects, but also allow us to predict the effective trapping distance for conventional light traps. The results suggest that the maximum distance for a typical "black" (i.e., ultraviolet) light trap is about 100 meters.

We are now measuring a third behavioral response: the flight response to various chemicals. Although we have determined that the corn earworm moth does not appear to respond to broadband infrared radiation, we have not ruled out the possibility that it is sensitive to a specific spectrum pattern, such as the absorption spectrum of its pheromones. In fact, the previously mentioned experiments of Bruce with spiny rat mites (which have no eyes) show that they respond selectively to incoherent IR radiation in the narrow band between 4.4 and 4.6 μm.[17]

Conclusions

In the present state of our research, we cannot unequivocally state that electromagnetic communication exists among moths, but neither can we wholly exclude the possibility of *some* response to infrared radiation. The most we can claim is to have reduced the problem to the point at which we can begin to make some substantive statements as to the likelihood of such a communication scheme.

From our systems analysis we have determined that the antennae indeed possess structures physically capable of infrared reception. But it is unlikely that moths use this scheme for long-distance mating communication. It is possible that environmental qualities such as temperature and humidity are sensed electromagnetically, but mating communication is more likely an olfactory process—which in turn may depend on some sort of electrical signal processing.

From our experimental analysis we have shown that *Heliothis zea* appears to be insensitive to broadband infrared sources. The infrared component in lights is not responsible for the insect's phototaxis. This attraction is attributable only to the compound eyes and not to the antennae. Nor has any behavioral response to polarized light been elicited. Mating behavior is mediated by the antennae. Infrared radiation does not seem to be responsible for the mating behavior; however, further experiments are under way to determine the nature of *H. zea*'s response to particular compounds.

In addition, several significant new techniques have been developed in the course of this research: photographic techniques using the scanning electron microscope, a procedure for staining and fixing transmission electron microscope sections, several novel methods of tethering and observing the behavior of flying moths, and an interesting strategy for mathematically describing the flight response of moths to light.

Design applications likely to result from this research include the following. Light traps for *H. zea* should concentrate energy in portions of the spectrum visible to the insect. Broadband infrared energy present in many existing light traps is wasted and will not increase the catch. The maximum trapping radius of any light trap may be estimated from its spectrum. Hence, in principle, the best spatial distribution of light traps in a given area for sampling or controlling insect populations may be calculated. The spectral sensitivity curve for *H. zea* suggests that the design of light traps might be improved by adding another color component to the light, with a re-

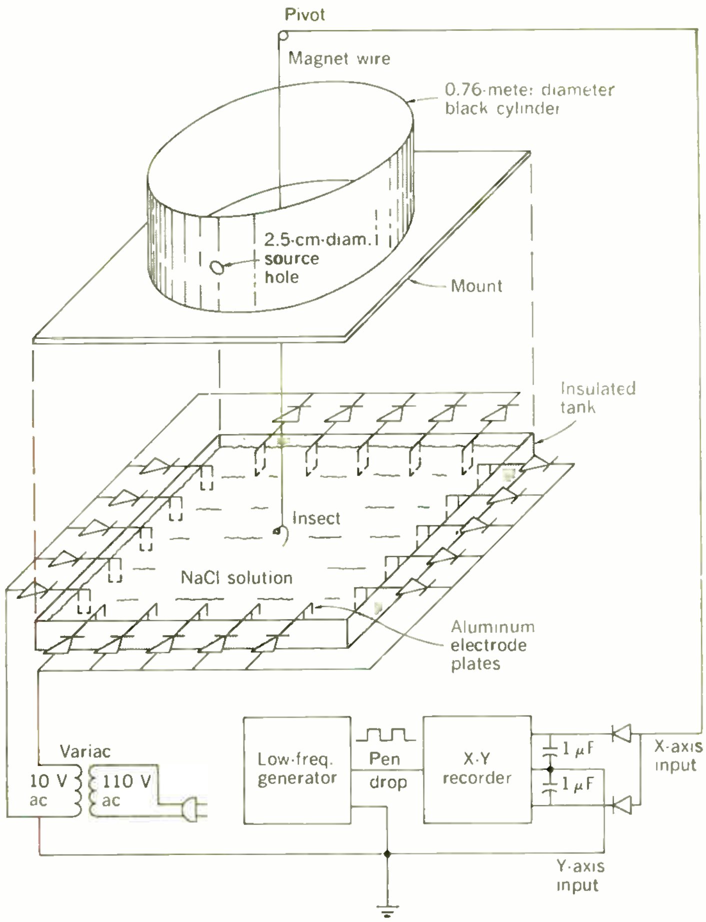
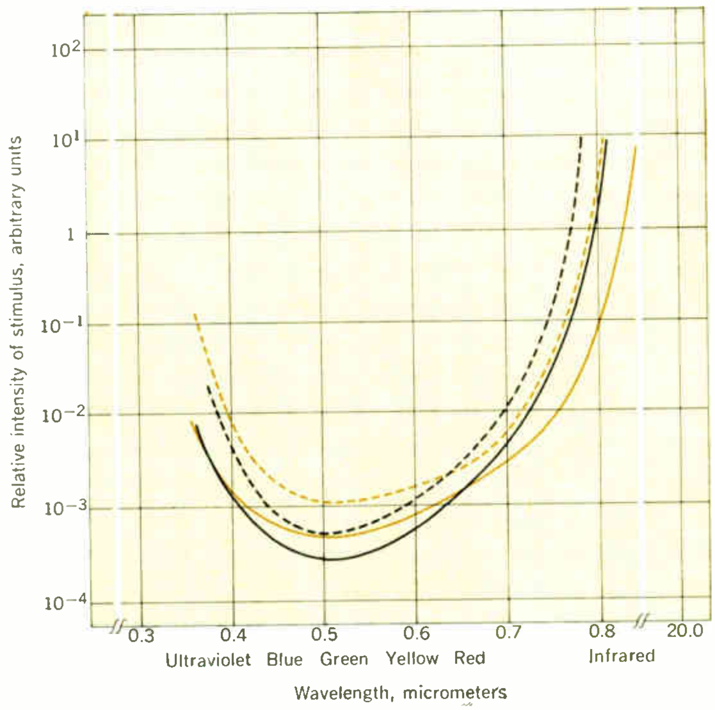


FIGURE 7. Apparatus used to measure absolute spectral sensitivity of corn earworm moth.[16] Mount is shown in raised position for observation purposes.

FIGURE 8. Absolute spectral sensitivity with and without antennae. (Bold lines, normal insect; fine lines, insect without antennae; solid lines, thresholds for continuing response; dashed lines, thresholds for the initiation of response.)



sulting improvement in its effective range. Other conclusions should be forthcoming as the work continues.

Bioengineers engaged in these investigations find the work particularly rewarding because they have the satisfaction not only of contributing to the store of basic knowledge in an important branch of the life sciences, but also of helping to solve problems of practical importance in environmental control by methods potentially less harmful than those now employed.

The aid of colleagues, notably Berkeley entomologist David L. Wood, in these investigations and in reviewing the manuscript is gratefully acknowledged. The investigation has been carried out under the terms of Cooperative Agreements 12-14-100-9044(33) and 10356(33) between the University of California and the U.S. Department of Agriculture. Y. Y. Zeevi assisted in the operation of the scanning electron microscope, which was purchased under National Science Foundation grant GB-6428 and whose operating costs are underwritten by grant GM-15536 from the National Institutes of Health.

REFERENCES

1. Riley, C. V., "The senses of insects," *Insect Life*, vol. 7, pp. 33-41, 1895.

2. See, for instance, the translation by Bernard Miall of part of this work: Fabre, J. H., *Social Life in the Insect World*. London: T. Fischer Unwin, 1912. However, Fabre concluded, after some experiments with the Great Peacock butterfly (Emperor moth) in which females were kept in containers of various nonmetallic materials, that "we must abandon the idea that the butterfly has any means of communication comparable to our wireless telegraphy" (p. 195).

3. Grant, G. R. M., "The sensory pits of insects considered as dielectric wave guide and resonators of infrared rays," *Proc. Roy. Soc. Queensland*, vol. 60, pp. 89-98, 1948.

4. Duane, J. P., and Tyler, J. E., "Operation Saturniid," *Interchem. Rev.*, vol. 8, pp. 25-28, 1950.

5. Callahan, P. S., "Far-infrared emission and detection by night-flying moths," *Nature*, vol. 207, pp. 1172-1173, 1965.

6. Callahan, P. S., "An infrared electromagnetic theory of diapause inducement and control in insects," *Ann. Entomol. Soc. Am.*, vol. 58, pp. 561-564, July 1965.

7. Callahan, P. S., "Intermediate and far-infrared sensing of nocturnal insects: Part 1, Evidences for infrared electromagnetic theory of communication and sensing in moths and its relationship to the limiting biosphere of the corn earworm," *Ann. Entomol. Soc. Am.*, vol. 58, pp. 727-745, Sept. 1965.

8. Callahan, P. S., "Intermediate and far-infrared sensing of noctural insects: Part 2, The compound eye of the corn earworm *Heliothis zea* and other moths as a mosaic optic-electromagnetic thermal radiometer," *Ann. Entomol. Soc. Am.*, vol. 58, pp. 746-756, Sept. 1965.

9. Callahan, P. S., "A photographic analysis of the moth flight behavior with special reference to the theory for the electromagnetic radiation as an attractive force between sexes and to host plants," *Proc. 12th Internat'l Congr. Entomol. London, 1964*, p. 302, 1965.

10. Callahan, P. S., "Are arthropods infrared and microwave detectors?" *Proc. N. Central Branch Entomol. Soc. Am.*, vol. 20, pp. 20-31, 1965.

11. Callahan, P. S., "Insect molecular bioelectronics: A theoretical and experimental study of insect sensillae as tubular waveguides, with particular emphasis on their dielectric and thermoelectret properties," *Misc. Publ. Entomol. Soc. Am.*, vol. 5, pp. 315-347, June 1967.

12. Callahan, P. S., "The scape and pedicel dome sensors: A dielectric aerial waveguide on the antennae of night-flying moths," *Ann. Entomol. Soc. Am.*, vol. 61, pp. 934-937, July 1968.

13. Callahan, P. S., "A high-frequency dielectric waveguide on the antennae of night-flying moths (Saturniidae)," *Appl. Opt.*, vol. 7, pp. 1425-1430, Aug. 1968.

14. Callahan, P. S., "The exoskeleton of the corn earworm *H. zea* (Lepidoptera: Noctuidae) with special reference to the sensilla as poly/tubular dielectric arrays," Res. Bull. 54, Univ. of Georgia College of Agriculture Experimental Stations, Jan. 1969.

15. Hsiao, H. S., and Hackwell, G. A., "The role of the antennae and compound eyes of the moth *Heliothis zea* in its attraction to light," *J. Insect Physiol.* (in press); and Griffith, P. H., and Süsskind, C., "Electromagnetic communication vs. olfaction in *Heliothis zea* (Boddie)," in preparation.

16. Süsskind, C., and Staff, "Response of insects to electromagnetic radiation," Rept. 68-2, Electron. Res. Lab., Univ. of Calif., Berkeley, 1968; and Griffith, P. H., "Investigation of insect electromagnetic communication with special reference to *Heliothis zea* (Boddie)," Rept. 68-3, Electron. Res. Lab., Univ. of Calif., Berkeley, Aug. 1968.

17. Bruce, A. W., "The perception of infrared radiation by the spiny rat mite *Laelaps echidnian* (Acari: Laelapidiae)," Ph.D. thesis, Ohio State Univ., 1969.

