VOLUME 3 □ NUMBER 9 SEPTEMBER 1966

IEEE spectrum

features

Departments: *Please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

the cover

An anatomical study of the eye, by Leonardo da Vinci, serves to illustrate the concept of the eye–brain perceptual mechanism, as described in an article on motion perception by C. F. Hempstead, beginning on page 128 of this issue. The squares of color may have subtle significance to readers familiar with levels of signal processing in our sensory systems.

The following mythematical problem is based upon one of our operations analysis activities (with changes in the cast of characters). You might have a similar problem. The playwright of this scenario is Harold Levenstein of our Space Systems Department.

A Shepherd's Tale

After many years of faithful effort and increasing responsibility in the care of his master's flocks, a certain shepherd fell upon evil days. His breath grew shorter, his bones more brittle, and his myopia only partially correctable. Pride blocked admission of these difficulties to even a benevolent master, and he sought a solution disguised in other terms. Finally he came to the master and said, "I have a recommendation for our cost improvement program. Currently we take the sheep out into the open fields for food and exercise. We lose many because of haphazard protection in inclement weather, because of the inroads of predators, because we are only 99 percent effective in rounding them up. Further, their fleece is torn by the gorse and their meat is lean and stringy because they exercise excessively. To correct these ills I propose to build a large stockade as in Figure 1, showing a primary area where the majority of the sheep will be at any time, and a feeding area, the two being connected by one-way turnstiles to form a closed system.

"We'll operate them as follows: we'll admit ten sheep to the feeding area, and thereafter, when one wanders through he will first enter the weighing and shearing station, and then enter the primary stockade. This will also automatically enable the entrance turnstile. A sheep in the primary stockade will immediately enter to refill the feeding area, impelled by hunger and curiosity. I'll attend to keeping food available and inspecting the sheep, and Wolf, my faithful sheep dog, will keep them moving in the stockade as well as provide early warning if a lion or eagle should try to penetrate. I'll retain one staff man to tend the weighing and shearing station.

"Since we own the land, the initial cost is that of fencing it in. Although we will have to bring feed in, we shall be able to dispense with three of my helpers, and our final product will bring a better price.

"I have carried out a little experiment on the side, and I estimate that each sheep will pass through the feeding area in accordance with a Poisson probability distribution, showing a mean time in the feeding area of one hour. Since the area's capacity is ten, we can expect to service our one-hundred-sheep flock in a ten-hour day."

On reflection, particularly when the financing of the stockade was compared with the savings achieved by staff reduction, it appeared to be an attractive plan, and it was put into effect.

Things worked pretty much as estimated and the shepherd grew lazy, sitting in his elevated chair overseeing the feeding area and occasionally adding to the fodder. However, the lack of challenge caused the one remaining member of his staff to become careless, and to allow, on the average, one in every one hundred sheep to escape to the outside meadow where the prospect of greener pastures elsewhere caused the sheep to depart forever. (There are those who claim that this was not carelessness but rather a deliberate act practiced by the staff man because several of his buddies had been laid off, leading him to suspect that continued perfection of operation would inevitably lead to automation and termination of his employment too—such matters, however, are beyond the scope of this paper.) The shepherd did not realize this until the day a sheep exited from the feeding area and none entered. How long was it before the shepherd discovered the loss of the sheep?

Analysis

Strictly speaking, there is no answer to the problem as posed. However, if we change it to read, "How long do we expect it to take before the shepherd realizes that the sheep have vanished?"—then we can find an answer.

Let the number of the sheep in the stockade be denoted by S, the number that have passed through the exit turnstile of the feeding area by E, and the number of escapees D. These are functions of time.

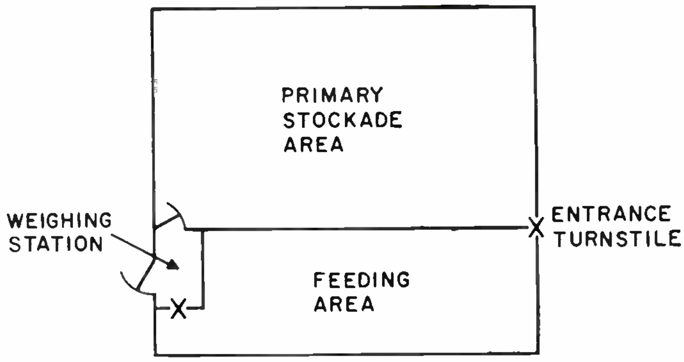


FIGURE 1

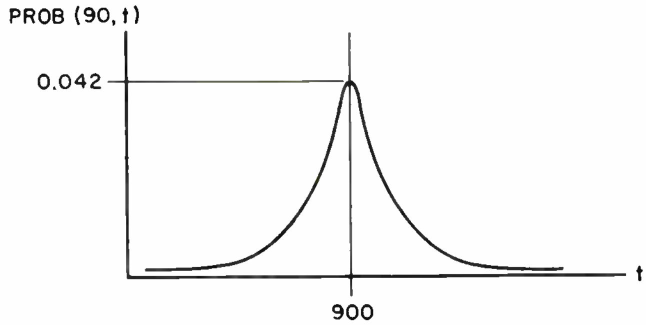


FIGURE 2

Then the performance of the system is such that we may write:

$$\text{prob}\,(E, t) = \frac{(m\lambda t)^E \exp\text{-}(m\lambda t)}{E!}$$

Thus, the probability that at any time t the number of sheep that have passed through the feeding area is E, is given by the above relation. λ^{-1} is the average time for one sheep to pass through (one hour by the shepherd's experiment), and m is the capacity of the feeding area, ten, as noted by the operating procedure.

Each of the sheep then makes a decision to escape (with probability q) or to return to the primary area (with probability $1-q$).

Thus, the probability distribution of sheep escaping is given as

$$\text{prob}\,(D, t) = \sum_{E=D}^{\infty} \text{prob}\,(D/E)\ \text{prob}\,(E, t) = \sum_{E=D}^{\infty} \frac{(m\lambda t)^E \exp - (m\lambda t)}{E!} \cdot \frac{E!}{D!\,(E-D)!}\, q^D (1-q)^{E-D}$$

Prob (D/E) means the probability of exactly D sheep escaping, given that E sheep appeared. For the binary decision that the sheep must make, this is just the binomial relation.

$$\text{prob}\,(D/E) = \frac{E!}{D!\,(E-D)}\, q^D\,(1-q)^{E-D}$$

With a little rearrangement we have:

$$\text{prob}\,(D, t) = \frac{(m\lambda t)^D \exp - (m\lambda t)}{D!}\, q^D \sum_{E=D}^{\infty} \frac{(m\lambda t)^{E-D}}{(E-D)!}\,(1-q)^{E-D}$$

The terms after the summation can be written:

$$\sum_{E=D}^{\infty} \frac{[(m\lambda t)\,(1-q)]^{E-D}}{(E-D)!}$$

which we should recognize as exp $[(m\lambda t)(1-q)]$, after some tinkering. Then the composite is:

$$\text{prob}\,(D, t) = \frac{\exp - (m\lambda q\ t)\,(m\lambda q\ t)^D}{D!}$$

If we set D equal to the initial number of sheep S_0, less the ten in the feeding area, we have the probability that the stockade is exactly empty written as a function of time and the characteristics of the feeding area.

Thus:

$$m = 10$$
$$\lambda = 1\ \text{hour}^{-1}$$
$$q = 0.01$$
$$m\lambda q = 0.1\ \text{hour}^{-1}$$
$$D = S_0 - m = 100 - 10 = 90\ \text{sheep}$$

$$\text{prob}\,(90, t) = \frac{(0.1\ t)^{90} \exp - (0.1\ t)}{90!}$$

This relation is sketched in Figure 2, as a function of t.

We note that the curve peaks at $t = 900$ hours, (or 37.5 days). A little differentiation indicates that more generally:

$$t_{max} = D/m\lambda q = 90/0.1 = 900$$

Thus, the most likely time at which the shepherd would discover his flock had vanished is 37.5 days. It is interesting to note that this fits our intuition very well: $m\lambda q$ is the mean rate of loss of sheep; most probable time corresponds to the expected time, and the probability of the discovery occurring exactly at that time is 4.2 percent.

IEEE forum

Correspondence relating to activities and policies of the IEEE

Basic research in the IEEE

When electrical engineering first split off from physics as an independent discipline nearly a century ago, the source of its strength lay in the reservoir of basic scientific knowledge that existed at the time. But the ensuing years have seen a deepening gulf form between the interests of modern physicists and the needs of modern engineers. Whether we like it or not, the responsibility for replenishing this reservoir has shifted largely from the physicist to the engineer. A tremendous opportunity now exists for the IEEE to share in this responsibility by consciously gearing itself to provide for the needs of a scientifically based engineering community. To meet this challenge I propose that we give serious consideration to strengthening the structure of the Institute on three specific points. All three have had an important part in the development of some of the world's most respected and successful scientific institutions.

The first point, which has been adopted by several of the leading scientific institutions, is the establishment of an honorary grade of membership whose qualification requirements are based solely upon scientific accomplishments of fundamental and enduring significance. The usual designation for this grade is Fellow. To be consistent it would be desirable that the same designation be used by the IEEE but the present grade of Fellow is now so deeply rooted that it may be impractical, and possibly not even in the best interests of the IEEE, to attempt to change it. Therefore, it is proposed that a new grade of membership be established to honor those who are making lasting contributions to the foundations on which our engineering profession is based. A possible designation might be Honorary Fellow. The new grade would be limited in number to a few tens, in contrast to the grade of Fellow, which is now numbered in thousands. And under no circumstances would it be conferred for other forms of service to the profession, as is the case for the Fellow grade, no matter how good and faithful that service may have been.

The second point, adopted by all of the leading scientific institutions, is the establishment of a journal devoted wholly to original scientific investigations on problems of fundamental importance. The *Physical Review* serves this function for the American Physical Society. It is proposed that the PROCEEDINGS OF THE IEEE be redirected to serve this function for the IEEE. Its purpose would be to stimulate interest in and to publish results of original research on the foundations of electrical and electronics engineering. The criteria for acceptability of papers would be twofold. First, the results presented should either open new areas of investigation by raising questions of fundamental importance or contribute new knowledge that is demonstrably necessary for the understanding of existing questions. Second, each investigation reported should be thorough and complete, with clear, definitive results. This is in direct opposition to the present trend of fast publication of partial results and loose conclusions, often based upon ill-conceived problems. It is anticipated that relatively few papers would be published in this journal but that a relatively high proportion of those published would be of such caliber that they would eventually become an indispensable part of the working literature. There must be no pressure to find material to fill space. It would be far better to have few papers, modeled more after the great classical memoirs of the distant past, than to compromise and find eventually that we have just another engineering journal.

The third and final point is to institute a reviewing procedure by which the high standards outlined above for the new PROCEEDINGS OF THE IEEE can be realized. Competent paper review has been and continues to be one of the most pressing problems of the IEEE. All who have ever experienced the frustration and sense of helplessness that comes after receiving an irrelevant review of a rejected paper are painfully aware of the shortcomings of our present system of anonymous reviewing. Relatively few reviewers seem to have sufficient depth of back-

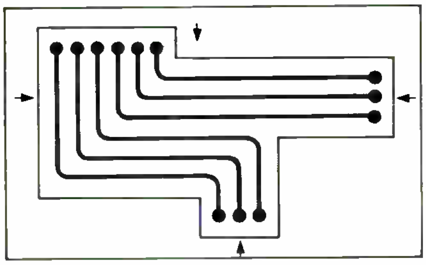
Schjeldahl | ELECTRONIC PRODUCT NEWS

Dimensional Accuracy in Schjel-Flex® Circuits

One in a series about flexible printed circuits for innovation-minded engineers. Interesting to you because you may soon use the materials, processes or types of circuits discussed. Interesting to us because they're our business.

Flexible printed circuits are being used in increasingly complex electronic systems. In the design of circuits for more sophisticated applications, dimensional accuracy has become more critical. Now, after research and refinement of materials and processes, flexible printed wiring can be produced to meet the dimensional tolerances of such applications.

Unlike rigid printed circuit boards, flexible printed circuits undergo dimensional changes during processing. By proper matching of materials and processes, shrink factors are controlled and compensation can be made for them in preparation of the art work.



Shrinkage

Two kinds of shrinkage occur in the processing of flexible printed wiring: etch shrink and thermal shrink. Both kinds of shrinkage vary with the substrates used, the way in which the substrate and conductor are laminated, the kind of circuit pattern involved and the circuit process employed.

Etch shrink results from stresses which develop in the flexible substrate at the time it is laminated to the conductor. When some of the copper is etched away during circuit processing, the exposed substrate returns to its original dimensions.

Thermal shrink is an inherent property of base materials which develops during manufacture. In the preparation of flexible printed circuits, it occurs when the laminate is exposed to heat during application of the coverlay.

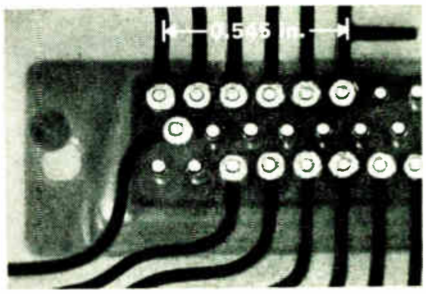
Shrink Control in Circuit Materials

Control of shrinkage begins with the manufacture of the flexible copper-clad. Schjel-Clad® materials are made with polyester, glass-epoxy type and polyimide (Kapton*) substrates. Each material is handled in a special way during lamination to minimize shrinkage during subsequent processing. By exercising very careful control of processing variables and using specially formulated and compounded Schjeldahl adhesives, we have been able to produce materials with reduced shrink factors. In some materials, shrinkage is as low as 0.02%. Using such Schjel-Clad materials to produce small circuits (up to 5-6") with reasonable tolerances, shrinkage can be overlooked.