18. Callahan, P. S., Private communication, Nov. 1969.

19. Jacobson, M., *Insect Sex Attractants*. New York: Interscience, 1965.

20. Laithewaite, E. R., "A radiation theory of the assembling of moths," *The Entomol.*, vol. 93, pp. 113 and 133, 1960.

21. Wood, D. L., *et al.*, "Response of *Ips confusus* to synthetic sex pheromones in nature," *Science*, vol. 158, pp. 1373-1374, 1968.

22. Shorey, H. H., and Gaston, L. K., "Sex pheromones of Noctuid moths: III, Inhibition of male responses to the sex pheromone in *Trichoplusia ni* (Lepidoptera: Noctuidae)," *Ann. Entomol. Soc. Am.*, vol. 57, p. 775, 1964.

23. Borden, J. H., and Wood, D. L., "The antennal receptors and olfactory response of *Ips confusus* (Coleoptera: Scolytidae) to male sex attractant in the laboratory," *Ann. Entomol. Soc. Am.*, vol. 59, p. 253, 1966.

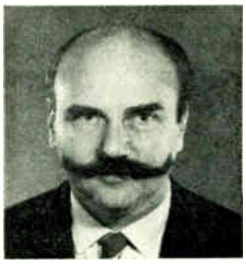
24. Schneider, D., "Insect antennae," *Ann. Rev. Entomol.*, vol. 9, p. 103, 1964; also "Insect olfaction: Deciphering system for chemical messages," *Science*, vol. 163, pp. 1031-1037, 1969.

25. Callahan, P. S., "Infrared stimulation of nocturnal moths," *J. Georgia Entomol. Soc.*, vol. 1, no. 4, pp. 6-14, 1966.

26. Roeder, R. D., "A differential anemometer for measuring the turning tendency of insects in stationary flight," *Science*, vol. 153, pp. 1634-1636, 1966.

Henry S. Hsiao, at present a graduate student at the University of California, Berkeley, received the bachelor's and master's degrees in electrical engineering in 1965 and 1967, respectively, from the Massachusetts Institute of Technology. At M.I.T. he was affilated with the Communication Biophysics Laboratory and worked in the field of psychoacoustics. He is now a member of the bioelectronics group at Berkeley, pursuing a doctorate under the terms of a National Institutes of Health training grant administered by the University of California College of Engineering.

Mr. Hsiao's current interests are in vision, eye movements, and electromagnetic communication among insects. He is a member of several societies, including Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.



Charles Süsskind (F) received the B.S. degree from the California Institute of Technology in 1948 and the M. Eng. and Ph.D. degrees from Yale University in 1949 and 1951, respectively, all in electrical engineering. He was a research associate and lecturer at Stanford University from 1951 to 1955 and then joined the faculty of the University of California, Berkeley, where he is now a professor of engineering science. Dr. Süsskind has made extensive contributions to the fields of microwave engineering, electron optics, and biomedical electronics; in the last, he has specialized in biological effects and perception of electromagnetic radiation. He is also known to Spectrum readers for his avocation, the history of technology, through a continuing series of articles on the early history of electronics, as well as other contributions. He edited The Encyclopedia of Electronics (1962), and coauthored (with Marvin Chodorow) Fundamentals of Microwave Electronics (1964) and (with Lynn Schell) Exporting Technical Education (1968). He is currently working on two other books, one on technology for the general reader and another on the history of radar.

Man–machine conversational data communication systems

With the increasing use of conversational data communication techniques, it has become necessary to set up standards to assure the most efficient implementation of existing designs—and to evolve new systems that will meet, and raise, these standards

Hideo Ohigashi Electrotechnical Laboratory, Tokyo

A brief discussion of the basic elements of data communication systems and a review of the current state of the art provide a prelude to the main theme of this article—a description of two inventions in this field by the writer. One is a data communication system; one a conversational information-processing system.

On-line, real-time systems, in which the computer is used via remote terminals, and time-sharing systems (TSS), in which a number of users share a computer through individual terminals, are becoming increasingly popular and thus it has become necessary to set up standards for conversational data communication techniques. The International Organization for Standardization (ISO) has established a specification for basic mode procedures, but this is designed primarily for off-line data communication and is not completely suitable for conversational communication. However, the consideration of standards for conversational-mode procedures is currently a prime topic on the ISO agenda.

This writer has designed a communication control system for the Electrotechnical Laboratory's TSS (ETSS). The system, which was developed under the sponsorship of the Japanese Government's National Computer Development Project, has been providing satisfactory service over the telephone lines of the Nippon Telegraph and Telephone Public Corporation.

The designs described in this article are based on experience with the ETSS. The proposed data communication system, which has been submitted to the ISO for consideration, will be used at Expo '70 in Osaka.

Current state of the art

Terminals. Among the requirements for on-line terminals are satisfactory performance answering the needs of the individual system, ease of operation, high reliability, low cost (particularly when a large number of terminals are involved), and other factors, such as low noise, small physical size, and ease of maintenance.

Input–output devices presently in use are listed in Table I. The most widely employed of these is a combination of a typewriter and a paper tape device.

In practice, the greatest problem for the terminal is the printer speed. To meet the requirement, a serial printer featuring a speed of 20 characters per second was recently developed and is now in use in Japan; it is shown in Fig. 1. The reason for this new printer lies in the printing mechanism; see Fig. 2. The serial printer uses the type-cylinder method, in which the types are selected by a combination of cylindrical and linear techniques (in contrast to the type-bar method, in which each type bar bears a piece of type on its tip, or the pallet method, whereby the types are selected by moving the pallet in the direction of the X- and Y-axes). In the type-ball method, types are chosen by two orthogonal angular selections and in the type-wheel method, selection is provided by moving and rotating a wheel or wheels.

Character-display terminals are in wide use today and offer a number of advantages as compared with typewriters, such as high speed, excellent editing capabilities, lack of noise, and high reliability. Their disadvantages are the difficulty of obtaining hard copy and their high cost.

The capability for high-speed reception of character-display terminals is a very attractive attribute when high-speed (10^4-baud) communication lines are available. As for editing capability, on a display terminal it is very easy to delete or change characters to be transmitted. In some cases, corrections or insertions are made per line. These are possible because characters are not printed as on a typewriter. Another asset of the display terminals is their quiet operation.

Type of communication. Both half-duplex and full-

I. Examples of terminal devices

Device	Application
Paper input–output	Typewriter
Tape input–output	Paper tape equipment, magnetic tape equipment
Card input–output	Card equipment
Paper output	Line printer (character), X-Y plotter (graph)
Display output	Character display (character), graphic display (graph)
Sound input–output	Telephone

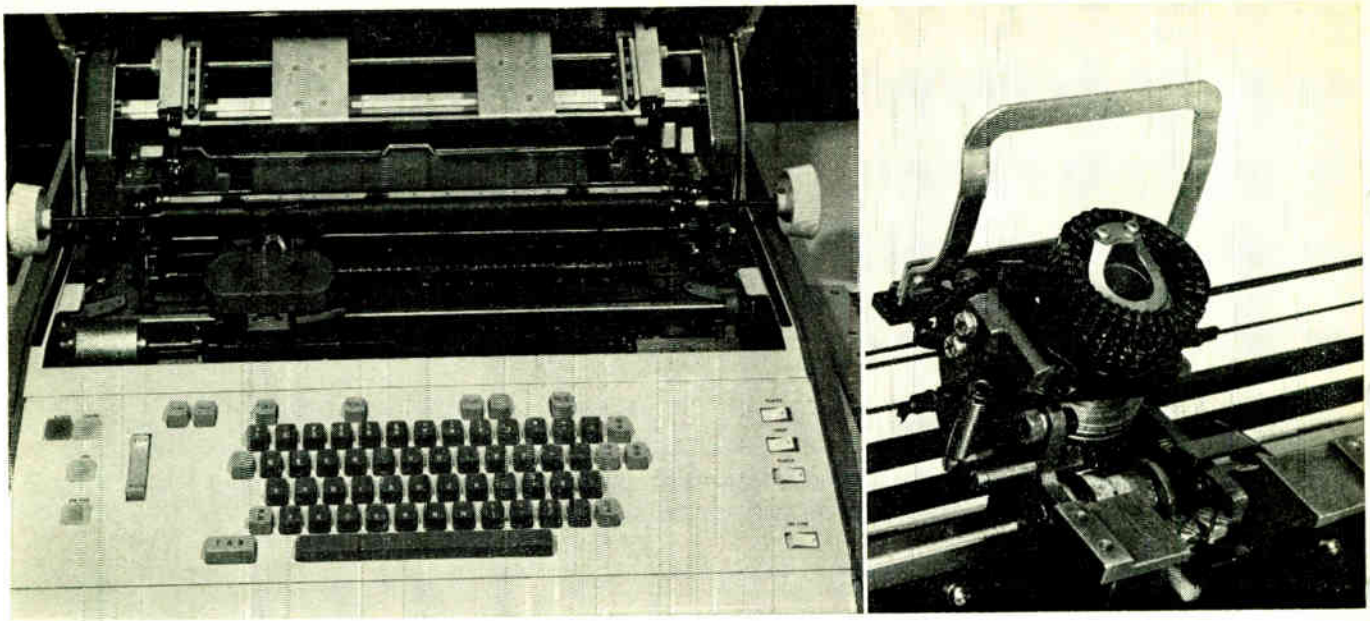


FIGURE 1. The 20-character-per-second printer. (Courtesy OKI Electric Ltd.)

duplex systems are used for on-line communication. In the half-duplex system, messages are sent in one direction at a time and two-way communication is provided by switching the direction of transmission. In a full-duplex system, messages can be sent in two directions simultaneously.

A full-duplex system is required for (1) bilateral message transmission and (2) a situation in which it is necessary to transmit a message back while it is being received if error control is provided by the information feedback system in a conversational system.

Error control. There are various error-control methods now in practical use; these can be roughly classified as follows (see Table II):

1. One in which errors in the message are detected at the receiving station and the sending station is notified so it can take proper action, such as retransmission.
2. One in which errors in the message are corrected at the receiving station.
3. One in which errors in the message are corrected at the sending station.

The repeated transmission system is a system in which a message is repeatedly (say three times) transmitted and each character in the message is determined at the receiving station by a process of comparing and taking the majority. However, the repeated retransmission of each message results in an effectively slow rate of transmission, and thus this method is not suitable for conversational systems.

In the information feedback system, the receiving station sends the message received back to the sending station, which compares it with the original copy for checking by identity. Redundancy in the transmission is intrinsically high in this system, as it is in the repeated transmission system. However, the redundancy does not necessarily constitute a shortcoming if simultaneous two-way conversation is not required—because it is possible to receive and send a message back at the same time over a full-duplex communication line. The communication control system, though, must be so designed that when the received message is returned the central processing unit (CPU) is bypassed, in order to reduce the load on that unit.

In the decision feedback system, a message is treated as a text, and a text is preceded by a start-of-text (or text-block) character (STX) and followed by an end-of-text character (ETX). A long text, however, is divided into a number of blocks, each block being followed by an end-of-text-block character (ETB) for the sake of efficient transmission. Upon receiving such a text or block, an acknowledgment character (ACK) is returned as the response character if the received information is error-free. A negative acknowledgment character (NAK) is returned as the response character if the received information is in error; this notifies the sending station of the need for retransmission. It is also customary to transmit an inquiry character (ENQ) from the sending station to ask for the response if the response character did not reach the sending station. Thus, although more efficient transmission can be expected, the control will be rather complicated as compared with the repeated transmission or information feedback systems.

The group counting method is an improved version o the horizontal parity scheme, in which two or more checking characters are used for greater dependability.