Shrinkage Compensation

In larger circuits or those with more critical tolerances, shrinkage is considered in the preparation of the art work. Shrink factors for each Schjel-Clad material are known, and by careful quality control with frequent checks, they are maintained at consistent levels. Applying these known factors to a specific circuit permits the designer to prepare art work with dimensions calculated to compensate for shrinkage.

By applying the same careful control to circuit processing that is used in manufacture of the Schjel-Clad materials, we are able to produce circuits which meet dimensional tolerances as tight as ± 0.003-inch per inch.



Schjel-Flex circuits can be produced to meet the dimensional tolerances required for accurate registration.

Result: Circuits That Fit

Schjel-Clad materials are manufactured to maintain consistent and low shrink factors. Schjel-Flex® circuit processing is also carefully controlled to maintain consistent shrink factors. Therefore, when these consistent factors are properly designed into circuitry art work, the etched circuits register with connector pins or with other layers in multi-layer circuits, and match production tooling for accurate punching, drilling and blanking.

Because Schjeldahl is involved in both material production and circuit fabrication, we are in a unique position to help you obtain accuracy—whether you choose to make or buy your circuits.

If you're in the early stages of considering flexible printed circuitry, let us send some information about our materials and processing. Write G. T. Schjeldahl Company, Electrical Products Division, Northfield, Minnesota 55057. If you have a specific application in mind and need answers to technical questions, call us now at (507) 645-5633. Don't let pronunciation stop you. Say "Shell-Doll."

*DuPont Trademark

ground to be capable of giving a really penetrating review of the kind of unconventional thinking that is present in any truly fundamental paper. Furthermore, the system of anonymous reviewing tends to relieve the reviewer of any real sense of responsibility toward either the author or the reader. A solution of these two problems—the selection of a competent reviewer, and the placing of part of the responsibility for the content of a paper squarely upon that reviewer—becomes absolutely essential when dealing with ideas that depart appreciably from the norm. Therefore, it is proposed that all papers published in the new PROCEEDINGS OF THE IEEE be communicated through one of the Honorary Fellows and that his name be published along with the author's as the communicator of the paper. This is a practice of long standing with the Royal Society of London and should go a long way toward correcting the most flagrant problems of our present publication system. First, it would make the world's top engineering scientists available as reviewers. If one were approached who felt that he would prefer not to review any particular paper he would be under no obligation to do so, of course, but from my own experience in corresponding with such people I have found them to be extremely cooperative. Next, it permits the author to select the man most qualified to judge his work and to derive the fullest benefit of that man's criticism by a direct exchange of ideas. Finally, by bringing the reviewer out from under the cloak of anonymity he is forced to assume an attitude of responsibility toward both the author and the reading public. When he communicates a paper for publication he does so knowing that his own professional reputation is at stake. His acceptance of the paper would be final and would ensure its publication.

This is one answer to the challenge: to establish an honorary grade of membership for recognition of those who are developing the scientific foundations on which our profession rests, to establish a scientific engineering journal that is devoted wholly to the publication of original work on these foundations, and to institute an open review procedure requiring that all papers published in this journal be communicated through one of these honorary members. It is not an easy answer. But I am convinced that a positive program in support of basic scientific engineering knowledge is neces-

high-speed

Tektronix Oscilloscope with general-purpose convenience

Type 585A / 82 unit features

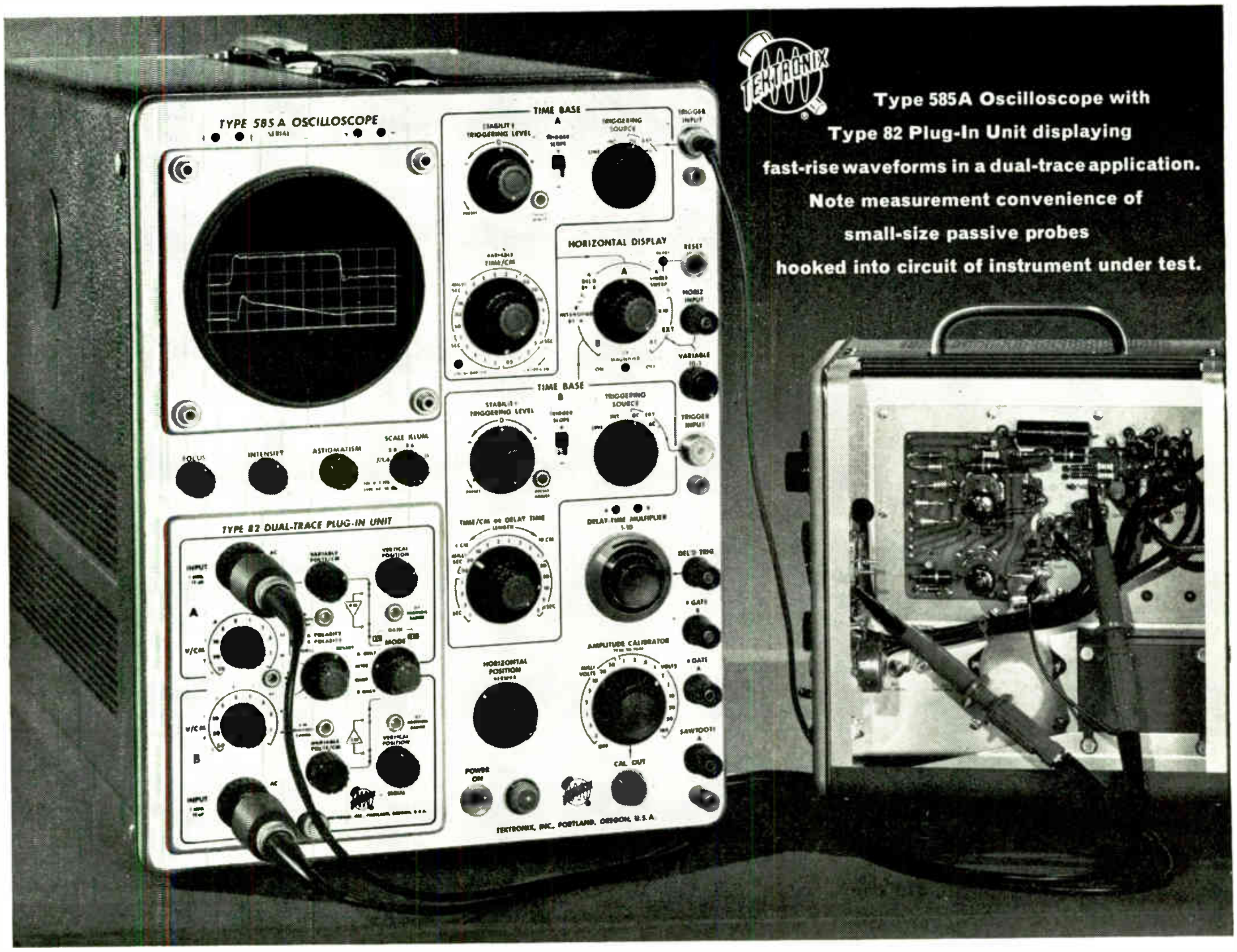
- **Dual-Trace Operation** with 4 operating modes and independent controls for each channel—for individual attenuation, positioning, inversion, and ac or dc coupling as desired.
- **Bandwidth DC-TO-85 MHz** (3-db down) at 100 mV/cm (12-db down at 150 MHz), and typically DC-TO-80 MHz (3-db down) at 10 mV/cm.
- **Calibrated Sensitivity in 9 steps** from 100 mV/cm to 50 V/cm, and in *10X Amplifier Mode*, from 10 mV/cm to 5 V/cm, variable between steps.
- **Internal and External Triggering** to 150 MHz.
- **Sweep Range from 10 nsec/cm** to 2 sec/cm.
- **Single-Sweep Photography** at 10 nsec/cm.
- **Calibrated Sweep Delay** from 2 microseconds to 10 seconds.
- **Bright, High-Resolution Display** with small spot size.
- **Conventional Passive Probes** for convenience.

plus

- **Compatibility with Letter and 1-Series Plug-Ins** to permit differential, multi-trace, sampling, spectrum analysis, other laboratory applications—when used with Type 81 adapter.

Type 585A Oscilloscope	$1725
Type RM585A Oscilloscope . . .	$1825
Type 581A Oscilloscope	$1425
No sweep-delay capabilities, but other features similar to Type 585A.	
Type 82 Dual-Trace Unit	$ 650
Type 86 Single-Trace Unit	$ 350
Type 81 Plug-In Adapter	$ 135
Allows insertion of 17 Tektronix letter-series plug-ins. Passband (up to 30 MHz) and Sensitivity depend upon plug-in used.	

U.S. Sales Prices, f.o.b. Beaverton, Oregon



Type 585A Oscilloscope with Type 82 Plug-In Unit displaying fast-rise waveforms in a dual-trace application. Note measurement convenience of small-size passive probes hooked into circuit of instrument under test.

Tektronix, Inc.

For complete information, contact your nearby Tektronix field engineer or write:
Tektronix, Inc., P.O. Box 500, Beaverton, Oregon 97005

sary if our profession is to realize its full potential.

Donald R. Rhodes
Raleigh, N.C.

Few will disagree with Mr. Rhodes' thesis that there is a deepening gulf between the interests of modern physicists and modern engineers. A gulf also exists between the interests of modern mathematicians and engineers. This situation has been developing over the past two decades, and engineers now generally accept the responsibility for developing the basic scientific and mathematical knowledge needed for the solution of engineering problems.

In deciding where the basic developments are to be described to the membership, we must keep in mind the development of the Group system within the IEEE. Each Group has served the professional interests of a segment of IEEE membership and has itself served to encourage publication of basic research results through its TRANSACTIONS. *The research-oriented* TRANSACTIONS *have attained worldwide reputations for excellence in their specialized fields. They would be seriously weakened by any plan that would deflect from the* TRANSACTIONS *to the* PROCEEDINGS OF THE IEEE *any and all papers judged to be fundamental or basic.*

The role that has developed for the PROCEEDINGS OF THE IEEE *is not unlike that envisioned by Mr. Rhodes. The emphasis is on new or emerging areas of interest to electrical and electronics engineers, areas not yet fully covered by the Groups. These areas are covered by special issues with guest editors of recognized reputation. Other areas are covered by review or synoptic papers written at a high level for research-oriented readers. A distinguished Editorial Board guides the* PROCEEDINGS OF THE IEEE *in the selection of areas of future importance to the IEEE.*

The reviewing system described by Mr. Rhodes has points in common with that presently used, with the Editorial Board for the PROCEEDINGS OF THE IEEE *playing a role similar to that he has described for the Honorary Fellows. Mr. Rhodes notes the frustration of an author upon receiving an irrelevant review. He should also note the joy to both author and editor that comes from a competent review by someone fully qualified for evaluation based on direct research experiences. Members of the Editorial Board are now called upon to select reviewers within the areas of their specialties. A small group cannot*

Up-to-date books that may be of importance to you professionally

10 DAYS FREE EXAMINATION—MAIL THE COUPON BELOW TODAY

1. AMPLIFIER HANDBOOK. By RICHARD F. SHEA. *New.* This giant volume covers virtually every phase of electronic amplification, with emphasis on practical application. The book's encyclopedic, nearly universal coverage of the subject ranges from direct current to gigacycles and angstroms, picowatts to megawatts, electrons to magnetrons. It discusses gases, solids and liquids; conductors, semiconductors, insulators, and dielectrics. It examines fundamental network theory and describes all major forms of amplifying devices, with a major portion of its pages devoted to amplifier circuits. Some up-to-the-minute topics include masers and lasers, tunnel diode amplifiers, logarithmic amplifiers, multipliers and dividers, microwave amplifiers, ceramics, and analog circuits. **410 illustrations, 1280 pp., $37.50.**

2. NOISE AND ITS EFFECTS ON COMMUNICATION. By NELSON M. BLACHMAN. *New.* This is a concise treatment of the fundamentals of random processes and their spectra, the effect of nonlinear transformation (such as demodulation) upon a signal and noise, the statistical theory of detection, and information theory. This comprehensive coverage, enhanced by some 200 problems, makes it valuable as a reference, workbook, or review. **224 pp., $13.50.**

3. ELECTROMECHANICAL POWER CONVERSION: Low Frequency, Low Velocity Conversion Processes. By ENRICO LEVI and MARVIN PANZER. In this single book, the reader will find, for the first time, a unified treatment of particle motion, plasmas and practical devices for electrical, mechanical, and aerospace engineering. Illustrations of many examples and problems clarify the basic principles and explain their application. **499 pp., $12.75.**

4. DIRECT ENERGY CONVERSION. By GEORGE W. SUTTON. This authoritative new text for graduate students and practicing engineers explains and correlates the fundamentals, potential, techniques and recent developments in solar photovoltaic cells, fuel cells, thermoelectric power generation, thermionic power generation, and MHD power generation. **336 pp., $14.50.**

5. MATHEMATICS OF PHYSICS AND MODERN ENGINEERING, 2nd Edition. By I. S. SOKOLNIKOFF and R. M. REDHEFFER. This extensively revised and updated volume is keyed specifically to the needs of scientists and engineers in today's rapidly changing technology. Contains expanded material on linear and nonlinear differential equations, the functions of several variables, and the coordinate-free formulation of the basic theorems of vector field theory. Over 300 problems and illustrative exercises. **768 pp., $14.75.**

6. FORTRAN IV PROGRAMMING. By ROBERT S. LEDLEY. This exceptional volume provides today's most comprehensive exposition of the language of Fortran IV. It clearly presents fundamental concepts as specialized applications in a complete series of selected programming examples. The best programming methods and problem formulation techniques are thoroughly covered in detail to provide clear, practical guidance in their actual use. Over one hundred illustrations. **229 pp., $4.95.**