In the constant mark scheme, only those codes that contain r 1's (there are ${}_nC_r$ such codes) are utilized and the $2^n - {}_nC_r$ codes are not used, resulting in high redundancy.

Also practicable are systems such as those in which special information is appended to the message; for instance, the number of digits when a numerical value is transmitted.

In the error-correcting-code systems, errors are corrected at the receiving station and the communication system is not concerned with error control. Here, however redundancy is high, and it also should be noted that the CPU may not process the data in time. This occurs be-

II. Error-control systems

	Detection	Correction
Receiving station	Decision feedback system Error-detecting code Vertical parity check Horizontal parity check Group count check „C_r code Special information	Error-correcting code Hamming code Cyclic code Repeated transmission system
Sending station	Information feedback system	

cause the message exchanged may be partitioned into a number of blocks (this process is called blocking) in order to achieve efficient line buffering for the system, and in those blocking systems in which a response is not made for each block, as in the case of this error-control scheme, it is possible that a number of blocks arrive at the computer within a short time and the computer cannot accommodate them. This indicates that blocking is not desirable when an error-correcting code is used.

Transmission control procedures. The ISO has established transmission control procedures for sending messages. These basic-mode procedures go through the following phases, as shown in Fig. 3: (1) Phase 1, in which linkages to the switch network are made. (2) Phase 2, in which the data link is established by polling and selecting. (3) Phase 3, in which the master station sends out the message. (4) Phase 4, in which the master station sends out an end-of-transmission character (EOT) and unlinks the data link when it has transmitted the message. The control returns to the control station in those systems in which the control station is independent of the communicating parties. (5) Phase 5, in which linkages to the switch network are canceled.

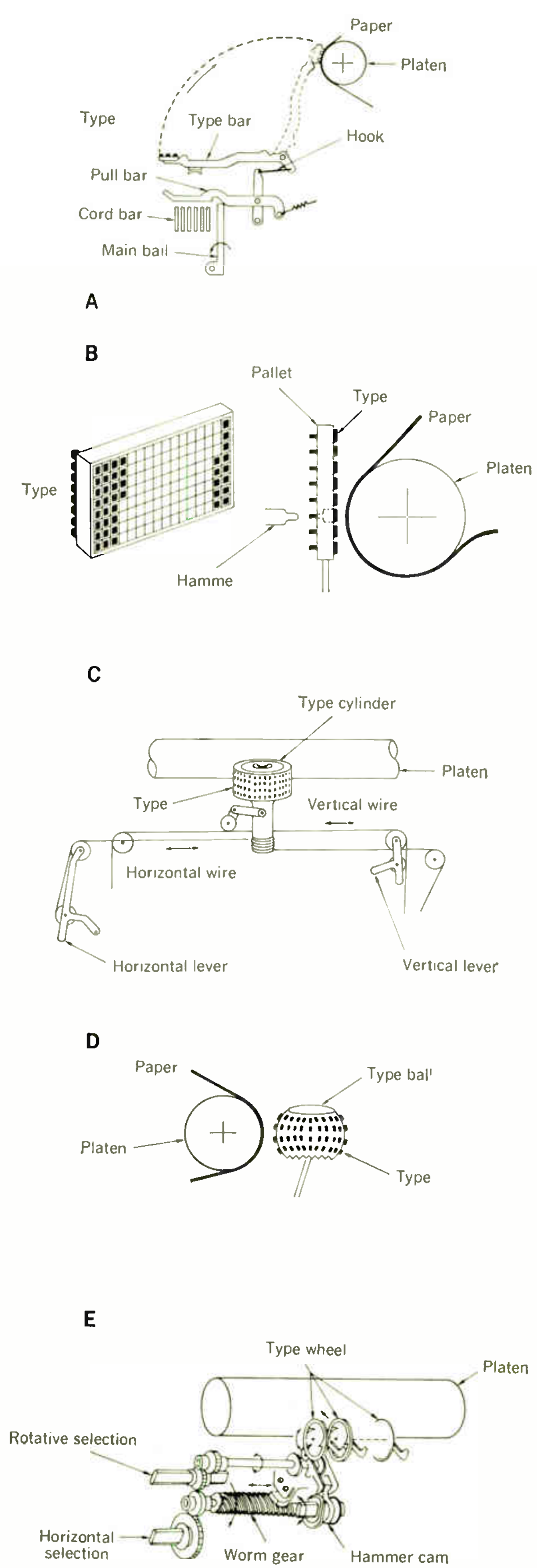
Polling, in phase 2, is a measure that can be taken only by the control station; the polled station becomes the master station if it has a message to be sent (a station that can transmit is called a master station or a send-mode station and a station that receives the message sent by the master station is called a slave station, or a receive-mode station in the half-duplex system).

Selection, as in phase 2, is a procedure carried out by a master station for selecting a slave station and requesting it to be ready for message reception. The selected station returns an ACK when it is ready for receiving and becomes a slave station, and it returns a NAK if it is not ready. Figure 3 shows a case in which the slave station is also the control station; systems in which the control station is independent of the master and slave stations are also feasible.

The ISO has not specified details of phases 1 and 5 as the matter is closely related to the communication network.

In the basic-mode procedures, phases 2, 3, and 4 are followed at every inversion of the direction of transmission. Therefore, the basic-mode system has the following disadvantages with regard to conversational data communication. Since the direction of message transmission is reversed frequently, losses in time and increases in the amount of processing will result if phases 2, 3, and 4 are repeated for every change of direc-

FIGURE 2. Example of printing mechanisms. A—Type-bar method. B—Type-pallet method. C—Type-cylinder method. D—Type-ball method. E—Type-wheel method.



tion. Also, the interruption function (QUIT) is not available.

Proposed systems

Conversational data communication systems. The system proposed in this article completely eliminates all of the shortcomings of the ISO basic-mode procedures.

An immediate countermeasure against the increase in processing and loss of time in the basic-mode procedures would be to execute phases 2 and 4 only at the beginning and end of a conversation, and to transfer only messages during the conversation. With this scheme, however, send/receive-mode switching is a problem.

In the writer's method, shown in Fig. 4, send/receive-mode switching is accomplished by a control character α. The send-mode station sends this α and becomes the receive-mode station. Aspects of timing of transmission will be described later.

For the sake of interruption, it is necessary that the system honor an interrupt at any time regardless of the send/receive mode. Various types are needed.

The scheme devised by the writer incorporates a particular transmission control character (called an interruption character and designated as β), and the reason for the interruption is specified by a text transmitted by the interrupting station following the interrupting action. To implement this scheme, the first requirement is that the system permit the sending of a β regardless of the send/receive mode, even during transmission. (For instance, the user should be allowed to depress the interrupt button on the terminal even if the terminal is in the receive mode and the keyboard is locked.) In other words, the system must be able to receive the β transmitted by the other party even when it is in the send mode or is transmitting in the send mode. The fulfillment of this requirement depends upon the type of communication system. In the full-duplex system, anything transmitted by the other party can be received at any time, and there is no problem. In the half-duplex system, however, the β cannot be transmitted while the terminal is receiving (although it is simple to implement hardware that will alloy the receipt of what is transmitted by the other party when the terminal is in the receive mode but not receiving). A method for coping with this problem is discussed in the following.

For interruption of the receive-mode station during reception, the fact is utilized that a long text is divided into a number of blocks. That is, when the receive-mode station wishes to interrupt the other party during reception of a text or text block, it transmits an interruption character (β) instead of an ACK at the time when an ACK is expected. Upon receiving a β, the send-mode station yields the right of transmission to the other station if there remain data to be transmitted.

One possible system that utilizes the β is one in which the receive-mode station autonomously switches to the send mode upon sending the β. Such a system, however, poses the problem that if the β does not reach the other party, both stations will switch to the send mode (when both have switched to the receive mode through some communication error, the system can get out of deadlock by issuing a reminder at intervals). In the proposed system, this problem has been solved by using the α described earlier for responding to the β as well.

In practice, the user depresses the interrupt button while he is receiving a message (Fig. 5), whereupon the transmission controller at the terminal transmits a β after having received the text block (or text). When the computer receives the β it replies with the α even if there remain data to be transmitted and yields the transmission right to the terminal. The text specifies the meaning of the interrupt. Also, the proposed system is so designed that the β can be transmitted when the terminal is in the receive mode but not receiving, and the send-mode station can receive the β as well if it is not transmitting. The terminal hardware is so designed that when the STX transmitted by the computer and the β transmitted by a terminal (or vice versa) occur simultaneously, the signal transmitted by the computer is honored and that transmitted by the terminal is nullified, and the terminal reacts.

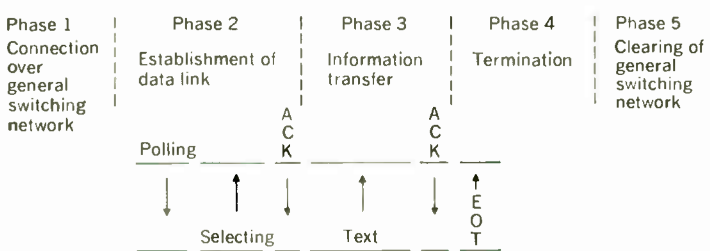


FIGURE 3. The ISO basic-mode procedures.

FIGURE 4. Proposed send/receive-mode switching.

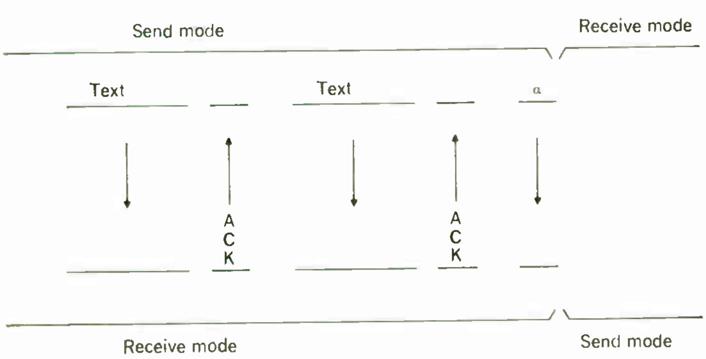
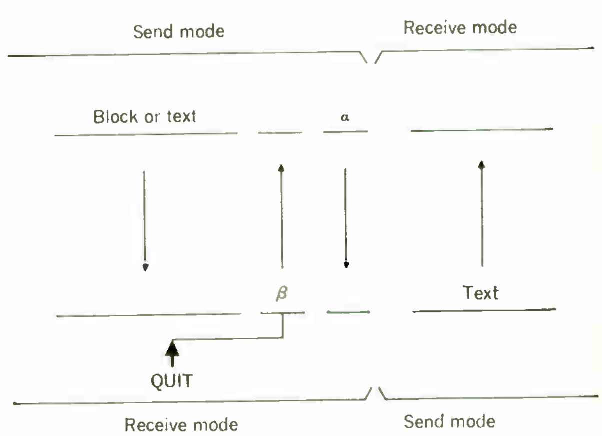


FIGURE 5. Interruption procedure.



Mode control system. The theory underlying the proposed system is pertinent to one desirable form of conversational information-processing or terminal control system. The philosophy is general, and independent of the system hardware, terminal device, and type of communication and error-control method. The actual form that implementation takes, however, does depend on the last two factors. To describe all of the possible error-control designs would be impractical here, and so the decision feedback system will be used as an example. As for the communication system, the half-duplex system is described as everything that can be done in it can also be done in a full-duplex system.

There are many conditions that a man–machine system must fulfill; primarily, it should (1) provide man with as much service as possible, and (2) be able to accept an instruction from man at any time, and to obey that instruction to man's satisfaction. The system in this sense is the complete entity of hardware and software. As a man–machine conversational data communication system is closely related to man, the user of the terminals, these requirements are important to him. It should be noted that the data communication system described here includes the communication and terminal controls.