7. BASIC TABLES IN ELECTRICAL ENGINEERING. By GRANINO A. KORN. Logarithm and trigonometric tables, physics data and design information, plus two easy-to-use tables of Laplace-transform pairs, appear in this unique book. Formulas, electronic-computer material and practical information useful for laboratory work are also included. The book comprises only that material directly and immediately useful in electrical engineering. **320 pp., $3.95.**

8. SEMICONDUCTOR ELECTRONICS. By JAMES F. GIBBONS. The author thoroughly examines the basics of transistor physics and circuit design in both their theoretical and experimental aspects. His book provides an invaluable foundation in modern circuit analysis, in which the design of amplifiers and the problems of switching circuits and devices are thoroughly discussed. **837 pp., $12.50.**

9. ELECTROMAGNETICS. By ROBERT S. ELLIOTT. A fundamental treatment of electromagnetic theory based on the *special theory of relativity*. The book begins with a thorough investigation and clarification of special relativity and then postulates Coulomb's law; it proceeds to derive all other segments of electromagnetic theory. An excellently written, clear exposition of concepts. Includes helpful historical introductions to most chapters. **544 pp., $15.50.**

10. SIGNAL DETECTION THEORY. By JOHN C. HANCOCK and PAUL A. WINTZ. This compact, detailed book collates recent important contributions in the area of signal detection. Centered around the problems of digital communications, it discusses the applications of communication theory with special reference to engineering problems. It is the first book to date to contain a section on sequential detection. **247 pp., $11.75.**

11. THE LASER. By PETER P. SOROKIN and WILLIAM V. SMITH. The first unified treatment of lasers to appear in book form. It examines the three major types of lasers — gas, optically pumped solid state, and semiconductor injection lasers — within the framework of the governing principle of stimulated emission. Much of the material is the direct result of original research by the authors and their colleagues. **544 pp., $15.50.**

12. BIOLOGICAL CONTROL SYSTEMS ANALYSIS. By JOHN H. MILSUM. This profusely illustrated book demonstrates that many biological systems can be mathematically modelled and analyzed through the use of techniques developed in control engineering. An interdisciplinary bridge between biology, medicine, and engineering, it presents stimulating case studies to motivate successive stages of dynamic analysis and control theory. **480 pp., $17.50.**

—10 DAYS FREE EXAMINATION—

McGRAW-HILL BOOK CO. Dept. 23-S-96
330 West 42 Street, New York, N.Y. 10036
Send me the book(s) encircled below for 10 days on approval. In 10 days I will remit for book(s) I keep, plus a few cents for delivery costs, and return unwanted book(s) postpaid.

1 2 3 4 5 6 7 8 9 10 11 12

Name (print) ____________________

Address ____________________

City ____________ State ________ Zip Code ________

For prices and terms outside U.S. write McGraw-Hill Int'l, N.Y.C. 23-S-96 *review all papers being written that might be basic contributions, but a knowledgeable group can pinpoint some of the best reviewers for each paper.*

I do thank Mr. Rhodes for his interest. I believe that all IEEE policy should be subject to continual criticism and evaluation in the spirit in which Mr. Rhodes' letter was written.

M. E. Van Valkenburg, Editor
PROCEEDINGS OF THE IEEE

Don't change IEEE spectrum

The IEEE SPECTRUM is the most read and most valued of the entire stable of IEEE publications, at least among my professional acquaintances.

The credit for this must go to your editorial staff and your authors, who seem well aware of the fact that the ability to communicate even abstract technical and academic concepts in plain rather than cryptographic text is a real contribution to technical growth.

The translation of a typical article from the PROCEEDINGS OF THE IEEE requires about as much labor as the translation of a page of ancient Egyptian cursive script, and in many cases one is left with some doubt as to what the author was really trying to say. Much of the material in the PROCEEDINGS OF THE IEEE becomes just so much background noise to readers who are faced with the common problem of absorbing a maximum amount of information in a minimum of time as a part of their continuing self-educational program.

My somewhat limited contact with many of the outstanding engineers and mathematicians of our time has shown me that they have one thing in common, the ability to present their ideas in readily understandable form, supplemented with the detailed material needed by the specialist in the field. One who can do this can be said truly to know the subject. Semantic complexity is often a refuge for the semi-knowledgeable pedant suffering from the "publish or perish" syndrome.

The continuance of your present style and information content will be a real contribution to the engineering profession.

Raymond W. Tackett
Paoli, Pa.

Give us more information on education

This is in response to your editorial in the May issue of IEEE SPECTRUM. I think this journal has done an excellent job since its inception. The balance and distribution of articles has been good; I hope it can continue at this high level.

I feel that this journal gives such a good view of the entire range of electrical engineering that it should be made a regular part of student memberships so that students would receive it on the same basis as the regular membership. Many times during the past few years articles have appeared that follow some course work my students have been doing. In most ways I think this is the best student-oriented publication the Institute publishes, although this is not its prime function.

From a professional point of view, I would like to see more attention given to trends and experimentation in engineering education. There should have been some discussion of the Goals Report along the lines of that which was carried out in *Mechanical Engineering* by ASME members. I think the general membership should be kept more aware of trends and thinking in education than they have been with IEEE SPECTRUM.

August E. Sapega
Hartford, Conn.

Six months prior to graduation, students begin receiving IEEE SPECTRUM *in addition to the* IEEE STUDENT JOURNAL. *Graduate students receive only* IEEE SPECTRUM.

Engineering obsolescence

In the July "IEEE forum," R. L. Nailen touched briefly upon a negative aspect of "obsolescence"; I would like to enlarge on his comment. Engineering obsolescence has become a warcry of management and is echoed through all of the major and minor professional journals. The profit squeeze has created an aberrated concept of the value of a good engineer. Novelty appears to be the keyword for evaluating the professional and its absence is being confused with obsolescence.

As long as the engineer was developing new concepts with old tools, the tools were incidental and ignored. When the market is saturated with a new concept, the company managers turn to new gimmicks, still ignoring the part played by basic tools. This pattern can be seen most acutely in the electronics industry in the switch from vacuum tubes to transistors and then to microcircuitry.

The stimulus of the dollar has fogged the vision of those who should know

Incomparable Performance at an Unbeatable Price ...the hp 427A Multi-Function Meter!

Time-saving features of the new Hewlett-Packard 427A Multi-Function Meter include: No need to change input connections to measure ac and dc volts—just change the function switch. Only one zero set for dc and resistance measurements . . . no need to re-zero when switching from dc to ohms measurements . . . seldom "zero" on the 1 v range and above. Instantaneous turn-on. Field-operate the compact hp 427A to typically 300 hours from one internal dry-cell battery (regulator included). AC-battery operation available at $35 additional. All Hewlett-Packard voltmeter accessories, such as dividers, adapters and leads, available for flexible performance.

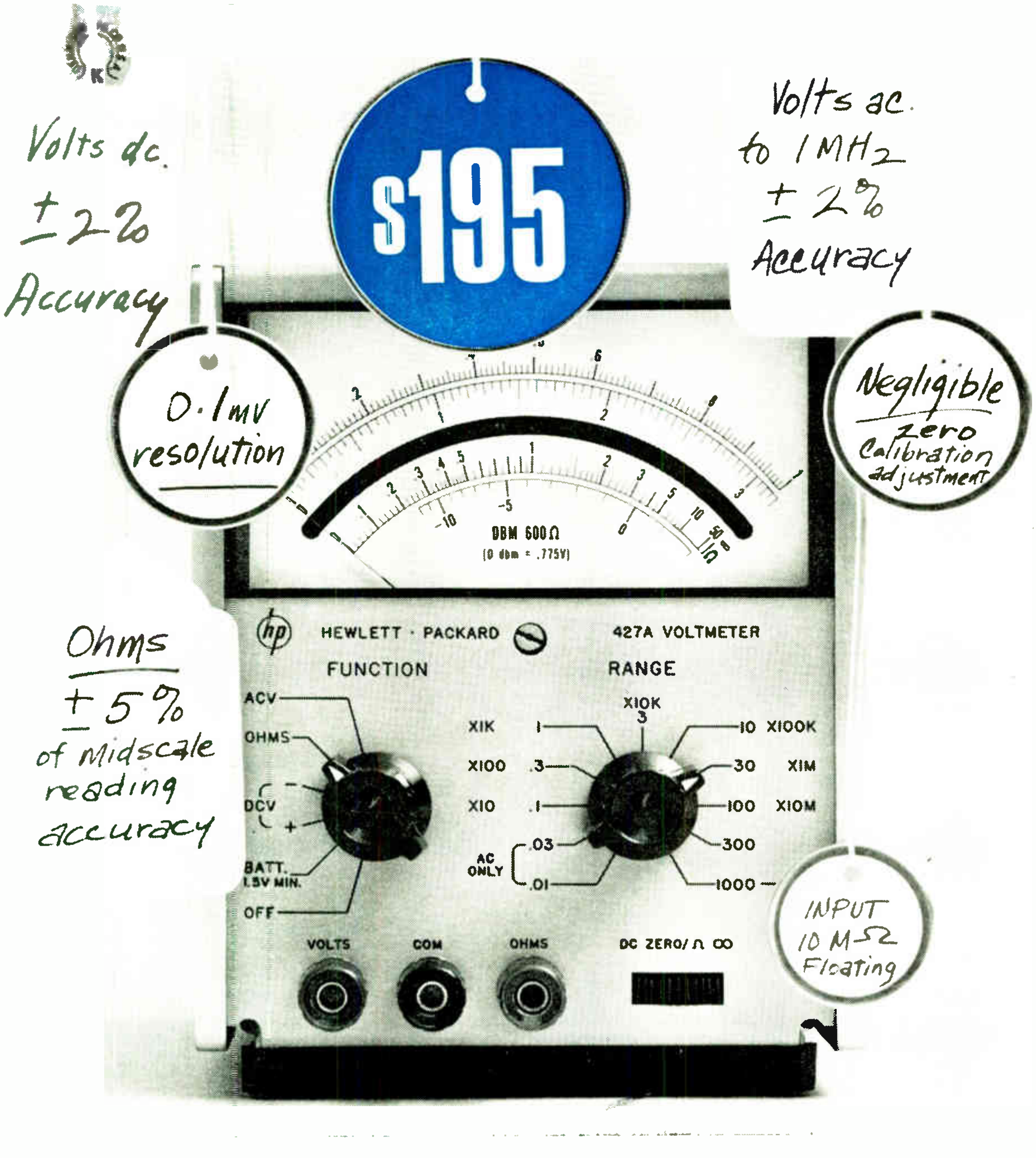
You can tell this is a measurement-packed instrument just by looking at the front panel. To see the 427A run through its paces, call your Hewlett-Packard field engineer. For complete performance details, write Hewlett-Packard, Palo Alto, California 94304, Tel. (415) 326-7000; Europe: 54 Route des Acacias, Geneva.

Data subject to change without notice. Prices f.o.b. factory.

HEWLETT hp PACKARD

An extra measure of quality

1481



better as to the value of educated people in their employ. When sales lag, the engineer is "obsolete," which is preferable to the more vulgar "overpaid." It is remarkable in an industry noted for its tremendous economic swings that the engineer himself can be swept up into the belief that he is truly obsolete. This phenomenon is especially acute in areas of narrow specialization.

Most of the muddled thinking today may be attributed to those who have the financial means to expound at length on principles that are distorted and biased by virtue of the position of men in management. Their interest, of course, lies primarily in those areas that can be equated to the dollar.

The great deception falls eventually on the lowly taxpayer who is told by management people that gimmicks must supplant all that has gone before—which is laughable to the engineer who coldly evaluates the pros and cons of all advances. As an example, the greatest curse of transistors is their temperature instability and finite interstage impedance, which cannot equal the advantages of vacuum tubes in this respect. And, of course, the microcircuit has absolutely no flexibility once released for production—which no engineer can ever believe is totally desirable (in addition to the drawbacks cited for transistors).

The obvious conclusion is that a cold evaluation of these new arts relegates new devices to *limited* areas where their worth admittably can be invaluable. An example is the extensive applications of these semiconductor devices to logic circuits where performance is generally limited to switch action.

In the face of lagging sales, management demands widespread applications of such devices regardless of limitations of the device itself. Hence, management and engineer are at odds, and up goes the warcry of obsolescence.

An example of management's studied ignorance of the value of basic concepts may be cited. An engineer was called into an organization to evaluate a production process utilizing basic but old techniques. His advice was to retain the old methods and base any future design on specification requirements no greater than the process required. Management's reaction was incredible, "This process must be antiquated; we have been doing it this way for years; a change on this basis alone is mandatory." From such reasoning, all automobile radiators should be eliminated since no change has been

made in principle since production of the Stutz Bearcat. Such is the reaction of many management people to the frenzied rush for gimmicks and innovations.

In conclusion, the swing of the pendulum is bound to return some day to a point where the professional will be evaluated by enlightened men on the basis of a sound and comprehensive grasp of the basic tools of the profession exclusive of his momentary economic value. These tools are never obsolete. May this day of enlightenment come soon.

John Gorman
Mount Holly, N.J.

"IEEE forum" is a welcome addition

Reading the "IEEE forum" in the July issue of IEEE SPECTRUM was a most refreshing experience. Not only do the members now have a way of expressing their opinions, but the Editorial Board actually responds and adopts changes in accordance with the members' preferences! This augurs a great future for communications between members and their society.

I would like to see more articles of historical interest in future editions. The feature on "World War II Radar: The Yellow-Green Eye" was most interesting. I don't expect to see this type of article in every issue, but something of the order of three to four articles a year seems reasonable. I also suggest a series of articles, not necessarily uninterrupted, on the great contributors to our profession. The article on Maxwell (IEEE SPECTRUM, Dec. 1964) is an example of the kind of feature to which I refer. I would also like to see more tutorial papers on various disciplines useful to the engineer, e.g., the articles on the electron by J. L. Salpeter (IEEE SPECTRUM, Mar., Apr. 1965) and on matrices by J. S. Frame (IEEE SPECTRUM, Mar., July 1964).

As for the advertising, I find ads about technical books very useful and only wish more publishers advertised their books in IEEE SPECTRUM.

My thanks and congratulations to past and present Editorial Boards for putting out such an excellent product. The new policies will undoubtedly maintain the present high standards and should also develop IEEE SPECTRUM's potential to become the finest magazine of its kind.

J. R. Martinez
White Sands, N. Mex.

Spectral lines

Communication of presence. "The telephone has failed again." The words gave me an odd feeling. Perhaps I shouldn't have said it, because the telephone call was of top quality, probably 50-dB SNR, 5-kHz bandwidth, loud and undistorted. I do not believe that 20 kHz and stereo would have changed the conclusion. Yet, the phone was not fulfilling its purpose.

After 120 years of electrical communication and more than 100 years of telephony, the typical conclusion of the typical brief telephone call was: "I'll come right over so that we can talk about it." The fellow I called would walk a quarter of a mile to discuss a trivial matter face to face rather than complete the business by telephone. There is nothing unusual in this, of course, but I wonder why one will go so far to achieve a face-to-face situation in preference to any presently available means of communication, even after the many decades and billions of dollars spent on the development of the telephone.

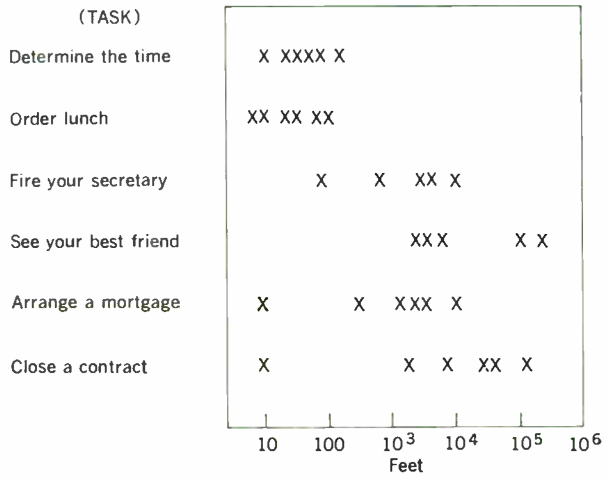
I asked some of my acquaintances how far they would travel to avoid having to use the telephone for a given task. To get a common measure, travel by foot was specified and rather ordinary tasks proposed. Some difficulty was found in relating inconvenience of the more difficult tasks to distance-of-foot travel, but such a relation was forced in order to find a common scale of inconvenience. The results are evident in the accompanying graph. The conclusions are:

1. As was surmised, people will go to considerable effort to achieve a face-to-face situation.
2. People do not like to walk. (On the more difficult tasks they admit they would go 100 to 1000 times as far by other means of transportation.)

The geometrical mean of the above data gives what I define as a "presence index" of about 1500 feet. A similar measure for messenger service is not much greater. How far would you walk to avoid having to use a (free) messenger?

The telephone has revolutionized our way of life and its value would be hard to measure, but clearly it is only a partial substitute for actual presence.

In the past few decades, much has been done to improve the cost, quality, reliability, speed, and universality of communication, but little has been accomplished that would affect the presence index. Will the addition of television make a difference? Very likely it will be some time before we know how much difference. Are there basic psychological barriers that will prevent much further reduction in the presence index? As an engineer, I would like to think that we can eventually come close to a complete illusion of presence. My instincts say otherwise. In any case, there is still a big challenge for the communication engineer.



Distances various persons would walk to avoid making a telephone call.

An article in the August issue of IEEE SPECTRUM suggested that eventually communication might be a substitute for travel. If so, we will have to solve the problem of communicating presence and beyond that we will still have some work to do. When we have perfected the mechanics of communications, what kind of channel will be required for communication with a 50-bit-per-second man? Will we need 120 megabits, as Goldmark suggested in the 50th anniversary issue of the PROCEEDINGS OF THE IRE? And that's not all. Deming Lewis, at the recent IEEE International Communications Conference, outlined a still more difficult problem: Even if we learn to communicate presence, we still have the problem of effectively and efficiently communicating meaning.

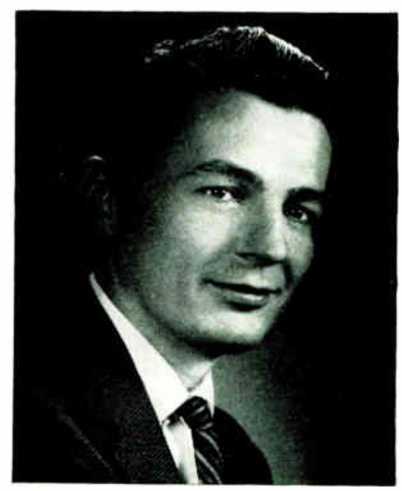
C. C. Cutler

Authors

Tropospheric scatter communications—Past, present, and future (page 79)

Frank A. Gunther (F), executive vice president of the Dynamics Corporation of America and president of its communications division, Radio Engineering Laboratories, completed engineering courses at Columbia University and Wagner College. He joined REL in 1925, advancing from the drafting room, through production and engineering, to marketing and upper management. From 1933 to 1947, as chief engineer, he was responsible for the development and manufacture of the first practicable two-way mobile radios for vehicles, a project that included field installation of many of the earliest mobile public service systems in cities throughout the United States. He also participated in the development and manufacture of the first commercially successful radio equipment employing frequency modulation, and he designed and supervised the installation of many of the first FM broadcasting stations in the United States. During the past 12 years, as vice president and later president of REL, he has been involved in the design, development, and production of tropospheric scatter equipment now in use in many parts of the world.

A dc transformer (page 117)



I. Giaever (M) received the Siv.Ing. degree in mechanical engineering from the Norwegian Institute of Technology in 1952 and the Ph.D. degree in physics from Rensselaer Polytechnic Institute in 1964. Before joining the General Electric Research Laboratory in 1958, he worked as a patent examiner in the Norwegian Patent Office and on the test program of the Canadian General Electric Company and the advanced engineering program of the General Electric Company in Schenectady. Since he became a staff member of the GE Research Laboratory, he has devoted a large part of his effort to studying the electronic properties of thin films. His work was concerned with tunneling through thin oxide film separating two normal metals, then tunneling from a normal metal to a superconductor, and finally tunneling between two superconductors. He was able to measure the energy gap in superconductors that was predicted by the Bardeen–Cooper–Schrieffer theory of superconductivity. In recognition of the work on tunneling, the American Physical Society awarded Dr. Giaever the Oliver E. Buckley Prize for Solid-State Physics in 1965.

A new theory of nerve conduction (page 123)

Ling Y. Wei (M) received the B.S. degree in electrical engineering from Northwestern College of Engineering (China) in 1942 and the M.S. degree in 1949 and Ph.D. degree in 1958, both from the University of Illinois. Early work included positions as radio engineer with the Chinese Air Force and telephone engineer with the Chinese Directorate of Telecommunications and service in various capacities with the Taiwan Telecommunication Administration. In 1958 he was appointed assistant professor of electrical engineering at the University of Washington, Seattle, where he performed semiconductor research and work on diffusion in 3–5 compounds. He joined the University of Waterloo, Ont., Canada, in 1960 and was named professor in 1963. At Waterloo, he initiated work on germanium–silicon heterojunctions using a surface-wetting technique. He has recently been engaged in theoretical studies of nerve conduction from the standpoint of solid-state physics and nonequilibrium thermodynamics, and is now working on a junction theory of nerve conduction and excitation and a theory of electron tunneling through a nerve membrane.

Motion perception using oscilloscope display (page 128)

C. F. Hempstead (M), who is a member of the technical staff of the Laboratory Measurements Department of Bell Telephone Laboratories, Murray Hill, N.J., received the B.S. degree in physics from Northwestern University, Evanston, Ill., in 1949 and the Ph.D. degree in physics from Cornell University, Ithaca, N.Y., in 1954. After receiving his doctorate, he entered the employ of the Bell Telephone Laboratories. His early work at the laboratories was concerned with backward-wave oscillators in the range of 30 to 120 GHz and, in addition, the use of these tubes for microwave spectroscopy of solids for maser applications.

In the early 1960s, Dr. Hempstead began to specialize in research work on type II superconductors and the application of these type II superconductors to the production of high magnetic fields. Then, in 1965, he was transferred to the Visual Systems Research Department of Bell Laboratories; here, his work focused on a study of the eye–brain perceptual mechanism. At the present time, Dr. Hempstead is employed in the Laboratory Measurements Department and is involved with work on high-precision automated microwave test equipment.

Experience with on-line monitoring in critical illness (page 136)

C. Vallbona received the B.A. and B.S. degrees in 1944 and the M.D. degree in 1950 from the University of Barcelona, Spain. After internship at the Hospital des Enfants Malades, Paris, France, and the University of Louisville School of Medicine, and residency at the University of Louisville School of Medicine and Baylor University College of Medicine, Houston, Tex., he became an instructor in the Department of Pediatrics of Baylor College of Medicine in May 1956. Later he took on additional work as an instructor in the Departments of Physiology and Rehabilitation. He was appointed assistant professor in the Departments of Pediatrics, Physiology, and Rehabilitation in 1958 and associate professor in 1962. In addition, he has served as program director of the General Clinical Research Center for Chronic Illness of the Texas Institute for Rehabilitation and Research since 1963.

W. A. Spencer, director of the Texas Institute for Rehabilitation and Research and professor and chairman of the Department of Rehabilitation, Baylor University College of Medicine, received the M.S. degree from Georgetown University in 1942 and the M.D. degree from Johns Hopkins in 1946. After interning at Johns Hopkins Hospital and taking his residency at Johns Hopkins and Brooke General Hospital, he served in the U.S. Army Marine Corps as assistant chief of medical service. From 1950 to 1959, he was medical director of the Southwestern Poliomyelitis Respiratory Center, Houston, Tex., and he became director of the Texas Institute for Rehabilitation and Research in 1959. In addition, he has been teaching at Baylor University College of Medicine since 1950, as instructor in the Department of Pediatrics, 1950–55; assistant professor in the Department of Pediatrics, 1955–57; and assistant professor in the Department of Physiology, 1954–57. He has been professor and chairman of the Department of Rehabilitation at Baylor since 1957.

L. A. Geddes (SM) has been professor of physiology at Baylor University College of Medicine since 1965, assistant professor of physiology at the University of Texas Dental College since 1958, and assistant professor of physiology at the Texas A.&M. College of Veterinary Medicine since 1965; he also serves as chief of the Section of Biomedical Engineering at Baylor and as consultant to the U.S. Air Force and NASA. He received the B.Eng. degree in 1945 and the M.Eng. degree in 1953 from McGill University and the Ph.D. degree in physiology from Baylor College of Medicine in 1958. He first became associated with Baylor in 1952 as a biophysicist. Subsequent positions included assistant professor of physiology at Baylor, 1956–61; director of the Laboratory of Biophysics of Texas Institute for Rehabilitation and Research, 1958–65; and associate professor of physiology at Baylor, 1961–65. At the present time, he also serves as consulting editor for the *American Journal of Medical Electronics* and for IEEE Transactions on Bio-Medical Engineering.

W. F. Blose of Baylor University College of Medicine and the University of Houston received the B.S. degree in mathematics from Oklahoma State University in 1958 and the M.A. degree in mathematics from Southern Illinois University in 1962. While attending Southern Illinois University, he served as manager of the Computing Division and, later, as research associate. He was involved in software development centered anound problems of implementation of algorithmic languages on small-scale computers. During this period, he presented several technical papers on algorithmic languages for the IBM 1620 computer.

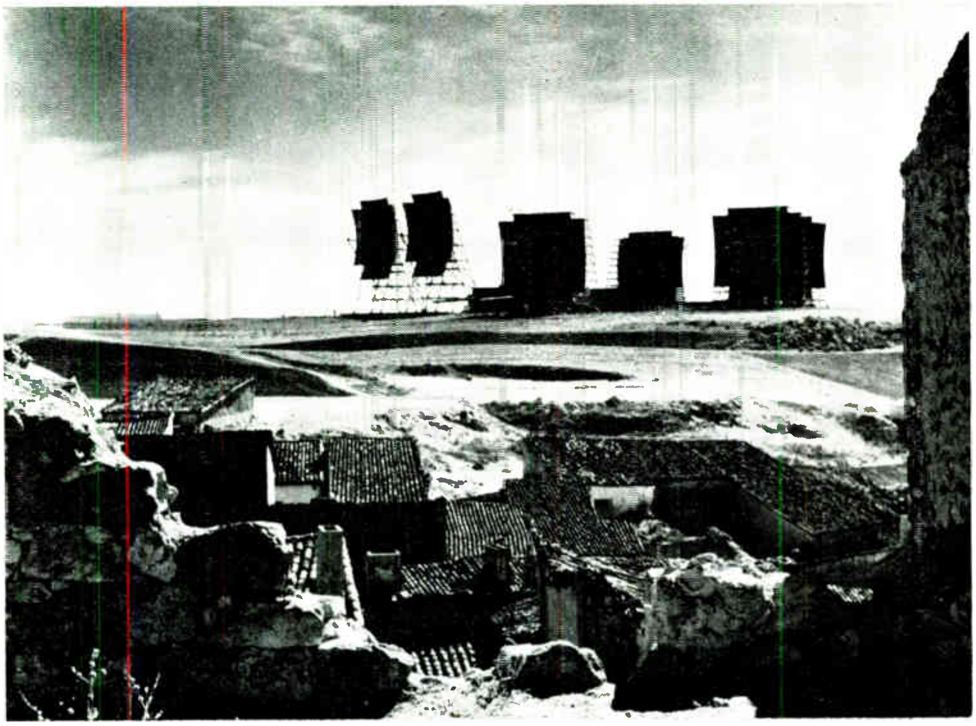
In 1963, Mr. Blose received a joint appointment to Baylor University College of Medicine and the University of Houston. At the present time, he is responsible for hospital data-processing support and is working on an information system supported by a grant from the National Institutes of Health. He has coauthored several technical papers dealing with data processing of medical research applications.