The most important characteristics of the man–machine conversational data communication system as compared with the interterminal data communication system is that the sequence and time of occurrence of the READ and WRITE requests is random. For instance, it is common for one WRITE request to follow another WRITE request. Also, occasionally it is desirable to let the computer interrupt the user's sending under a READ request. Based on this line of reasoning, the writer believes that the most important function with respect to conditions 1 and 2 of a man–machine system such as the TSS are as follows:

1. When the computer does not want the user to send a message, it prevents him from doing so. (In a system in which the user is not even allowed to prepare a message for future sending while he is being prohibited from sending, the prohibition takes the form of keyboard lock.) And the computer permits the user to send from one terminal only when the user is supposed to send under a READ request (whereupon the ready-to-send signal is turned on, and the keyboard is unlocked in a keyboard lock system). In other words, it is necessary that the computer control the send/receive mode as specified in Table III, depending upon the READ/WRITE request, and the IDLE state (a state not under a READ or WRITE request). Such a system, as developed by the writer, is called a mode control system.

2. The user should be able to depress the interrupt button on the terminal (or to take an equivalent action) at any time (even during reception), and the computer should be able to recognize the interrupt immediately and answer the user's request.

Figure 6 shows an example of a mode control system. When a terminal sends a text under a READ request and the computer returns an ACK, the terminal always returns an α and switches to the receive mode. When the computer receives an ACK but the computer does not transmit the α, it remains in the send mode (IDLE state). When a READ request is issued, the computer transmits an α to yield the transmission to the terminal.

It should be noted that, although the foregoing description has been on the basis of a half-duplex system, the design can be implemented on a full-duplex system if a circuit for distinguishing the reception of an α is incorporated.

The mode control system has a number of advantages:

1. The ready-to-send signal on the terminal is turned on when a READ request is issued and thus it is easy for the terminal user to transmit.

2. The terminal switches to the send mode only under a READ request. Therefore, when the type and sequence of requests are known, the terminal can communicate automatically with the computer, with outgoing data previously prepared on the tape or card equipment. With such an arrangement, the user need not stay at the terminal, thus eliminating the labor of manually switching to and from the keyboard and the tape or card reader every time switching is required. On-line time will be saved by making full use of the tape or card equipment at remote terminals as described in the foregoing. This means that the mode control system is suitable for data communication between machine and machine.

3. The computer controls the mode of a terminal in the mode control system; therefore, the computer could yield the transmission right to the terminal after the line buffer has been prepared in the case of a READ request (a WRITE request does not engender a problem) so that buffering for the communication line would be easily implemented.

4. When it is desirable to cancel a READ request from the computer side while the user is sending under a READ request, the computer can interrupt the transmission.

Discussion

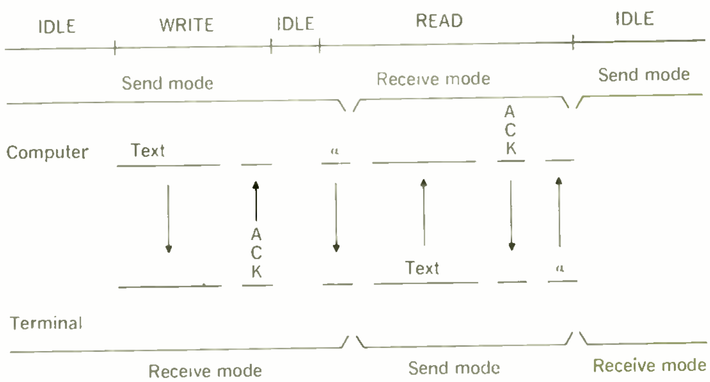
In the foregoing we have presented the conversational system developed by the writer. However, other designs for such a system may be feasible also; some of these are discussed here.

Send/receive-mode switching. 1. In scheme 1, the receiving station switches to the send mode, responding by an ACK upon receiving a text, and the party originally in the sending mode switches to the receive mode upon re-

III. Mode control system

I/O Request	Computer	Terminal
READ	Receive	Send
WRITE	Send	Receive
IDLE	Send	Receive

FIGURE 6. Example of a mode control system.



ceiving an ACK after transmitting a text. When the response is NAK, however, neither side switches to the other mode. In the case of consecutive WRITE requests, WRITE 1, WRITE 2, the computer switches to the receive mode upon transmitting a message in the form of a text under the first request WRITE 1, whereby the computer end cannot transmit under the next request WRITE 2 unless the terminal transmits a dummy text STX ETX to yield the transmission right to the computer.

In this system, the receive-mode station autonomously switches to the send mode upon transmitting an ACK after it has received a text. This leads to various difficulties: (1) When an ACK has been mistaken for a NAK, both stations will be trapped in the sending mode. (2) When an ACK does not reach the other party, control becomes complicated.

2. In scheme 2, send/receive switching information is appended to the text. Therefore, when a message is transmitted in the form of a text, the next I/O request is examined, and an appendix indicates if the next I/O request is READ or WRITE, or if it is a part of consecutive WRITE requests. In such a system, a WRITE request cannot be executed in most cases until the next I/O request has been examined; this is considered a deficiency of the system. Furthermore, the process of switching the send/receive mode cannot be clear cut or rational with this method. For instance, it is a disadvantage of the method that both stations will be switched to the send mode if an STX is used for the switching and the STX does not reach the other station.

3. The only difference between schemes 2 and 3 is that the send/receive-mode switching information is contained in the text in scheme 2 whereas it is in the heading in scheme 3. Thus, scheme 3 has the same disadvantages as scheme 2, in addition to the fact that where a heading is added the text assumes the form SOH Heading STX Data ETX, instead of STX Data ETX, as in other designs. The SOH stands for "start of heading."

4. In the scheme shown in Fig. 4, a rather complicated algorithm is necessary because of the use of the α.

Mode control system. 1. Shown in the following is a method of incorporating the mode control system in a communication system in which the send/receive mode is switched each time a text has been sent out. When a message is sent from a terminal under a READ request, it is transmitted as a text, whereas the computer transmits a message under a WRITE request as text blocks and finishes when the data to be transmitted are finished without transmitting an ETX; thus, the computer remains in the send mode (IDLE state) with respect to the communication. If the next request is WRITE, the computer simply repeats the operation and yields the transmission right to the terminal when the next request is READ by transmitting a dummy text STX ETB.

This system has a disadvantage when the received message tape is to be relayed to the same or other systems. For instance, if the sequence of the I/O requests is WRITE 1, WRITE 2, followed by a READ, and an attempt is made to transmit only the tape produced by the WRITE 1, a dummy text STX ETX must be appended as the last tape punch is an ETB, and this is awkward.

2. A method of incorporating the mode control system in scheme 2 is introduced in the following. A message is sent in the form of a text when it is sent under a WRITE request. In doing so, however, a special character (designated γ) is appended to the text, specifying that the send/receive mode is not to be switched (in other words, the computer remains in the send mode). The computer transmits a dummy text without a γ to the terminal and switches the terminal to the send mode only when it encounters a READ request after the WRITE. With this design, if the sequence of the I/O request is WRITE 1, WRITE 2, followed by READ, the tape received at the terminal will be STX M1 γ ETX STX M2 γ ETX STX ETX (in this example, γ is included in the text), therefore, the terminal hardware must be modified so that the γs are not punched when the terminal is receiving, allowing a user to transmit the received tape only for the M1 without modifying it and without adding the dummy text STX ETX.

This system has the following disadvantages as well as those previously described in relation to scheme 2: a γ must be appended to the text for every WRITE request; terminal hardware must be designed so that the γs are not punched; and the tape or card contains redundant dummy texts STX ETX.

Conclusion

Recent work in the field of man–machine conversational data communications has been discussed. The practicality of one of the systems described, the data communication system invented by the writer, has been proved; however, he would appreciate criticism and further comment. A proposal for a conversational system is also presented, and at least one practical application of such a technique discussed.

It is to be hoped that the development of these systems will contribute in some measure to the establishment of improved international standards for the conversational mode.

Essentially full text of a paper presented at the 1969 IEEE Computer Group Conference, Minneapolis, Minn., June 17–19.

With regard to the development of the ETSS communication system, the writer wishes to acknowledge the support of workers at the Engineering Bureau of the Nippon Telegraph and Telephone Public Corporation, of Dr. S. Hayasaka and associates at the OKI Electric Industry Company, and of Drs. K. Noda, H. Nishino, and H. Aiso and Messrs. H. Kawai and K. Fuchi and other members of the ETSS staff.

REFERENCES

1. "Basic mode control procedures for data communication systems," Draft ISO Recommendation No. 1745, Dec. 13, 1968.

2. "Proposed USA Standard for communication control procedures for the USASCII," Apr. 1968.

3. "Contribution by ECMA on data communication control procedures," May 1968.

4. Lynch, W. C., "Reliable full-duplex file transmission over half-duplex telephone line," *Commun. ACM*, vol. 11, pp. 407–410, June 1968.

Hideo Ohigashi was born in Higashiosaka, Osaka, Japan, in 1936 and was graduated from Osaka Electrical Communication Engineers University in 1961. He has been with the Electrotechnical Laboratory, Tokyo, since 1962, where he has been chiefly engaged in work in the areas of arithmetic and communication control systems. Mr. Ohigashi is a member of the Institute of Electronic and Communication Engineers of Japan and of the Information Processing Society of Japan.

The metal–semiconductor contact: an old device with a new future

Recent improvements in Schottky-barrier diodes revived interest in these components. Their greatest potential application appears to be in integrated circuits for performing various gating and clamping functions

A. Y. C. Yu Fairchild Semiconductor

Advances in process technology are making possible the fabrication of Schottky barriers with reliable, ideal electrical characteristics. With these advances, one can anticipate a rapid increase in the utilization of Schottky barriers, not only as discrete devices but also as a new component in integrated circuits. This article presents a unified picture that quantitatively characterizes both the Schottky barrier and ohmic contacts on silicon. Fabrication techniques and various applications are also discussed.

Metal–semiconductor contacts can be classified into two groups according to their current–voltage characteristics: Those with rectifying characteristics are called Schottky barriers; those with linear characteristics are called ohmic contacts.

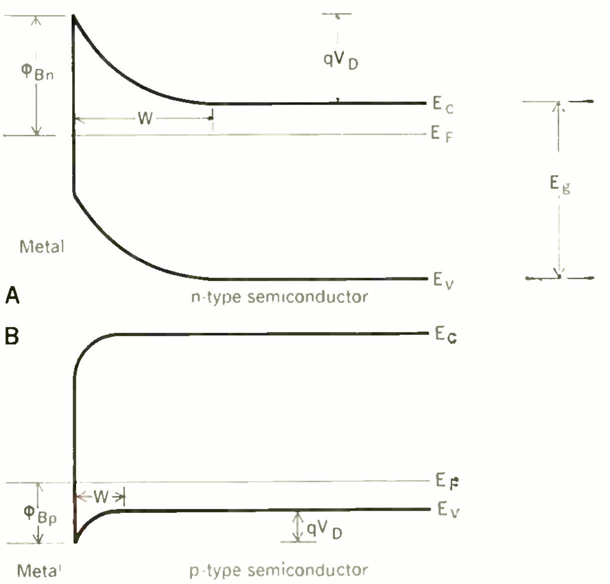
Metal–semiconductor contacts with rectifying characteristics were discovered and investigated by Braun at the end of the 19th century. Despite many attempts to understand their mechanism of current flow, the correct physical model—involving thermionic emission—was not discovered until half a century later by Schottky; hence the name, Schottky barrier.[1] Researchers at the Bell Telephone Laboratories in the late 1940s were actively pursuing the understanding of metal–semiconductor interfaces when they accidentally discovered the transistor. Since then, almost all efforts in the semiconductor industry have been directed toward p-n junction devices. It is not surprising, then, that as late as the early 1960s commercial Schottky barrier diodes were in a rather primitive stage compared with p-n junction diodes, although the former were discovered some 70 years earlier. Their uses, therefore, have been very limited because of their poor electrical characteristics and high processing cost.