J. Canzoneri, III, received the B.S. degree from the University of Southwestern Louisiana in 1955 and the M.S. degree from the University of Houston in 1964. He joined Westinghouse Electric Corporation's Educational Center in 1955 as a graduate student engineer and, in 1958, became an associate engineer at Westinghouse's Air Arm Division in Baltimore, Md. During the period from 1959 to 1966, he served successively as senior electronics engineer with Dresser Electronics, development engineer with Schlumberger Well Surveying Corporation, group head with Lockheed Electronics Company, and engineering supervisor with Geo Space Corporation's Digital Data Acquisition Systems Department. He is now an instructor in the Department of Rehabilitation of Baylor University College of Medicine and director, biomedical engineering, at the General Clinical Research Center for Chronic Illness of the Texas Institute for Rehabilitation and Research. His major research interest has been the development of engineering systems for clinical investigation.

Reactive power control in a metro system (page 141)

W. J. Balet (M), assistant division engineer in Consolidated Edison Company's System Engineering Department, received the bachelor of electrical engineering degree from Cornell University in 1959 and subsequently attended the Polytechnic Institute of Brooklyn and the General Electric Power System Engineering course. He joined the Consolidated Edison Company of New York, Inc., in 1958 and, since then, he has held a variety of assignments in various fields of utility engineering. During the past five years, he has been mainly associated with system planning. Since 1964 he has been responsible for the studies made to determine optimum methods of reactive control and studies related to planning and operation of the interconnections with other utilities. After the Northeast power interruption of November 9, 1965, he was appointed to two Federal Power Commission study groups to investigate the interruption. He is presently an assistant division engineer in the Consolidated Edison Company's System Engineering Department and, in this position, he is responsible for the planning of the electric generation and transmission system.

R. L. Webb (F) received the B.S. degree in electrical engineering from Rice University in 1926. Following graduation, he joined the General Electric Company and served in the Test and Switchgear Engineering Departments. In 1929 he became associated with the Brooklyn Edison Company, a predecessor of Consolidated Edison Company of New York, Inc. He has been responsible for the development of a number of protective relays and various features in switching equipment, such as an ac generator loss of field relay, shaft vibration relay, three-phase voltage relay, double protective system for 345-kV circuits, and turbine–boiler–generator overall unit protective systems. He has been active in high-voltage switchgear standardization work for the past 25 years and was a member of the group that prepared the new symmetrical method of rating high-voltage power circuit breakers as covered by present ASA standards. In his present position as division engineer in the Consolidated Edison Electrical Engineering Department, he is responsible for the electrical systems protection as well as technical studies concerning station and substation design equipment ratings and station operation capabilities of machines and circuits.



Tropospheric scatter communications

Past, present, and future

An effective "gap filler," tropospheric scatter techniques provide a means for radio communications at distances lying between the short-range ultrahigh-frequency and superhigh-frequency systems and the long-range low-frequency systems

Frank A. Gunther *Radio Engineering Laboratories*

In the 11 years since the first practicable operational tropo systems were developed, there has been a tremendous growth in their number and complexity. This article describes the principles and applications of tropospheric scatter communications and discusses current development trends. Included is a table listing practically all of the tropo systems installed around the world, with pertinent technical data, together with maps showing the geographic location of most systems. The maps are keyed to the table for easy reference.

Tropospheric scatter communication may be defined as a method or system of transmitting, within the troposphere, microwaves in the UHF or SHF bands to effect radio communication between two points on the earth's surface separated by moderate distances of from 70 to 600 miles. Such a span or hop may be augmented by other spans in tandem to permit end-to-end or through circuits up to many thousands of miles. More specifically, this method of communication is now generally understood to embrace a radio system that permits communication over the distances indicated, with excellent reliability and good information capacity, using relatively high transmitted power, frequency modulation, and highly sensitive receiving apparatus. The name "tropospheric scatter communication" is now usually referred to in the engineering vocabulary simply as "tropo."

Figure 1 is a conventional sketch of a tropo span. This graphic representation is merely symbolic, and indicates in a general way that the circuit utilizes high power and large directional antennas, may be duplexed, can surmount ground obstacles of considerable magnitude, and will span a relatively large distance over the earth's surface.

(Above) Three-branch tropo relay station in Spain, forming part of a large U.S. Air Force defense communications system in Western Europe and the Mediterranean.

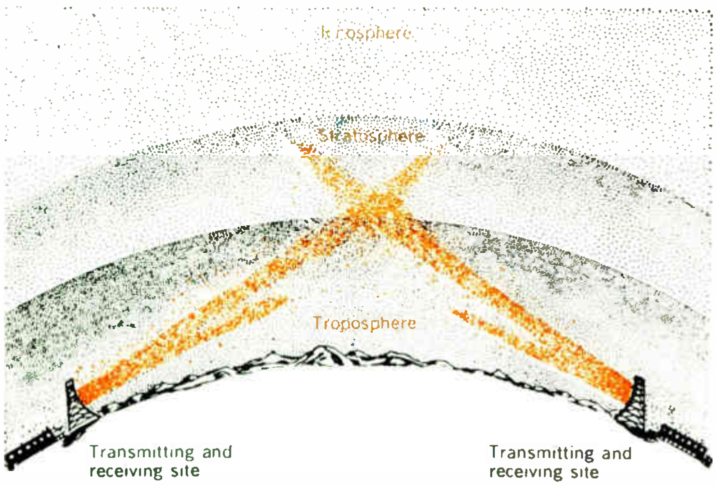


Fig. 1. Graphic representation of a tropo span.

With accuracy, tropo can be termed a "gap filler." It provides a means of radio communications at distances not covered either by the short-range UHF and SHF line-of-sight systems and the medium frequencies, or by the HF and long-range LF systems. Figure 2 summarizes the principal characteristics of "microwave" communications as utilized today over the surface of the earth, and indicates the relationship of tropo to the other modes in the microwave band (70 MHz to 20 GHz). Satellite communications are intentionally omitted from Fig. 2, as the propagation mechanism is somewhat different in space.

Systems of multiple spans in tandem, composed of tropo links and extending even to thousands of miles, are now in operation in many areas of the world, providing reliable multichannel communications. Tropo is also used in several special applications. The ability of tropo to span hundreds of miles of inhospitable terrain with circuits of relatively high traffic density insures a continuing need for this type of transmission.

History

It appears fitting to give a short account of the history of the early detection of the tropospheric scatter phenomenon and to describe the events that led to the first practicable operational systems in 1955.

Perhaps the first recorded investigation of tropospheric scatter was made in 1933 by Marconi,[1] who described 550-MHz tests conducted in 1932, first aboard the yacht *Elettra* and finally over a 168-mile path at 550 MHz from Rocca di Papa, near Rome, to Cape Figari, Sardinia. These results were the culmination of many earlier experiments, which had consistently demonstrated propagation 50 to 75 percent beyond the optical horizon at shorter ranges (15 to 50 miles).

André Clavier[2] reports on tests performed in 1941 with 10 watts at 3000 MHz (10 cm). The tests were conducted from Toulon, in southern France, to specially equipped ships sailing on a course to remain within the transmitted beam. These experiments duplicated those of Marconi in some respects, but updated information gained in Marconi's investigation, both in scientific refinement and in achievable frequency.

It should be noted that these programs were limited in range, by the low power available, to what is now known as the diffraction field; in a few cases of longer ranges, they were limited by anomalous propagation (possibly atmospheric ducting).

Tropo would probably still be a scientific speculation but for the tremendous impetus given to the development of high-powered microwave tubes during World War II. Radar, FM broadcasts, and television broadcasts all experienced reception beyond the distances that could be explained by the "smooth sphere" theories of propagation then current. In 1943 Terman[3] noted a probable tropospheric involvement in observations of propagation far beyond the horizon unreported in the unclassified literature of that time.

By 1949 a "freeze" had to be imposed on the issuance of licenses in the United States for new television stations because the cochannel interference caused by propagation beyond expected boundaries had proved to be much greater than anticipated.

The combined pressures of this nuisance factor and the growing need for reliable communications on the "Arctic perimeter" led, in the early 1950s, to the application of serious effort on the part of many individuals[4–6] and several organizations to unravel these mysteries. Experimental data were relatively slow in accumulating, because tests far beyond the horizon required high power, large antennas, and relatively long and expensive testing programs.

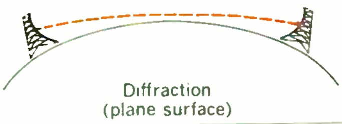
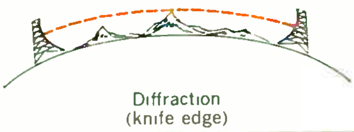
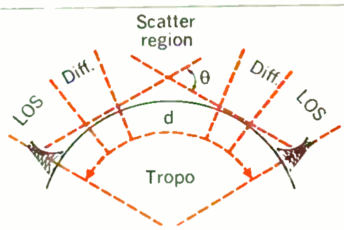
In 1952 Bell Telephone Laboratories, supported by much theoretical speculation and a convincing amount of empirical evidence, ventured to propose a practical beyond-the-horizon communications system, which operated entirely in the tropo mode and became known as the "Polevault" system. Armstrong,[7] in previously unreported work, verified independently the existence of usable fields in the 150- to 300-mile range from his laboratory at Alpine, N.J., at about this same time.

The explosion of interest created by the success of the Polevault system resulted in a special issue of the PROCEEDINGS OF THE IRE in October 1955, with contributions from major segments of the scientific community. It is to the credit of the U.S. Air Force that, based upon the evidence and theoretical considerations presented, it decided to take the risk and make the plunge into the new field of tropo by embarking on the Polevault system, rather than utilizing a line-of-sight system over a stretch of difficult and unfriendly terrain with the accompanying logistic problems produced by a much greater number of stations.

The tropospheric propagation mechanism

As a result of experience to date, tropo has emerged as a propagation mechanism that can endow communications systems with certain advantages not available through any other means, particularly where other types of systems are not practicable.

The phrase "tropospheric forward scatter" describes a hypothesis that has been suggested to explain the mechanism that enables the propagation of usable microwave radio signals well beyond line-of-sight distances. Diffraction theory does not account for the substantial electromagnetic fields that are produced at these ranges, nor does atmospheric ducting explain the time availability of these fields. One of the hypotheses attempts to explain

	Useful Propagation Range	Equipment Characteristics	Comments
	0 to 35 miles, depending on h	0.1 to 10 watts, 2- to 10-foot antennas	Low-cost, high-performance wide-band system; replaces costly right-of-way maintenance of coaxial or multiple cable or overhead wiring.
	30 to 70 miles, depending on h and N_s	0.1 to 100 watts, 6- to 28-foot antennas	Diffraction mode is very specialized form of UHF used only rarely where rugged terrain prevents use of direct LOS and permits longer path with obstacle gain.
	30 to 120 miles, depending on h, N_s, and G_o	0.1 to 100 watts, 6- to 28-foot antennas	Great attention is being given to refining propagational computation in the diffraction region because of need for utilization in tropo path predictions.
	30 to 120 miles, depending on h, N_s, G_o, and A_o	0.1 to 100 watts, 6- to 28-foot antennas	
	70 to 600 miles, depending on many factors	1 to 100 kW, 10- to 120-foot antennas, refined modulation and receiver techniques	Only practical wide-band, reliable ground-based method of achieving 70- to 600-mile hop where unsuitable intervening territory prevents use of LOS or diffraction modes.

h = height of antenna center
N_s = refractive index
G_o = obstacle gain
A_o = obstacle absorption
d = distance between stations
θ = scatter angle

Fig. 2. Principal ground-to-ground communication modes utilizing microwave (70 MHz to 20 GHz) region of radio spectrum; characteristically wide-band service (100 kHz to 20 MHz).

these apparent anomalies as resulting from atmospheric turbulence. This turbulence is thought to produce "blobs" of atmosphere whose refractive indexes are sharply different from those of the surrounding atmosphere. When irradiated by a microwave signal, these blobs reradiate the signal, scattering it in all directions. Some of this scattering is in the forward direction, which produces the field at a receiving location. The entire process is thought to occur in the region between the stratosphere and the surface of the earth—that is, the troposphere. This hypothesis forms the basis for the National Bureau of Standards methods[8] for predicting the strength and time availability of beyond-the-horizon fields. Although the hypothesis itself remains unproved, the method has been widely accepted; the hypothesis combined with massive empirical information ordinarily affords a usable solution to the problem of predicting performance.

Another promising hypothesis explains these fields as being the products of mode propagation based on partial reflections. The gravitational stratification of the atmosphere, where the lowest layer is the densest and each succeeding layer is less dense, with sharp boundaries between the layers, is thought to be responsible. When an electromagnetic wave is propagated through such an atmosphere, ray-bending occurs and is accompanied by partial reflections. This hypothesis has been advanced by Carroll and others.[9–12]

A third approach, long a subject of speculation by many, has been formalized through a derivation by Bullington.[13] This concept is based primarily on the average value and the standard deviation of the index of refraction at the earth's surface, both of which decrease exponentially with height. The result is a quantitative explanation of tropospheric radio propagation, derived without the use of arbitrary numerical factors.

Regardless of the true mechanism, much is known concerning the characteristics of the microwave energy field propagated beyond the horizon. It has been learned through observation of a large mass of empirical data collected from operational tropo links.

First, it is known that the average amplitude of the field propagated beyond the horizon is greatly attenuated with respect to the transmitted field. The amount of attenuation can be calculated as a function of the angular distance between the transmitting and receiving sites. Angular distance is a parameter that takes into account curvature of the earth, terrain configuration, and climatology. For all but the most unusual circumstances, it is closely related to the linear distance between the two sites. This relationship is the basis for the general statements that appear later regarding the traffic capacity vs. distance capabilities of the tropo scatter medium.