Since the early 1960s, a great deal of progress has been made in Schottky barriers. The development of metal–silicide and refractory-metal contacts[2] made possible the high-temperature stability of these Schottky barriers. Surface effects on Schottky diodes[3] have been studied recently and are clearly shown to be responsible for most of the nonideal characteristics of these diodes. If p-n junction guard rings are incorporated around the periphery of the Schottky diode, the undesirable effects of the surface can be eliminated, and the resultant *I-V* characteristics are nearly ideal.[4-6] It was shown recently that passivated Schottky barriers with ideal characteristics can be made on n-type silicon with an extremely simple process in which no elaborate surface cleaning, p-n junction guard ring, or rare metal are needed.[7] It is now not only practical and economical to incorporate Schottky barriers into integrated circuits to take advantage of their unique electrical characteristics, but also feasible to manufacture reliable discrete Schottky barriers with excellent performance at a cost even lower than that of p-n junctions. These recent developments indicate that Schottky barriers undoubtedly will be utilized extensively.

Although ohmic contacts are made on all solid-state devices by various metals or alloys, the physical nature of these contacts is poorly understood. Within the past few years extensive work has been done to characterize the metal contacts on heavily doped semiconductors. Extensions of this work are beginning to elucidate the nature of ohmic contacts.

The purpose of this article is to review some of the more important features of metal–semiconductor contacts, such as their energy-band diagrams, their current–

FIGURE 1. Band diagram of Schottky barriers on (A) n-type and (B) p-type semiconductors.



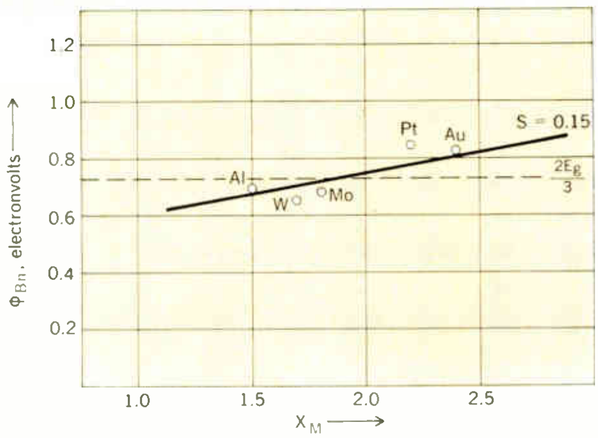
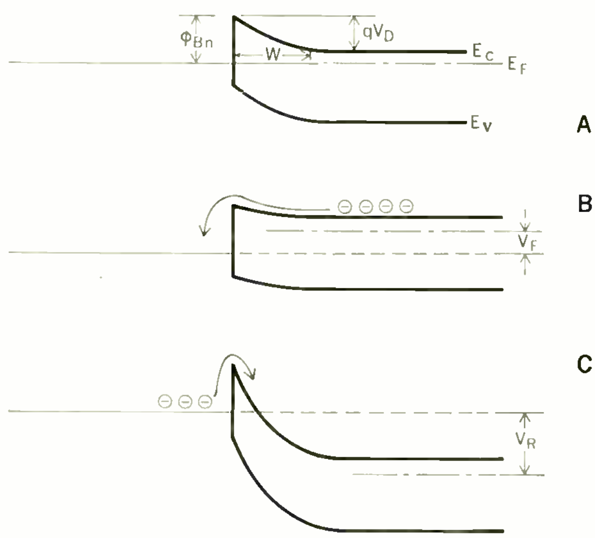


FIGURE 2. Plot of ϕ_{Bn} of several metals on silicon vs. the metal electronegativity. Data are taken from Yu and Mead[7] (Al); Crowell et al.[10] (W); Zettler and Cowley[6] (Mo); Kahng and Lepselter[2] (Pt); and Archer and Yep[11] (Au).

FIGURE 3. Band diagrams of Schottky barrier under con ditions of (A) thermal equilibrium, (B) forward bias, and (C) reverse bias.



voltage characteristics, their processing, and various applications. Most of the discussion is restricted to Schottky barriers on silicon, since silicon is by far the most important semiconductor material at this time. However, the majority of the statements also apply to Schottky barriers on other semiconductors.

Band diagrams of Schottky barriers

The thermal-equilibrium-band diagrams of Schottky barriers on n- and p-type semiconductors are shown in Fig. 1(A) and (B), where E_c is the conduction band edge, E_v is the valence band edge, E_g is the band gap, E_F is the Fermi level, ϕ_{Bn} and ϕ_{Bp} are the barrier heights for n- and p-type semiconductors, respectively, qV_D is the total band bending in the semiconductor, and W is the space-charge layer width. The band bending occurs in such a manner that it retards the flow of majority carriers into the metal; thus a "barrier" is present at the contact. This barrier arises because the metal and the semiconductor have different work functions (or electronegativities). For ionic crystals (such as ZnS) in which the surface-state density is low, the barrier heights with different metals vary as the electronegativity of the metal x_M.[8] That is, if we plot ϕ_{Bn} versus x_M of the metals, the slope of this plot is

$$\frac{d\phi_{Bn}}{dx_M} = S \tag{1}$$

and for ionic crystals S is very close to unity. However, for covalent semiconductors (such as Si, Ge, and GaAs), the disruption of the crystal lattice at the surface perturbs the electronic structure so much that high densiti of surface states exist in the forbidden gap. These states effectively screen the semiconductor bulk from the exterior charges such that the barrier heights are essentially independent of the metal electronegativity [i.e., S is very small in Eq. (1)] and have been found to be about two thirds of the band gap energy.[9] This is clearly shown Fig. 2, where ϕ_{Bn} values of several metals on Si are plott versus the electronegativity of the metals. Also shown are a solid line (with $S = 0.15$) and a dashed line (which is $2E_g/3$).

The barrier height is found to be essentially independent of semiconductor doping levels,[11] and therefore

$$\phi_{Bn} + \phi_{Bp} = E_g \tag{2}$$

Assuming that the doping in the semiconductor is uniform, the space-charge layer width of the Schottky barrier is exactly analogous to that of a one-sided step p-n junction:

$$W = \sqrt{\frac{2\epsilon_s(V_D + V_R)}{qN}} \tag{3}$$

where ϵ_s is the permittivity of the semiconductor, V_R is the reverse voltage,* and N is the doping level. The capacitance per unit area is then

$$C = \frac{\epsilon_s}{W} = \sqrt{\frac{q\epsilon_s N}{2(V_D + V_R)}} \tag{4}$$

This equation can be rewritten as

$$\frac{1}{C^2} = \frac{2}{q\epsilon_s N}(V_R + V_D) \tag{5}$$

Thus a plot of $1/C^2$ vs. V_R will be a straight line. The intercept will give V_D and the slope will give N. Capacitance–voltage measurements of Schottky barriers hav been commonly used to determine the barrier heig and the semiconductor doping.

Current–voltage characteristics

Metal–semiconductor contacts on lightly doped semiconductors. When the space-charge layer width is relatively thick (for example, $N_D < 10^{17}$ cm^{-3} i n-Si), the only means of electron transport at the interface is by thermionic emission over the barrier. Under forward bias (metal positive on n-semiconductor), there are many electrons with enough thermal energy to climb over the barrier into the metal; see Fig. 3(B). Since the

* Reverse bias is in such a direction as to increase the band bending in the semiconductor; metal is biased positive for p-type and negative for n-type.

barrier from the semiconductor side is reduced by the forward bias, an increase in forward bias will result in a rapid increase in the forward current. When the diode is reverse-biased, as in Fig. 3(C) (metal negative), electrons in the metal have to climb a barrier with height ϕ_B, which is essentially independent of bias voltage, in order to get into the semiconductor. Thus, in the reverse direction, the current is expected to be small and essentially bias-independent until the electric field in the depletion region is so high that avalanche breakdown occurs. This qualitatively explains the rectifying characteristics of a Schottky barrier.

The thermionic-emission current of majority carriers of a Schottky barrier may be described by the equation

$$I_{TE} = I_0[\exp(qV_F/kT) - 1] \tag{6}$$

$$I_0 = [RAT^2]\exp(-\phi_B/kT) \tag{7}$$

where V_F is the applied forward bias; k is Boltzmann's constant; T is the absolute temperature; q is the electronic charge; I_{TE} is the thermionic-emission current; I_0 is the straight-line extrapolation of I_{TE} to $V_F = 0$ (saturation current); R is Richardson's constant, which is $4\pi m^* qk^2/h^3$ (where m^* is the majority-carrier effective mass); and A is the diode area.

In addition to the majority-carrier thermionic-emission current, the minority-carrier current flowing in the Schottky barrier is the same as that of a p-n junction. Consider a Schottky barrier on an n-type semiconductor (for a p-type semiconductor the equation is completely analogous); then the hole current is given by

$$I_p = I_{p0}[\exp(qV_F/kT) - 1] \tag{8}$$

$$I_{p0} = \left[\frac{AqD_pN_vN_c}{N_DL_p}\right]\exp[-E_g/kT] \tag{9}$$

where D_p and L_p are the diffusion constant and the diffusion length for holes, and N_v and N_c are the effective density of states of the valence and conduction bands, respectively. Since ϕ_B is always less than E_g for a covalent semiconductor, the thermionic-emission current is usually much larger than the minority-carrier current [compare Eqs. (7) and (9)]. For a numerical example consider an aluminum Schottky barrier on n-Si with $N_D = 10^{16}$ cm^{-3}, $\phi_{Bn} = 0.69$ eV, and $L_p = 10$ μm at room temperature:

$$I_{TE} = (3 \times 10^{-5} A)[\exp(qV_F/kT) - 1] \text{ amperes} \tag{10}$$

$$I_p = (3 \times 10^{-11} A)[\exp(qV_F/kT) - 1] \text{ amperes} \tag{11}$$

(Note that diode currents are insensitive to N_D if $N_D < 10^{17}$ cm^{-3} for n-Si.) We see that the majority-carrier (thermionic-emission) current is about six orders of magnitude greater than the minority-carrier current, which can be neglected in most cases for Schottky barriers. This has been confirmed experimentally by various measurements.[7,12]

This comparison brings out the most important differences between the *I-V* characteristics of Schottky barriers on covalent semiconductors and those of p-n junctions:

1. In Schottky barriers the current is carried by majority carriers, whereas in p-n junctions it is by minority carriers. Since minority-carrier storage limits the switching time of a p-n junction when it is switched from forward to reverse, Schottky barriers will have a negligible storage time, because the current is carried by majority carriers. Thus, Schottky barriers are ideal for high-frequency and switching applications.