Second, the amplitude of the received field varies substantially with time over a given path. The National Bureau of Standards has gathered and analyzed a mass of data on these variations, so that it is now possible to predict the amplitude–time distributions for most paths with a high degree of accuracy. For convenience these amplitude variations are separated into short-term and long-term distributions. Short-term amplitude–time distributions are those measured over periods shorter than a few minutes. During such periods the amplitude variations are described by a Rayleigh distribution. The long-term distribution represents the variation of hourly median levels over a longer period of time—usually a month, a season, or a year. If amplitude variations are sampled over a period of one hour, an hourly median amplitude can be found, and if the hourly median amplitudes are observed over a longer period of time, the distribution becomes apparent. Since it has not yet been possible to derive exact theoretical expressions for these long-term distributions, empirical relations, based on the large mass of available data, are used.

It must be emphasized that the short-term and long-term distributions of signal amplitude vary considerably with the season of the year and with geographical location, and therefore when a system is being designed these factors must be taken into consideration. Determination of parameters and design are usually based on the worst propagational month of the year—usually February or March in the Northern Hemisphere.

Third, it is known that when a wide band of the frequency spectrum is utilized for the transmission of intelligence, any existing multipath propagation introduces a factor that is evidenced as a signal distortion. It is a result of the phase differentials that appear across the frequency band when signals are received over the longer paths of the multipath routes as well as over the direct path. This distortion places an upper limit on the bandwidth that can be transmitted successfully. To date, no generally accepted method of predicting the magnitude and distribution of the distortion has been formulated. Several researchers, including Medhurst,[14] Booker and de Bettencourt,[15] Chisholm *et al.*,[16] and von Baeyer,[17] have studied the problem. Several others[18–20] have published tentative methods for predicting its effects.

System considerations

All high-quality, long-distance, multichannel communication systems have two requirements: (1) reliability and (2) delivery of a favorable SNR (signal-to-noise ratio) to the customer. The signal components usually are amplified, as required, to overcome the various losses in the system. The big problem for the communicator is the accumulation of noise. Every time the signal is processed there is some contribution to noise, which adds on a cumulative basis. The result is that long-haul systems require sharp attention to the details of noise accumulation if the resulting SNR is to meet requirements.

Much thought has been given to this general subject and international objectives of performance have been formulated. This work has culminated in the CCITT standards, for telephone/telegraph, and the CCIR standards, for radio practice. In addition, the Defense Communications Agency of the U.S. Department of Defense has formulated specialized versions of these standards to apply to the Defense Communications System.

Signal-to-noise ratio is a complex subject, and a technical discussion would be out of place in this type of article. For a high toll quality on a telephone voice circuit in a multichannel system, the Defense Communications Agency prescribes that a minimum SNR of 45 dB be delivered to customers at the termini of 6000-mile systems, with the noise contributions from individual spans prorated on the basis of mileage as a percentage of 6000 miles. In commercial telephone systems one may occasionally find an SNR of 33 dB, which is quite usable; however, the ratio for toll quality is generally 39 dB or better.

Tropo scatter links are frequently installed in tandem to provide long-distance multichannel communication circuits. Theoretically, the relay stations involved could be termed "circuit drop-and-reinsert" stations, where all or a number of circuits must be used at the stations for local traffic requirements, or they could be called "through" repeaters, where there are no traffic or control requirements and the incoming signal could be heterodyned to the frequency of the outgoing signal without demodulation to baseband. In actual practice, regardless of traffic requirements, all relay stations bring the signals down to baseband, since it is necessary at least to utilize the order wire channels for circuit control purposes. Demodulation to baseband and subsequent remodulation for transmission to the next station in the chain may produce some intermodulation and multiplex noise to be added to the overall circuit noise. As a practical matter, however, equipment has long been available in which radio equipment intermodulation and multiplex noise have been reduced sufficiently that these factors in multiple span systems are no longer of major concern.

It was pointed out previously that tropospheric forward scatter signals are subject to amplitude–time variations, separated into short-term and long-term distributions. In practice, the two types of signal fading are not compensated individually, since the resultant fading is a combination of both types, usually referred to a monthly median level. Methods known as diversity techniques have been developed to counteract this fading and afford the prescribed propagational reliability.

Diversity can be defined as the utilization of more than one independent and uncorrelated transmission path over a single span of tropo to afford greater reliability than that provided by a single transmitter and receiver at each end. In each type of diversity now in common use, certain additional equipment is required.

One of the most effective methods is known as space diversity, in which two antennas, separated in space by 100 wavelengths or more, are used to receive the signals and reduce the effect of fading, since the signals separated in space by this distance show an almost complete absence of correlation. Another method in wide use is known as frequency diversity, wherein two frequencies, separated by about one to ten percent depending upon the frequency band in use, are transmitted over the span from the transmitting to the receiving station. Here, again, there is a minimum of correlation between the signals received on the two frequencies. As an example, a separation of 10 to 12 MHz at a frequency of 1000 MHz should be adequate. Other types of diversity, which are little used at the present time, include circular polarization and differences in azimuth orientation of antennas (angle diversity).

Although space and frequency diversity may be utilized independently (dual diversity), both may also be used simultaneously to afford quadruple diversity. For high-reliability circuit requirements and in applications where the span length otherwise might produce a marginal circuit, quadruple diversity is the preferred solution despite the additional equipment necessary. It is therefore used in the vast majority of fixed station systems installed to date.

To produce an optimum signal from a multiple receiver installation involved in either dual or quadruple diversity, a "combiner" is provided for adding and effectively utilizing the received signals. The combiner, into which all the receivers are fed, may be operated at baseband frequencies (postdetection combining) or at intermediate receiver frequencies (predetection combining). This subject will be discussed further in a later section.

Many variables affect any specific tropo path and, therefore, considerable engineering effort is usually expended to extract the most favorable compromises for a proposed system. Ultimately, system performance is traded for cost, within limits imposed by the path loss and supportable bandwidth of the propagation mechanism. Path loss can be directly attacked by higher power and greater antenna gains but there is a very real limitation on supportable bandwidth, at least as systems are designed today.

In many cases system design is not limited entirely by the cost or investment factor, since the importance of the system, its reliability, and its channel requirements may be overriding. Many special systems exist today in which the propagational reliability may vary from 99.9 to 99.99 percent for 120- to 300-voice-channel capacity over path lengths of 90 to 175 miles. There are also systems of 48 to 96 voice channels with the same reliability operating up to span lengths of 300 miles, and a few circuits that span 300 to almost 600 miles but are limited to 12 to 48 channels.

Tropo system parameters. The system parameters for tropo are the same as for any other type of communication system; that is, the power available at the receiver is equal to the transmitted power, plus all system gains, less all system losses. The system design engineer must consider all parameters (and there are many) in seeking the best solution for any system, and select the best possible values for these parameters within the physical or cost limitations imposed by the problem.

At this point, instead of advancing the mathematical presentation of the general system equation with appropriate detailed treatment of all parameters involved, the reader is referred to the National Bureau of Standards Technical Note 101, issued in 1965.[8] This thorough examination of all the parameters of a tropo circuit, based upon an extensive mass of data gathered over the years and embracing the frequency range of 40 to 10 000 MHz, offers the most complete analysis known on the subject. It is believed that the methods offered by Technical Note 101 for the solution of tropo system design and propagation problems are more widely used than any others.

In the planning of any tropo system much time can be spent in refining the values of the various controllable parameters to provide a maximum of power at the receiver with an acceptable customer SNR. Nevertheless, the overall path propagation loss is by far the most serious loss in the system, and exceeds by a wide margin the propagation losses in other types of radio systems.

The large path attenuation is overcome by transmitters capable of high power output, extremely directional antennas, and very sensitive receiving systems. The long-term amplitude distribution of received signals is counteracted in the same way. The short-term variations, on the other hand, usually are only partially offset in this manner, because in most cases total compensation for the wide excursions that characterize this distribution would require inordinate amounts of power. Diversity techniques, employing advanced signal combining schemes, are used instead. Although it is undoubtedly true that better understanding of the propagation mechanism would yield more advanced techniques and equipment, the tropo scatter mode can be utilized very effectively on the basis of what is already known.

Modulation and detection techniques. Important areas in which known techniques are readily usable in the tropo mode are modulation and detection. Frequency modulation (FM) is practically the only method used in the transmission of tropo signals today because of its superior performance when compared with single-sideband suppressed-carrier amplitude modulation (SS-BSC-AM, or simply SSB). It remains to be proved that various schemes of pulse code modulation (PCM) for some purposes may ultimately be better than FM.

The reasons for the recognized superiority of FM over SSB are frequently misunderstood, or overlooked. First, of course, is the fact that much higher SNR performance is available from an FM system operating well above threshold. The seldom-reproduced right-hand portion of Fig. 3, shows that the SNR values achievable at high signal strengths are far greater with an FM system than with an SSB system. In both cases the limiting factor is intermodulation noise caused by equipment nonlinearities.

The second feature favoring FM can be seen if we examine Fig. 4, which shows the distribution of energy about the carrier frequency. It should be noted that these occur in balanced pairs and that the process of detection produces a summation of all the individual components. Frequency-selective fading can alter the detected result only slightly, thus providing a sort of built-in frequency diversity action.

These features are obtained at the cost of increased spectrum space required, but are unobtainable by any other presently known means of achieving the extremely high performance required for toll quality of transmission. Current practice sets the FM index at about 3, requiring a bandwidth of about eight times the highest baseband frequency, although in most systems optimization of deviation usually takes place after installation to achieve best results.

In computing the SNR deliverable to a customer's handset, many factors must be considered, such as:

- Thermal noise
- Highest baseband frequency
- Peak per channel test tone deviation before pre-emphasis
- Channel top frequency
- Watts per cycle of bandwidth
- Receiver noise figure
- Pre-emphasis improvement
- Diversity improvement
- Multiplex noise
- Intermodulation distortion noise
- Transmission-line echo distortion noise
- Line loading factor

Antennas and transmission lines. Because of the high gain requirements encountered in tropo scatter transmission, highly directional antennas must be used. They usually take the form of large parabolic reflectors with horn feeds. The larger the parabolic surface, the higher the gain. However, a point of diminishing returns is reached. Above a certain gain, depending upon the particular radio path, the increase of gain with increasing parabolic diameter proceeds at a slower and slower rate until a practical limit on the maximum gain capability over that path is reached, primarily because of the medium-to-aperture coupling loss experienced in tropo scatter systems. This loss occurs only in the tropo scatter mode; in the line-of-sight applications, gain is not affected in this manner. The directly proportional relationship between gain and frequency limits the size of antenna that can be used for any particular frequency. Furthermore, since path attenuation increases with increasing frequency, the optimum frequency for a particular path is a compromise between path attenuation and net antenna gain. Another limiting factor to the size of antenna for a particular frequency is the surface tolerance that must be maintained in construction of the parabolic reflector. This limitation, of course, applies to the other modes of propagation as well.

The transmission lines used in tropo scatter work are usually the best available, since they represent only a small part of the total equipment cost and contribute heavily to the performance of the system. Large-diameter, rigid coaxial cable is normally used for transmitting power from the transmitter to antenna, particularly at powers of 10 kW and greater. It has high power-handling capability, low attenuation, and low-voltage standing-wave ratio (VSWR). In many instances, rectangular waveguide is used to provide even better performance in these respects. The transmission line from the antennas to receivers can be lighter, and is usually flexible coaxial cable with low loss at low VSWR. When transmitter output powers of one kilowatt or less are employed, flexible coaxial cable may also be used.

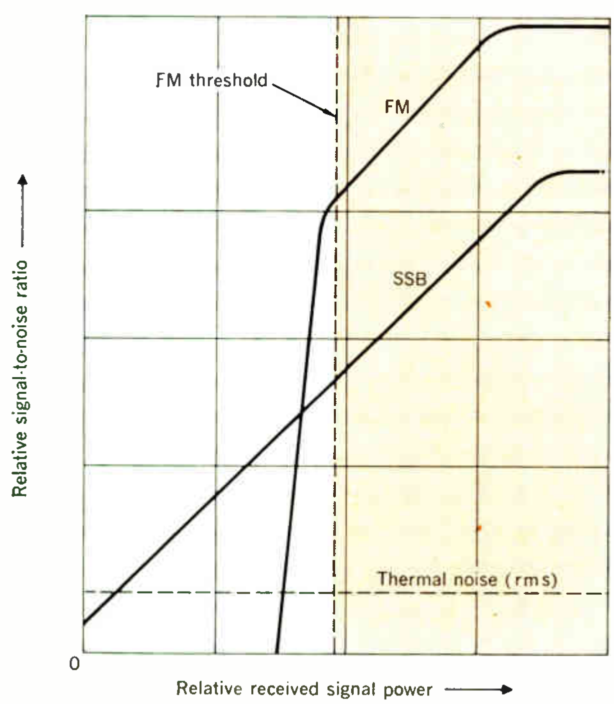
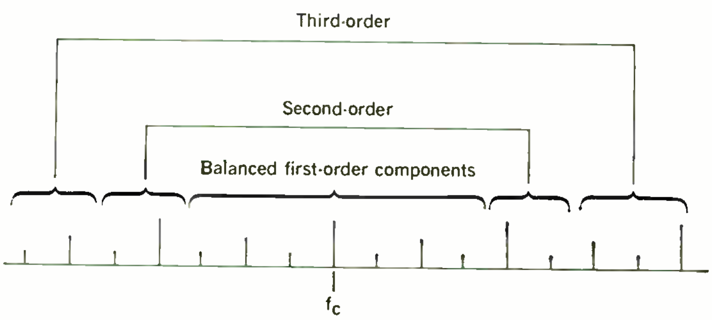


Fig. 3. Comparative performance of FM and SSB.

Fig. 4. Relative distribution of energy in FM signal.