2. For diodes of similar area, Schottky barriers have much larger currents at the same forward bias; in other words, the forward voltage drop will be much less across a Schottky barrier than across a p-n junction for the same current. This is clearly shown in Fig. 4, where the forward characteristic of a Schottky barrier is compared with that of a p-n junction. For a current of 1 μA the forward drop is 0.24 volt for the Schottky barrier, whereas it is more than 0.5 volt for the p-n junction even though the p-n junction diode area is almost 30 times that of the Schottky-barrier area. This low turn-on voltage feature makes Schottky barriers useful in various clamping applications. Because of the high thermionic-emission current, oxide-passivated Schottky barriers are relatively insensitive to ionizing and nuclear radiation.[13]

The forward-voltage temperature dependences of Schottky barriers and p-n junctions are different. For Schottky barriers on n-Si, solution of Eq. (6) for V_F yields

$$V_F = \frac{\phi_{Bn}}{q} + \frac{kT}{q}\ln\left(\frac{I_{TE}}{RAT^2}\right) \tag{12}$$

and

$$\frac{dV_F}{dT} = \frac{1}{q}\frac{d\phi_{Bn}}{dT} + \frac{k}{q}\ln\left(\frac{I_{TE}}{RAT^2}\right) - \frac{2k}{q} \tag{13}$$

For n-Si, assuming that $d\phi_{Bn}/dT = dE_g/dT \approx -0.2$ mV/°C, and $R = 120$ A/cm$^2 \cdot$°K^2, we have

$$\frac{dV_F}{dT} = -1.76 + 0.086\ln\left(\frac{I_{TE}}{A}\right) \text{ mV/°C} \tag{14}$$

A similar calculation for Si p-n junctions yields*

$$\frac{dV_F}{dT} = -2.0 + 0.086\ln\left(\frac{I_{TE}}{A}\right) \text{ mV/°C} \tag{15}$$

The two curves are plotted in Fig. 5. This difference in the temperature coefficient may be important in the design of circuits utilizing both types of diodes.

The reverse current, according to Eq. (6), should be a constant and equal to I_0. However, the top of the barrier [see Fig. 3(C)] will be slightly rounded off when electrons approach the barrier, since there will be induced charges of the opposite sign in the semiconductor. This is commonly called the image-force-lowering of the barrier, and the observed slight voltage-dependence of the reverse current agrees well with this model.[5-7] For high reverse bias, such that the maximum electric field causes impact ionization in the depletion region, avalanche breakdown will occur just as it does in p-n junctions. The breakdown voltage of both Schottky barriers and p-n junctions increases with semiconductor resistivity.

Metal–semiconductor contacts on heavily doped semiconductors. As pointed out before, when the semiconductor doping is low, the space-charge layer width is so wide that carriers can cross the barrier only

* $I_{p0} = (Aq/N_D\sqrt{\tau_p})\sqrt{D_p}N_cN_v\exp(-E_g/kT)$. Since $D_p = (kT/q)\mu_p \approx T^{-1.6}$, where μ_p is the hole mobility and τ_p is the hole lifetime, and $N_vN_c \approx T^3$, then I_{p0} is reduced to $I_{p0} \propto T^{2.2}\exp(-E_g/kT)$. It should be noted that this calculation is for diffusion current only. Recombination current in the space-charge region is not considered here.

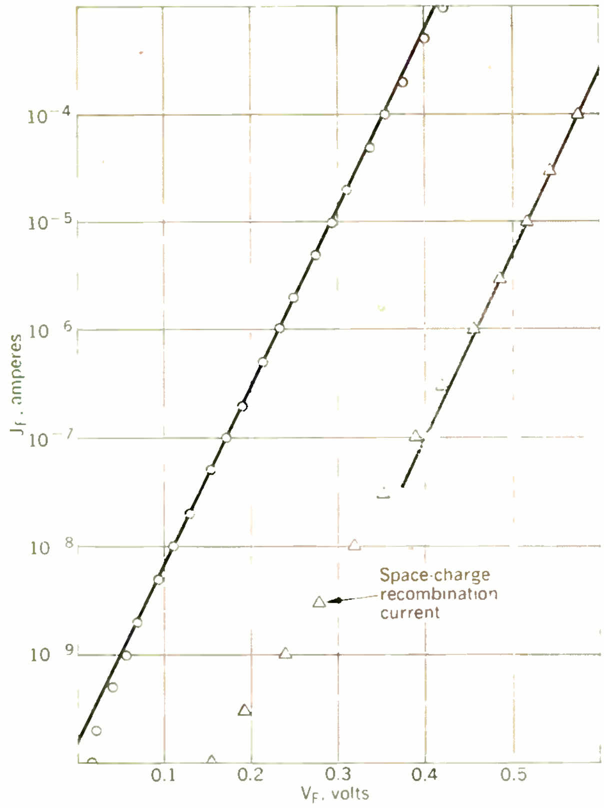
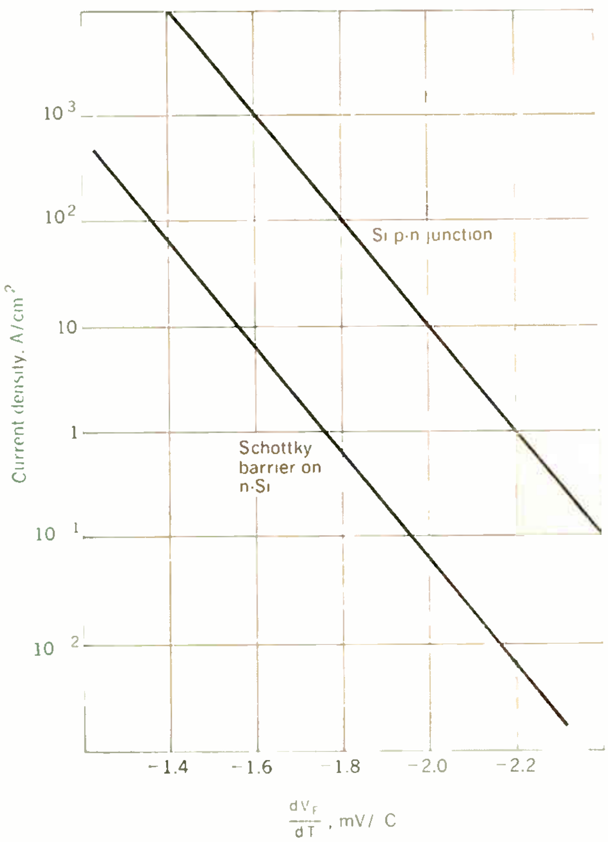


FIGURE 4. Forward current–voltage characteristics of an aluminum-n-Si Schottky barrier (circles): area of diode = 0.85×10^{-5} cm², $N_D = 10^{16}$ cm^{-3}; and that of a p$^+$-n junction (triangles): area of diode = 2.5×10^{-4} cm², $N_D = 10^{16}$ cm^{-3}. Straight lines represent $I_F \propto \exp(qV_F/kT)$.

FIGURE 5. Plot of current density vs. dV_F/dT for n-Si Schottky barrier and p-n junction.



by thermionic emission over the barrier. This situation is depicted in Fig. 6(A), where the band diagram of a Schottky barrier on an n-type semiconductor with forward bias V_F and the resultant *I-V* characteristics are shown. (Analogous figures can be constructed for a p-type semiconductor.) When the semiconductor doping is heavier (for n-Si, for example, $N_D > 5 \times 10^{17}$ cm^{-3}) the space-charge layer width is so thin—since $W \approx 1/\sqrt{N}$, from Eq. (3)—that carriers can "tunnel" through the middle part of this barrier, in addition to thermionic emission over the barrier; see Fig. 6(B). This added tunneling current of carriers with some thermal energy is commonly called thermionic field emission.[14]

If the semiconductor doping is very heavy (for n-Si, for example, $N_D > 10^{19}$ cm^{-3}), the depletion width is so thin that carriers at the Fermi level can tunnel through the barrier; see Fig. 6(C). This tunneling of carriers at the Fermi level is called field emission.[14] This additive component of current flow makes the current–voltage characteristics linear over a large current range, thus giving rise to ohmic behavior (aluminum on n-Si, for example). Since the tunneling process depends sensitively on the barrier thickness, it is expected that the resistance of an ohmic contact such as aluminum on n-Si should be a sensitive function of the doping of the semiconductor. Since the tunneling probability is proportional to exp $[1/W]$, the tunneling current is expected to be proportional to exp $[\sqrt{N}]$. Therefore, the zero-bias specific resistance (in ohm·cm²) should be

$$R_c \equiv A\left.\frac{dV}{dI}\right|_{V\to 0} \propto \exp\,[1/\sqrt{N}] \qquad (16)$$

Thus, a plot of ln R_c vs. $1/\sqrt{N}$ should be a straight line when tunneling current dominates[15]; it should become less dependent on the semiconductor doping when the doping is lighter, since dV/dI in Eq. (6) is independent of N. Comparisons of experimental data and the theoretical calculations of R_c vs. $1/\sqrt{N_D}$ are shown in Fig. 7, and the agreement is quite good.[16] The temperature dependence of R_c for $N_D > 10^{19}$ cm^{-3} has also been measured and it is found to be very slight,[16] as expected, for the tunneling current. Therefore, for ohmic contacts on n-Si, the tunneling model is able to account for all the experimental data.

To summarize, for metal on lightly doped semiconductors, the space-charge layer width is wide enough that the only carrier transport mechanism across the interface is thermionic emission over the potential barrier at the contact. This results in a rectifying characteristic, and such contacts are called Schottky barriers (for example, aluminum, gold, or platinum on n-type silicon with $N_D < 10^{17}$ cm^{-3}). However, not all Schottky barriers in this case are highly rectifying at room temperature. If the barrier is very low (<0.2 eV), the thermionic-emission current can be so large that the current–voltage characteristic may appear linear over a large current range (for example, for Pt Si on p-Si, $\phi_{Bp} \approx 0.2$ eV). But since this characteristic is extremely temperature-dependent, it can easily be distinguished from tunneling contacts.

For metal on heavily doped semiconductors, the depletion region is so thin that tunneling of carriers through the barrier is possible. This often results in a linear characteristic over a large current range when the semiconductor doping is very heavy, and such contacts are called ohmic contacts*—Al, Au, and Pt on n-type Si with $N_D > 10^{19}$ cm^{-3}, for example. Since the current flow is by tunneling in this case, the I-V characteristic is relatively temperature-insensitive.

Silicon Schottky-barrier technology

Three commonly used silicon Schottky-barrier structures are shown in Fig. 9. The n epitaxial film on an n$^+$ substrate is thermally oxidized, windows are opened by the standard photoresist technique, and the wafer is placed in a vacuum system for metal deposition. In order to achieve an intimate contact between the metal and the semiconductor, the silicon wafer is often chemically cleaned just prior to insertion into the vacuum system to remove any oxide and other residue. Nevertheless, a thin layer (~10 Å) of SiO_2 is usually unavoidable. This layer can be removed if the silicon surface is cleaned further in the vacuum system by inert-gas ion sputtering[5] or by high-temperature heating in ultrahigh vacuum.[17] However, these surface-cleaning procedures often are not feasible or compatible with present-day integrated-circuit technology. In the vacuum system, metal is then deposited by either evaporation or sputtering. Another photoresist step defines the metal geometry, such as that shown in Fig. 8(A). This particular structure, however, is not satisfactory for several reasons. Because of the positive fixed charge Q_s, at the Si–SiO_2 (or other dielectrics) interface, the depletion region of the Schottky barrier will be narrower near the periphery of the contact, leading to excess current at the corner. The resultant reverse characteristic is usually "soft," and the breakdown voltage is very low.[3] Furthermore, because of this excess current, the 1/f noise extends even to the megahertz range. Since these undesirable features are all attributable to corner effects, they can be eliminated if a

*It should be pointed out that not all ohmic contacts belong to this class. Contacts on semiconductors other than silicon have not been extensively investigated and thus are not well understood.

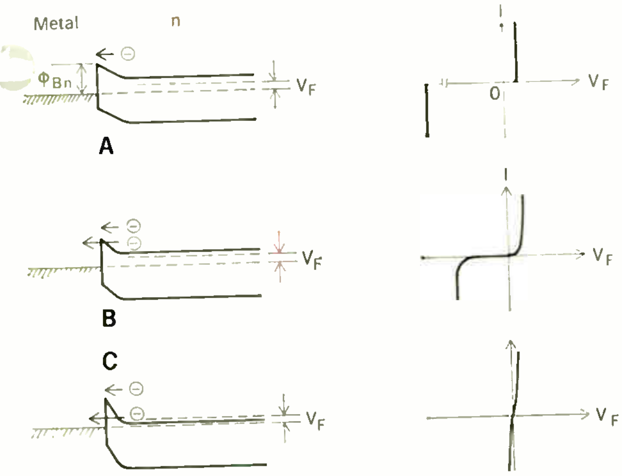


FIGURE 6. Band diagrams under forward bias and current-voltage characteristics of Schottky barriers on n-type semiconductor. A—Low doping, current by thermionic emission. B—High doping, current by thermionic field emission. C—Very high doping, current by field emission.

FIGURE 7. Plot of R_c vs. $1/\sqrt{N_D}$ for aluminum-n-Si contact at room temperature.

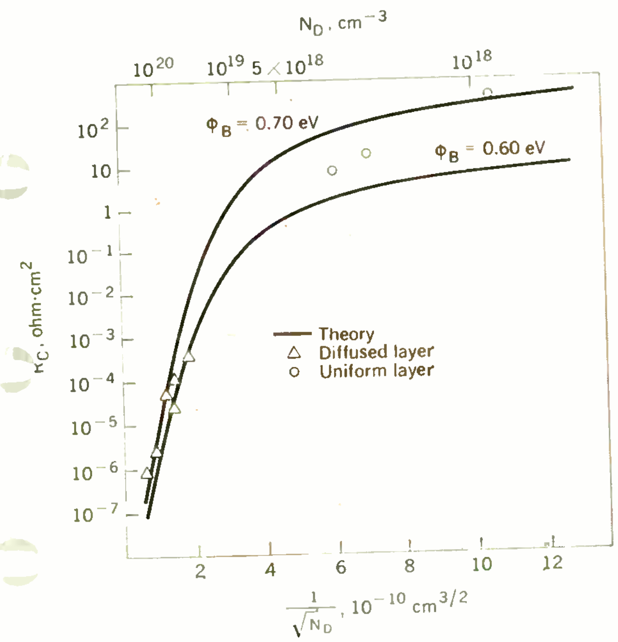
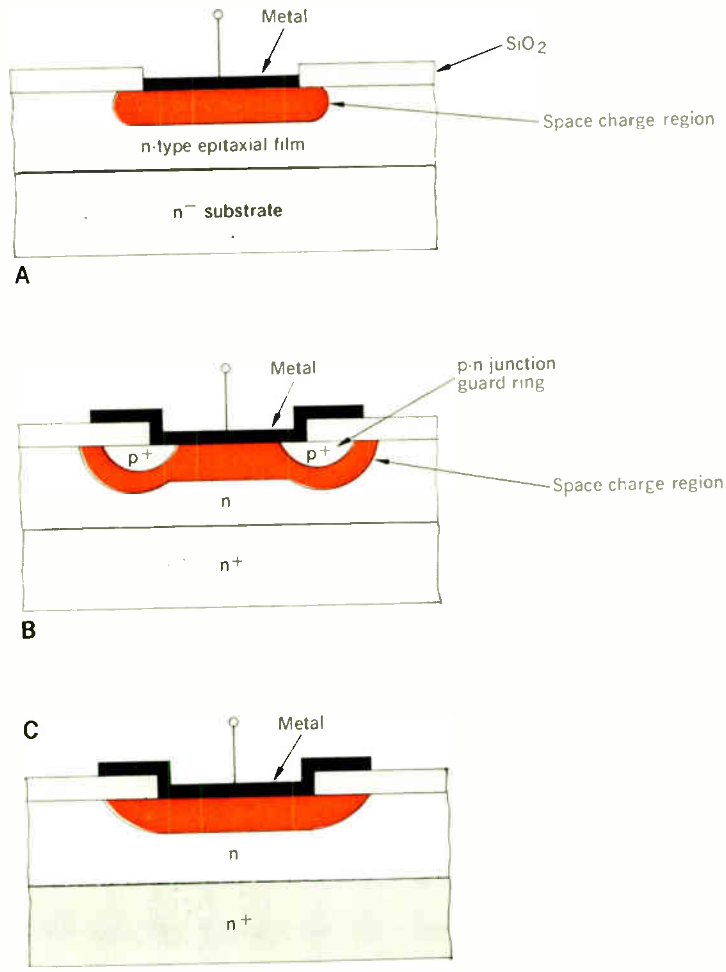
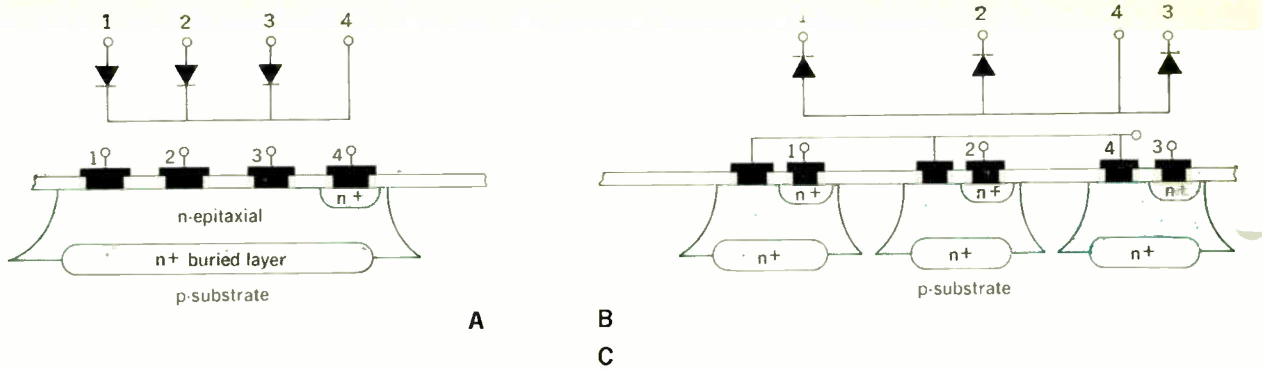


FIGURE 8. Three Schottky-barrier structures: (A) without metal overlap, (B) with p-n junction guard-ring structure, (C) with metal overlap.





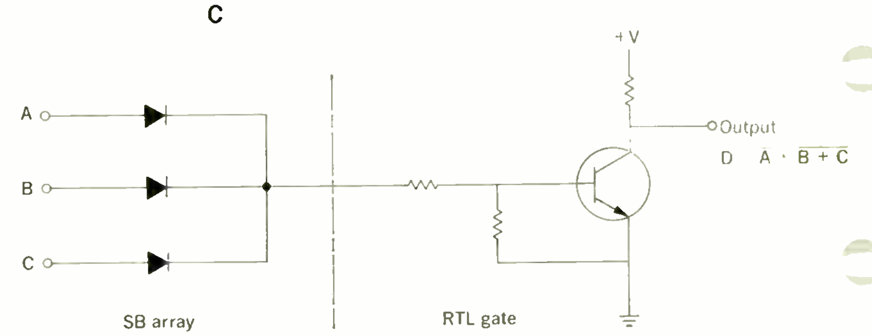
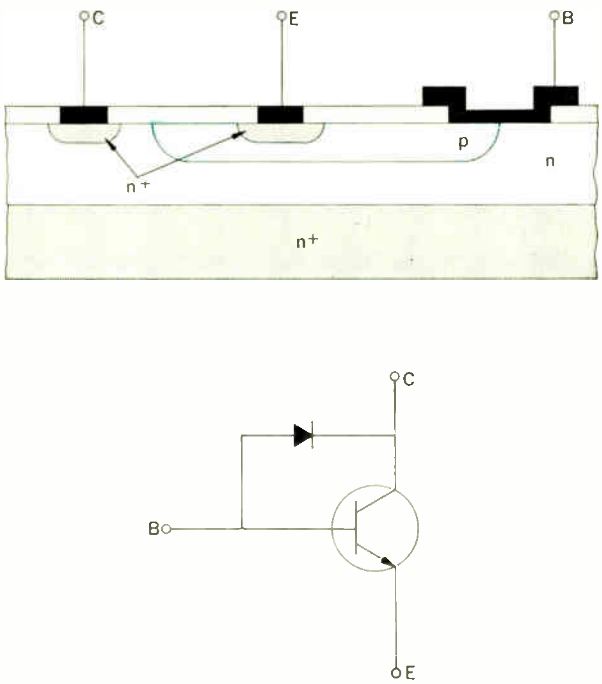


FIGURE 9. Gating application of Schottky barrier (SB). A—Common-cathode structure. B—Common-anode structure. C—SB's used as OR gate.

FIGURE 10. A Schottky barrier clamped n-p-n transistor.



p-n junction guard ring is employed, as shown in Fig. 8(B).[4-6] The doping profile of the guard ring can be arranged such that its breakdown voltage is higher than that of the plane Schottky barrier.[5] Therefore, this guard-ring structure exhibits near-ideal *I-V* and noise characteristics. Minority-carrier injection of the guard ring can be inhibited by choosing a metal that forms a reasonably high barrier on p-Si (such as molybdenum or titanium) and by reducing the surface concentration over the p^+ region sufficiently that a reverse-biased Schottky barrier will always be in series with the p-n junction guard ring.[6] Thus the guard ring will never be highly forward-biased and no minority carrier injection will result.

From the foregoing discussion, we can see that the technology of Schottky-barrier fabrication had advanced so considerably in recent years that Schottky barriers with excellent characteristics can be achieved by using the p-n junction guard-ring structure. Recently, a new technique of fabricating Schottky barriers on n-Si has been developed.[7] This method requires no special surface cleaning (such as ion sputtering), p-n junction guard ring, or rare metal. A simple structure, like that shown in Fig. 8(C), is sufficient. Aluminum, which is commonly used for making contacts and interconnections in integrated circuits, is employed. The aluminum is allowed to overlap the oxide so that the underlying surface will be kept in depletion when the diode is reverse-biased, thus eliminating soft breakdown caused by surface accumulation at the corner.[3] This overlap can be quite small (of the order of the space charge width under reverse bias), so that this parallel MOS capacitance will not be excessive.

The structure is heat-treated at 400°–550°C for a short time. This heating step allows the aluminum to reduce the unavoidable thin SiO_2 layer at the interface and make intimate contact to the silicon. The resultant Schottky barrier, though without a p-n junction guard ring, has an ideal *I-V* characteristic and very little low-frequency noise, and has been shown to be reliable and reproducible.[7] The processing involved is much simpler than that for a guard-ring structure [Fig. 8(B)]; and most important, since aluminum is ordinarily used in integrated circuits Schottky barriers over lightly doped regions and ohmic contacts over heavily doped regions can be made by one single step. Thus, the incorporation of aluminum Schottky barriers into integrated circuits imposes no extra processing.

Applications

Because of the absence of minority-carrier storage, Schottky-barrier diodes are ideal for microwave applications, such as mixers and detectors. The guard-ring structure, despite its excellent electrical characteristics at low frequencies, has excessive capacitance as a result of the parallel p-n junction. The aluminum Schottky barrier with oxide overlap, because of its simplicity in processing and excellent noise performance, has been found to be superior. Other applications in discrete devices—for example, as the gate in a field-effect transistor[18,19] and as an IMPATT diode,[20] have also been considered and demonstrated.