System performance. In defining the performance of any communication system, two features are of paramount importance: the quality of the communication channels and the reliability with which this quality is achieved. In voice communications, channel quality is measured in terms of SNR; in teletypewriter and data channels, it is measured in terms of binary error rate; in both, the ultimate criterion is intelligibility. The two measures of quality may be combined under the single heading, grade of service. In statistical communications, which is what multichannel communications really are, there is no absolute level of service that can be considered to exist for all time. As a result, it is necessary to talk about the time availability for a particular grade of service.

When these definitions of system performance are applied to the tropo scatter situation, it is possible to separate grade of service and reliability on the basis of the two amplitude–time distributions of the received field. Since grade of service has been defined as voice-channel SNR or binary error rate, it can be related to the short-term distribution. Then, it follows that the long-term distribution establishes the time availability, or reliability, of the grade of service resulting from the short-term distribution.

Reliability has another facet, too. It is the reliability of equipment, expressed in terms of mean time between failure (MTBF). Although much complicated circuitry is involved in tropo scatter equipment, the state of the art with respect to components and circuit design permits very good MTBF values, certainly as high as any that can be realized in other communications media. And new, transistorized equipment provides even better MTBF values.

Like all communications media, tropo scatter has certain limitations, which in this case restrict the number of channels that can be transmitted and the distances that can be spanned. The two most stringent limitations encountered in tropo scatter transmission are the large path attenuation and the bandwidth of the medium. The first, as we have seen, is related to angular distance between the transmitting and receiving antennas. Even though it is mitigated by the use of high radiated power, antenna directivity, and sensitive receiving systems, there is a range beyond which the combination of the utmost capabilities of these devices is not adequate to provide a signal level above the system threshold. It is perhaps too obvious to mention that the traffic capacity of a system is maximum at the minimum range and decreases to minimum capacity at the maximum range. The second major limitation, the bandwidth of the medium, is not well understood. It is thought that the bandwidth restrictions that have been observed over tropo scatter paths are a result of frequency selective fading caused by multipath propagation of the signal. This theory is partially borne out by the ability of highly directive antenna beams to reduce the multipath effects—that is, to increase the phase-coherent bandwidth. It has also been observed that multipath effects increase with increasing path length and, further, that they are inversely proportional to the RF carrier. In any event, the net effect is to limit the RF bandwidth that can be successfully transmitted and, consequently, to limit the traffic capacity of the system.

A less serious limitation, though one still worthy of consideration, results from choice of a less than optimum frequency. It has already been pointed out that optimum frequency is the result of a compromise between path loss and realized antenna gain. But in most cases another factor, spectrum availability, is involved in frequency selection. Often a frequency that is not optimum for a particular path must be used, with a consequent reduction in received signal strength. This situation, of course, limits the amount of traffic that can be transmitted with a specified performance.

In addition to these technical limitations, cost and size represent two other factors that limit system capability. The cost factor needs little explanation. If a system with a certain capacity is considered too expensive, one of reduced capacity or less reliability must be considered. Physical size of equipment is usually only a limitation when it conflicts with the intent of the system. It rarely ever imposes any restrictions on a fixed-plant installation, but if an installation must be made transportable, or at least mobile, then very definite limitations are imposed. These usually dictate the maximum transmitter power output that will be available, and the maximum antenna sizes and order of diversity that can be used. And when really tactical transportability is required, such as a helicopter lift requirement, size and weight are the most important of all limitations.

To reiterate, the factors that limit tropo scatter system performance are path attenuation, bandwidth, frequency allocation, number of tandem links, physical size, weight, and cost. Nevertheless, the effects of these factors are offset to a remarkable degree by the skillful utilization, with necessary compromises, of engineering techniques and advanced equipment designs, including the use of FM, low-noise receivers, diversity reception with maximal-ratio combining, threshold extension techniques, high-power transmitters, diversity transmission, high-gain antennas, low-loss transmission lines, and appropriate siting of stations. These enable the system engineer to exploit the tropo scatter medium, imperfectly understood though it may be.

Typical tropo equipment

The development and subsequent experience with tropo during the past 11 years has tended to "standardize" to a great extent the equipment used today on these circuits, regardless of point of manufacture. Frequency modulation is used almost exclusively on tropo circuits because of its inherent capability to produce a greater SNR. There are very few proponents of amplitude modulation. At this time, it appears unlikely that AM will receive further attention in the foreseeable future.

Almost all of the tropo circuits utilize an output power of one kilowatt or greater, and in practically every case a klystron power tube is used. With the continual development of more efficient and larger klystrons, output powers up to 100 kW are now being used. With the advent of high-power traveling-wave tubes additional flexibility in equipment design may be forthcoming.

Diversity equipment configurations

In dual diversity, either space or frequency diversity is usually employed. Referring to Fig. 5, it should be noted that for dual space diversity, in theory only one transmitter need be provided per terminal. However, in the event the transmitter goes down, communication is lost. It may, therefore, be practical to provide a second transmitter, maintained in a standby condition to take over immediately and automatically in the event the other transmitter fails. Thus, for this type of dual diversity, there would be available both the two regular receivers with combiners and two transmitters to insure continuous communication, even though one receiver or one transmitter is down and the resultant circuit reliability is somewhat less than that designed. This arrangement also permits regular preventive maintenance on receivers or transmitters without complete loss of communication. In the dual space diversity arrangement (Fig. 5), the advantage of frequency diversity is not available; but space diversity is more conservative of the frequency spectrum.

In the case of dual frequency diversity (Fig. 6), the desired redundancy of equipment is provided, even though the advantage of space diversity is gone, since only one antenna is required. Although there appears to be a drift toward the use of dual space diversity rather than dual frequency diversity, there are occasions where the latter is the only practicable solution at the present state of the art, as in the case of a ground site or shipboard site that does not permit the employment of two antennas spaced at a distance of at least 100 wavelengths apart, or where cost factors are overriding.

Figure 7 shows a typical configuration of a quadruple diversity terminal. Obviously, this type of diversity offers the maximum in equipment redundance and system performance, assuming proper design. Also, it is obvious that quadruple diversity is more costly, both in initial capital outlay and maintenance. But a reference to Table I will show that nearly all of the systems in which a propagational reliability of 99.9 percent or greater is desired utilize quadruple diversity. Its use is practically mandatory in long systems and when digital data transmission is required with low error rate.

Figure 8 demonstrates a typical dual diversity terminal using predetection combining, which is now being employed on a few systems, as can be seen from an inspection of Table I. In the example shown, it is evident that if the demodulator or baseband amplifier were to experience a failure, the circuit would be down and would remain so until repairs were effected. This lack of equipment redundancy has been recognized, and it is understood that the latest dual diversity equipment employing this method of combining provides for the addition of a back-up demodulator and baseband amplifier. They would naturally be eliminated in the case of quadruple diversity terminals using this method of combining.

Figures 5 though 8 indicate certain optional equipment, to be employed when the system designer determines that the equipment is desirable or essential to produce the required RF carrier-to-noise ratio. Depending upon the improvement needed in the receiver, a low-noise preamplifier can be inserted between the preselector and down converter. It can take the form of a parametric or a tunnel diode amplifier, although transistor amplifiers are now becoming practical. Generally speaking, most modern tropo receivers have a basic noise figure of about 8 dB. The use of a tunnel diode amplifier can bring this figure down to about 4.5 dB, and the use of a good parametric amplifier can improve the figure to about 2.5 dB or even somewhat lower.

Figures 5 through 7 also show optional equipment labeled "threshold extension demodulator" for use on marginal circuits and where a very high degree of reliability is necessary. This device usually takes the form of an FM feedback demodulator connected in parallel with the regular demodulator, with provision that this threshold extension device will automatically take over when the signal approaches the receiver threshold. Although some additional distortion takes place in the threshold ex-

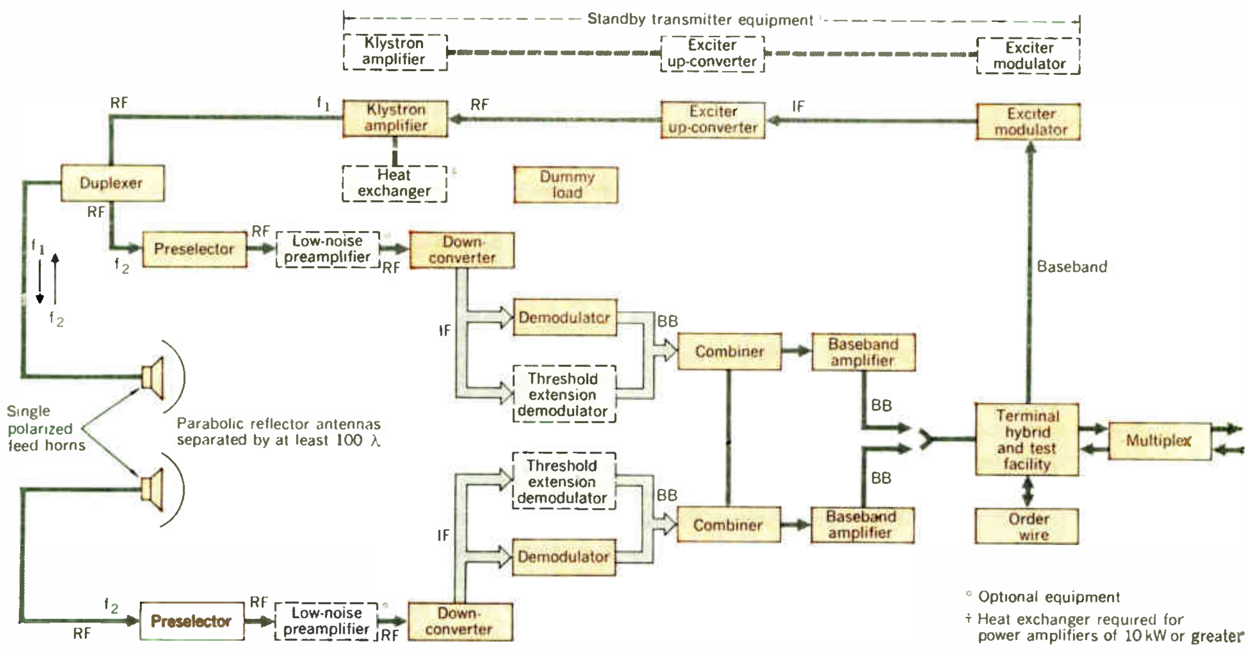
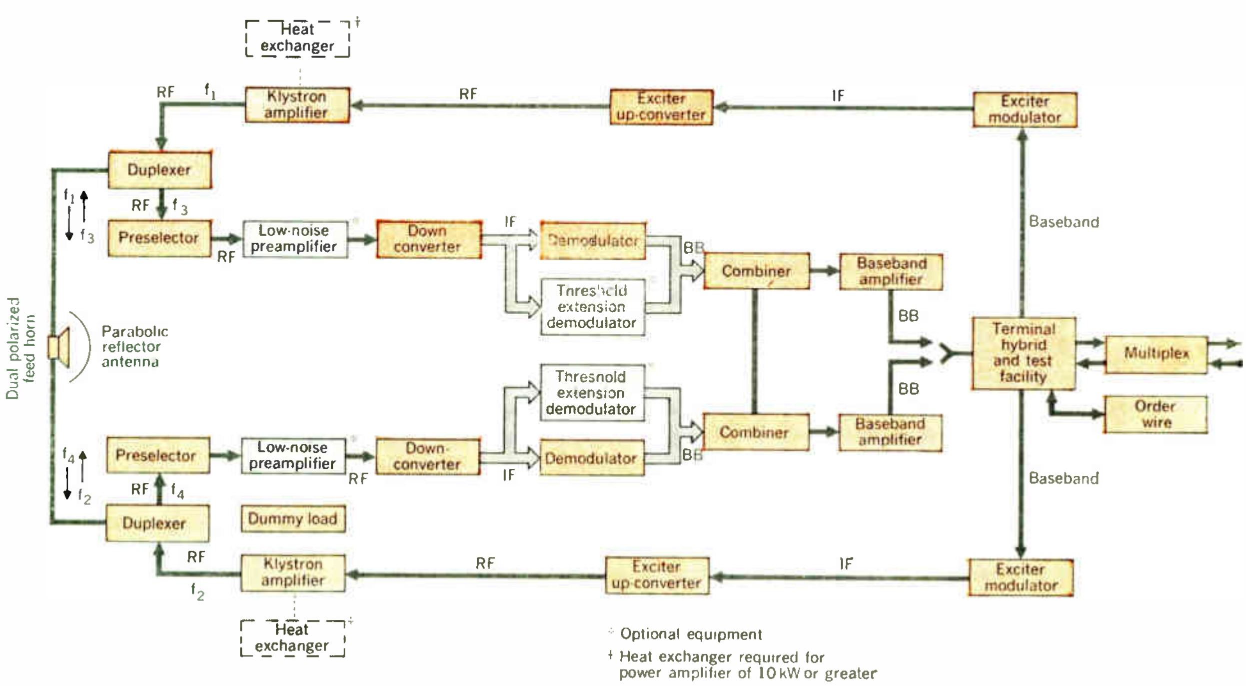


Fig. 5. Typical dual space diversity terminal.

Fig. 6. Typical dual frequency diversity terminal.



tension mode, the device is highly useful in maintaining communications on marginal circuits, particularly when propagation conditions are unfavorable. Since thermal noise usually controls system noise under these conditions, increased equipment noise is of no consequence.

In Fig. 8 it can be seen that predetection combining is performed at intermediate frequencies. For a number of years a controversy has existed over the relative merits of postdetection and predetection combining. A discussion of the detailed technical claims of the opposing sides is beyond the scope of this article. However, it is brought to the reader's attention that this controversy still exists, even though years of field experience would appear to give postdetection combining the edge.

Regardless of the merits of the claims made for both types of combining, two obvious facts should be considered. First, from theoretical analysis and laboratory controlled experiments, combining of diversity signals on a voltage basis at IF seems to have a distinct advantage over postdetection combining, which occurs after the inherent power loss in the detection process. However, the requirement that the signals to be combined at IF must be held accurately in phase with each other, in both voltage and frequency, has not been proved practical in the field for any substantial percentage of time for signals randomly scattered by the tropo mechanism.