Probably the most important application of Schottky barriers will be in integrated circuits to perform various gating and clamping functions. In digital logic circuits, functions such as the logical AND and logical OR are normally performed by diode arrays. Most of the present-day integrated circuits make use of the simple structure of the multiple-emitter n-p-n transistor to form common-anode diode arrays. When using Schottky barriers, the common-cathode diode array is more desirable because it requires less silicon area than does the common-anode Schottky-barrier array [Fig. 9(A) and (B)]. The smaller area is important since it means a lower potential cost to the customer. The gating application shown in Fig. 9(C) illustrates that the Schottky-barrier diode array can be used as an OR gate at the input to an inverting resistor–transistor logic (RTL) gate to form a NOR function. The low forward voltage drop of the diode results in a reduced gate fanout, but with an increase in noise immunity.

In clamping applications, Schottky barriers can be connected in parallel to the collector–base junction of an n-p-n transistor to keep the transistor from going into heavy saturation and thereby reducing switching time.[21,22] The conventional approach to reduce switching time is by gold diffusion at the expense of low yield (caused by the nonuniformity of gold diffusivity) and low transistor gain (caused by the short minority-carrier lifetime in the base). If a Schottky-barrier clamped structure (as shown in Fig. 10) is employed, the gold diffusion can be completely eliminated. Therefore, one can achieve high yield and high transistor gain, in addition to short switching time, by utilizing clamping action of the Schottky barrier. Furthermore, since aluminum forms an excellent Schottky barrier over the lightly doped collector n region and at the same time forms a good ohmic contact over the heavily doped base p region, these two contacts can be made by one single metalization step (Fig. 10). No extra processing is needed. It should be noted that other metals, such as platinum and molybdenum, can be used instead of aluminum to form Schottky barriers and ohmic contacts to lightly doped and heavily doped regions of the semiconductor, respectively, if the silicon surface is sufficiently clean. However, the simplicity and the compatibility of the aluminum Schottky-barrier process with the present-day metalization scheme cannot be overemphasized. We have mentioned only a few of the more important applications of these devices. Many other applications are, of course, feasible.

The author wishes to thank R. A. Aldrich for his valuable assistance in the initial preparation stage of this article, and E. H. Snow and R. J. Whittier for their critical readings of the manuscript.

REFERENCES

1. Henisch, H. K., *Rectifying Semiconductor Contacts*. New York: Oxford, 1957.

2. Kahng, D., and Lepselter, M. P., "Planar epitaxial silicon Schottky barrier diodes," *Bell System Tech. J.*, vol. 44, p. 1525, 1965.

3. Yu, A. Y. C., and Snow, E. H., "Surface effects on metal–silicon contacts," *J. Appl. Phys.*, vol. 39, p. 3008, 1968.

4. Sah, C. T., U.S. Patent pending.

5. Lepselter, M. P., and Sze, S. M., "Silicon Schottky barrier diode with near ideal characteristics," *Bell System Tech. J.*, vol. 47, p. 195, 1968.

6. Zettler, R. A., and Cowley, A. M., "p-n junction Schottky-barrier hybrid diode," *IEEE Trans. Electron Devices*, vol. ED-16, pp. 58–63, Jan. 1969.

7. Yu, A. Y. C., and Mead, C. A., "Characteristics of aluminum–silicon Schottky-barrier diode," *Solid-State Electron.*, in press.

8. Kurtin, S., McGill, T. C., and Mead, C. A., "Fundamental transition in the electronic nature of solids," *Phys. Rev. Letters*, vol. 22, p. 1433, 1969.

9. Mead, C. A., "Metal–semiconductor surface barriers," *Solid-State Electron.*, vol. 9, p. 1023, 1966.

10. Crowell, C. R., Sarace, J. C., and Sze, S. M., "Tungsten–semiconductor Schottky barrier diodes," *Trans. AIME*, vol. 233, p. 478, 1965.

11. Archer, R. J., and Yep, T. O., "Control of thin-film interface barriers," Tech. Rept., Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Mar. 1969.

12. Yu, A. Y. C., and Snow, E. H., "Minority-carrier injection of metal–silicon contacts," *Solid-State Electron.*, vol. 12, p. 155, 1969.

13. Yu, A. Y. C., and Snow, E. H., "Radiation effects on silicon Schottky barriers," *IEEE Trans. Nuclear Science*, vol. NS-16, pp. 220–226, Dec. 1969.

14. Padovani, F. A., and Stratton, R., "Field and thermionic field emission in Schottky barriers," *Solid-State Electron.*, vol. 9, p. 695, 1966.

15. Mead, C. A., "Physics of interfaces," in *Ohmic Contacts to Semiconductors*. New York: The Electrochemical Society, 1969, p. 3.

16. Yu, A. Y. C., "Electron tunneling and contact resistance of metal–silicon contact barriers," *Solid-State Electron.*, in press.

17. Spicer, W. E., and Simon, R. E., "Photoemission studies of the band structure of silicon," *Phys. Rev. Letters*, vol. 9, p. 385, 1962.

18. Statz, H., and Munch, W. V., "Silicon and gallium arsenide field-effect transistors with Schottky barrier gate," *Solid-State Electron.*, vol. 12, p. 111, 1969.

19. Hower, P. L., Hooper, W. W., Tremere, D. A., Lehrer, W., and Bittmann, C. A., "The Schottky barrier gallium arsenide field effect transistor," *Gallium Arsenide—Proc. Second Internat'l Symp.*, Institute of Physics and Physical Society, London, 1969.

20. Reddi, V. G. K., "Avalanche transit-type oscillation in gallium arsenide Schottky-barrier diode," presented at the Solid State Device Research Conf., 1969.

21. Aldrich, R., "Low-storage Schottky-barrier diode transistor," presented at the IEEE Internat'l Electron Devices Meeting, Oct. 1968.

22. Price, J., "A high-speed integrated Schottky-diode transistor logic circuit," presented at the IEEE Internat'l Electron Devices Meeting, Oct. 1968.

Albert Y. C. Yu (M) received the B.S. degree in 1963 from the California Institute of Technology and the M.S. and Ph.D. degrees in 1964 and 1967, respectively, from Stanford University, all in electrical engineering. His doctoral dissertation involved research work on photoemission and optical studies of the electron structure of transition metals. He joined the Physics Department of Fairchild Semiconductor's Research and Development Laboratory, Palo Alto, Calif., in 1967. His chief interest has been in the physics and technology of metal–semiconductor barriers.

He has contributed to studies of surface effects, minority carriers, and radiation effects on silicon Schottky barriers as well as ohmic contacts to silicon. More recently his efforts have been concentrated on the technology of contacts to various compound semiconductors. Dr. Yu is a member of the American Physical Society and Sigma Xi.

New product applications

Phase limiter makes phase comparisons of two channels with unequal input signal levels

The ITLM Constant-Transmission Phase Limiter is an IF limiter that provides nearly constant transmission phase angle throughout its dynamic range. Typically, the total phase shift is less than 7.5 degrees over a 70-dB dynamic range. Each of the two models can handle either pulse or CW signals.

With the new units, phase comparison between two channels having unequal input signal levels can be made easily. When an ITLM is used in a system, amplitude differences, due to factors such as different antenna patterns, will not degrade the received phase information.

The new limiter permits the development of new kinds of monopulse and pulse-signature-identification receivers. Center frequency for the limiter is 30 or 60 MHz and 3-dB bandwidth is 10 or 20 MHz. Figures 1 through 4 show some typical applications.

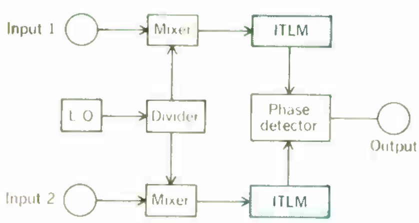


FIGURE 1. Phase measurement system.

Phase measurement

Figure 1 shows a phase measurement system where the two signals to be compared are connected to inputs 1 and 2. The signals are passed through mixers and converted to the center frequency of the ITLMs. The ITLMs amplify and limit these signals before feeding them to a phase detector. The relationship between the phase angles of the two signals is preserved in the conversion process. The response of the ITLM is limited only by its bandwidth and, therefore, it can be used for the phase measurement of pulse-modulated signals.

"Phase" monopulse

The block diagram of a phase-type monopulse system is shown in Fig. 2 For this processor it is assumed that the sum and difference signals are in quadrature. When the sum difference signals are added in quadrature in the combiners, the resultant signal has a phase angle with respect to the sum signal given by Eq. (1):

$$\tan \phi = \left|\frac{\Delta}{\Sigma}\right| \qquad (1)$$

In the combiners the amplitude information contained in the sum and difference signals is changed to phase information. This signal is then amplified and limited in the ITLM before being phase-detected. In this system, the signal levels are never more than 3 dB apart and, thereby, the phase mistrack error is reduced further.

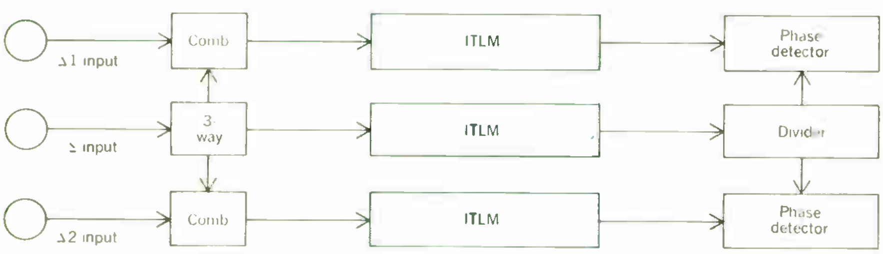


FIGURE 2. "Phase" monopulse IF processor.

Phase-tracked receiver

In Fig. 3 a block diagram of a phase-tracked receiver is shown. This system could also be considered to be a monopulse system. It reduces cost by eliminating the sum and difference hybrid and processing required with standard monopulse systems.

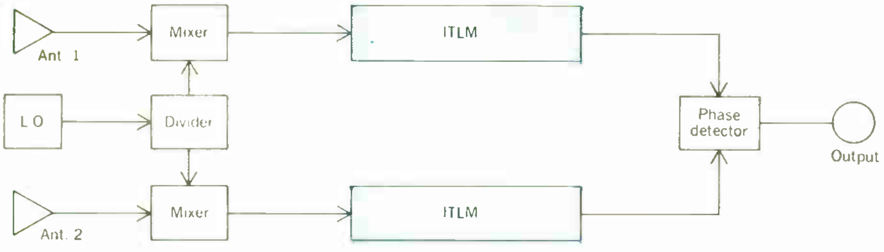


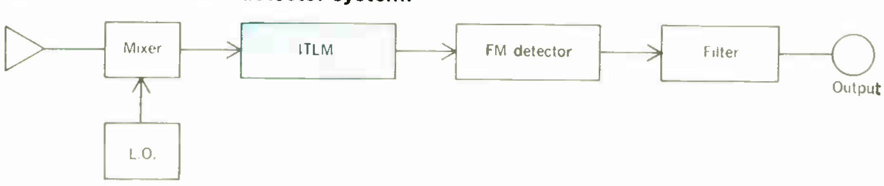
FIGURE 3. Phase-tracked receiver.

FM signal content detection

In Fig. 4, a block diagram of an FM signal content detector receiver is shown. A signal with no FM content will cause a constant output to be obtained from the FM detector. The FM content will cause an ac output. The amount of FM and its spectrum can be determined through the filters. The ITLM is essential to prevent extraneous signals from being generated due to variations in signal levels.

Figure 4, with the exception of the filter, is an FM receiving system with low AM-to-PM conversion.

FIGURE 4. FM content detector system.



More information is available from Varian LEL Operation, Solid State Division, Akron St., Copiague, N.Y. 11726.

Circle No. 85 on Reader Service Card