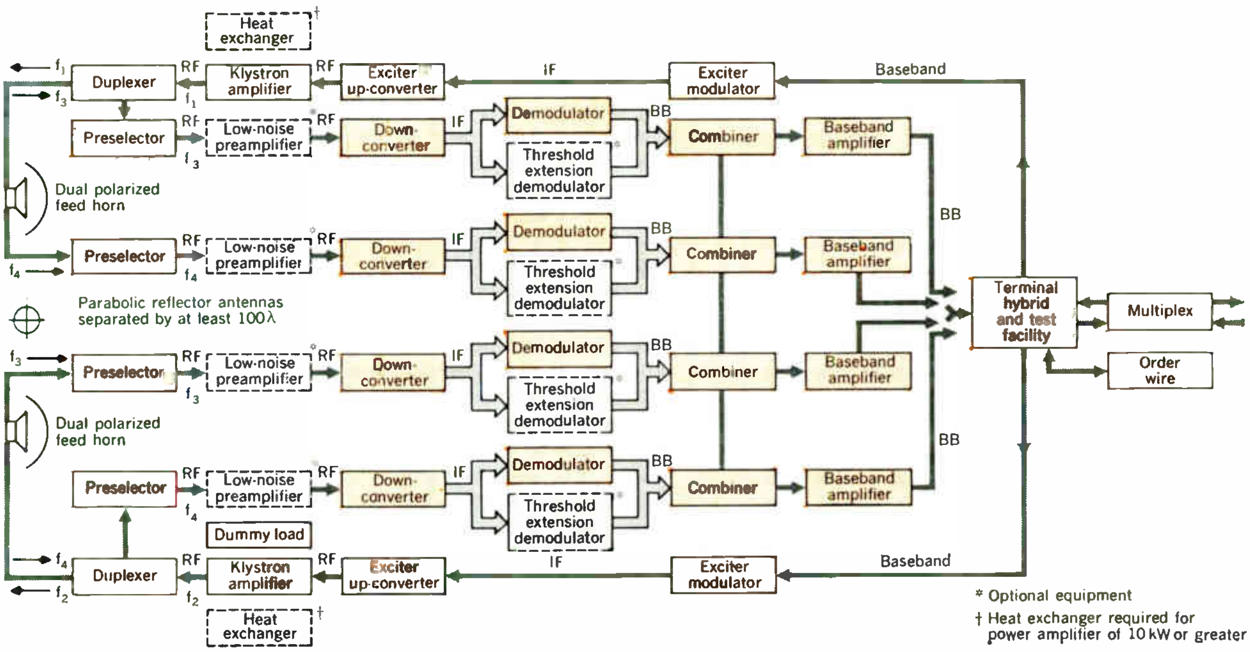
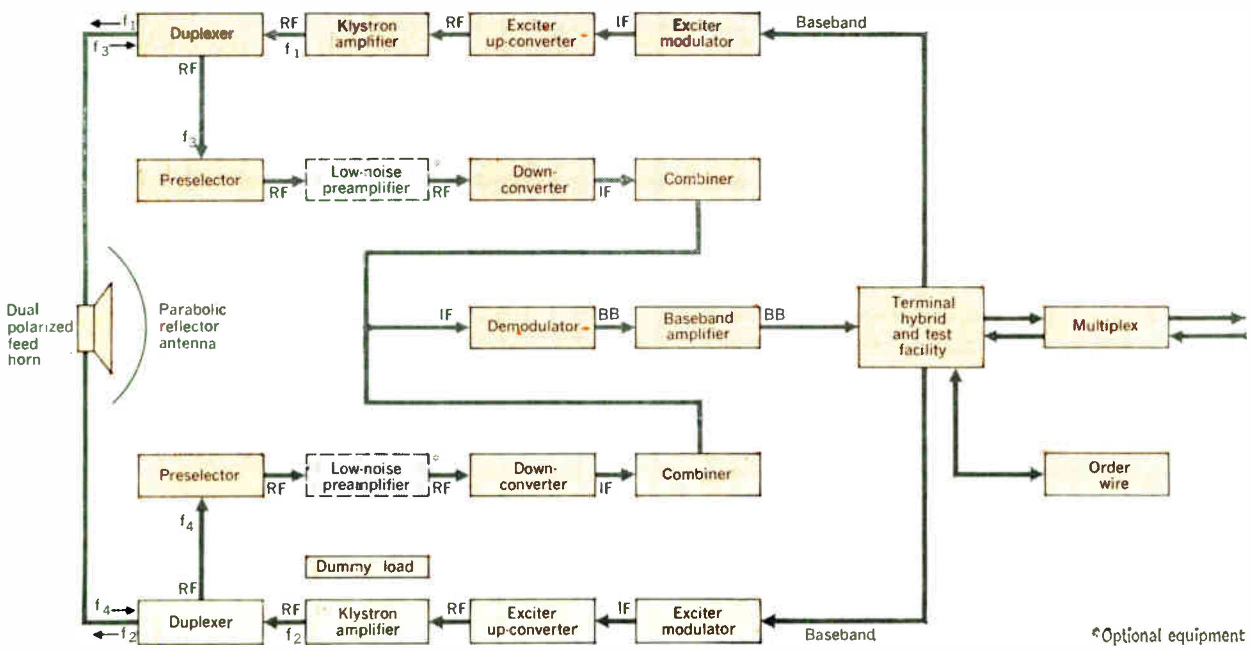


Fig. 7. Typical quadruple diversity terminal (space and frequency).

Fig. 8. Typical dual frequency terminal using predetection combining.



Second, insofar as the writer is aware, no multispan system (spans in tandem) that uses predetection combining has yet afforded operation of the quality and reliability required for toll circuit performance, particularly for data transmission. A few short single-span systems have evidently afforded satisfactory results for voice and teletypewriter circuits. From information at hand, no properly engineered system using postdetection combining has failed to offer results meeting rigid military or toll circuit requirements for both voice and data. The reader is referred to Table I, which shows the relative extent of use of postdetection combining versus predetection combining.

Advantages of tropo

Tropo has been developed over the past ten years into a highly successful method of radio communication, which offers certain advantages not possessed by other modes:

1. It provides a high-grade multichannel service over distances between 70 and 600 miles in a single span, thus reducing the number of stations or terminals required in a line-of-sight system.

2. With a properly designed system, it will offer a circuit of high propagational reliability on a year-round basis. This reliability compares very favorably with that of line-of-sight circuits, and is generally far superior to that afforded by current high-frequency techniques.

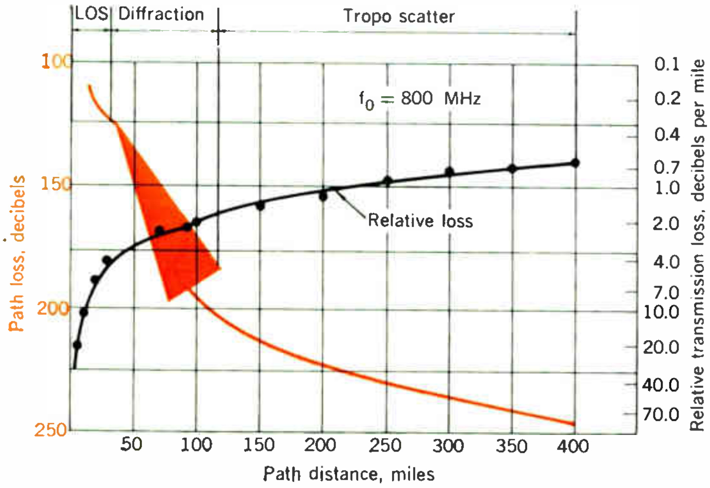
3. It can be utilized in rugged or otherwise inhospitable terrain where it is impractical or impossible to provide other means of communication.

4. It provides a high degree of spectrum utilization while simultaneously minimizing frequency allocation problems incident to radio interference in a high-density location.

5. It offers a relatively high degree of security as compared with other methods of communication. Radio interference, deliberate or otherwise, is reduced to a minimum unless the interfering transmission is within the beam and range of any one station in the tropo system. Further, there are in a tropo system fewer stations that are subject to jamming or link impairment. From a physical viewpoint, the surreptitious destruction of a tropo link is more difficult than the cutting of a submerged cable or unattended line-of-sight span. As for security obtained in a tropo link versus that obtainable in a satellite circuit, it remains to be seen whether or not the latter can be developed to the point where destructive radio interference or radio interception can compare with the security achievable on tropo systems.

Undeniably, and under certain conditions, tropo offers certain overriding advantages. When the cost of a tropo system is considered, these advantages must be weighed against the cost of other systems that may lack the advantages peculiar to tropo.

Fig. 9. Relative transmission loss.



Discussion of Table I and the maps

In Table I an attempt has been made to list all of the tropo systems or projects from the first ones completed in 1955 up to July 1, 1965, plus a few additional systems scheduled to be completed beyond that date and which are a matter of public knowledge. Unquestionably other systems, both military and common carrier, are in the planning, design, or implementation stages, but at present information is not available in sufficient detail to include these in the table.

Although a strong effort was made to make the table as complete and correct as possible, undoubtedly there are some errors in the data accumulated on a worldwide basis. Apologies are extended for such errors or any omissions. The reader will also note a number of blank spaces in the several columns where the information was either not readily available, or it was not feasible to expend additional time in further attempts to obtain it.

The table is a compilation and condensation, containing the most important data for each system. An attempt was made to accord due credit to all companies or organizations involved in each project, but no assessment has been made of the overall success of each system, since official field operational test reports are very difficult to obtain for obvious reasons. From the writer's personal knowledge, however, most, but not all, of the listed systems have been quite successful, and it is a credit to the engineering profession and to the manufacturing capabilities of the various producers that nearly all of the successful systems installed, including many of the earliest vintage, are still operating with a very satisfactory performance.

It is interesting to note from a study of Table I that most of the systems installed until the last few years were designed primarily for defense purposes. Now, many of the newer systems are utilized exclusively as common-carrier circuits, or in some cases, as a combination for government and common-carrier services. Common carriers and certain government telecommunication agencies have now recognized the importance of tropo systems, particularly where other systems are unsuitable or impossible to utilize.

Other notable facts from Table I include the preponderant use thus far of fixed stations for the various systems as compared with transportable or mobile types; the use of frequencies principally in the 900- or 2000-MHz bands; and the employment mainly of quadruple diversity particularly in systems with fixed stations, to achieve maximum reliability. With few exceptions, the power amplifier tube is a klystron with an output power of one kilowatt or greater. Frequency modulation is used on all systems with one exception.

Maps 1 and 2 should be viewed in conjunction with Table I, since the system numbering is identical. They indicate the general distribution of tropo systems throughout the world. In practically every case, the various individual systems with maps have been published over the years in public prints and technical journals, although technical details or specifications of equipment have generally been omitted in the interest of brevity or under the assumption that technical features lack public interest. In Maps 1 and 2, used with Table I, an effort has been made to consolidate the worldwide picture of tropo to provide a comprehensive view of what has happened during the past 11 years in the free world, and since the inception of this type of radio communication. Due to world political conditions, most of the tropo equipment manufactured since 1954 has been destined for military use, and the locations of the various military systems around the world as shown on the maps confirm that these defense systems have been in consonance with political policy and military deployment of forces. Maps 1 and 2 do, however, also show the various common-carrier or commercial circuits installed before July 1, 1965, and, in addition, some late ones not yet cut over.

The following figures derived from Table I may be of interest. Although approximate, they are considered fairly accurate, based on data received. The figures do not include transportable equipment, particularly that made for defense purposes, since data on such circuits are not available and vary widely depending on military deployment. It is estimated that approximately 500 000 channel miles are in use by the military on transportable equipment.

Route miles of tropo as of July 1, 1965

Total route miles established since 1955, including experimental circuits	53 600
Operational system route miles installed or under installation	51 300
Operational system route miles dismantled	1 200
Operational system route miles now in commission or being installed	50 100
Total experimental route miles established	2 300
Experimental route miles dismantled	700
Experimental route miles now in commission	1 600

Voice channel miles of tropo

Total channel miles installed since 1954, including experimental circuits	2 687 000
Operational system channel miles installed or in process of installation	2 600 000
Operational system channel miles dismantled	30 000
Operational system channel miles in commission or being installed	2 570 000
Total experimental channel miles established	87 000
Experimental channel miles dismantled	27 000
Experimental channel miles in commission	60 000

Using the above data in conjunction with Table I and counting only operational systems using fixed stations, or transportable equipment now used as fixed stations, the following figures are derived:

Average route miles per system	642
Average voice channel miles per system	33 000
Average number of tropo spans per system	4.1
Average length per span (miles)	157

Tropo costs

It has been loosely stated for years that where a choice exists as to the utilization of tropo over line-of-sight, submarine cable, or any other means of communication, tropo costs more. This statement requires examination.

First, it is true that in tropo large values of propagational path losses must be compensated by higher power transmitters, very sensitive receivers, and relatively large antennas. This may involve greater initial capital outlays for equipment. However, as a matter of interest in considering costs per channel mile for any proposed system it is interesting to study Fig. 9, which reveals that the relative loss in dB per mile for tropo is quite low when compared with diffraction path and line-of-sight propagational losses. This in itself does not indicate that tropo is cheaper to use than other modes of communication for any particular situation. There are many factors involved in a cost comparison and, for a specific system, all must be considered. They include:

1. Initial cost of all communication equipment, and its installation, required for the system—not just the cost of the radio equipment.
2. Whether or not the terrain is such that systems other than tropo can be utilized.
3. Cost of construction, including building, access roads, antenna towers, protection features, etc.
4. Estimated maintenance costs, including allowances for spare parts, power, fuel, repairs, inspections, tests, transportation, and living quarters and salaries of operating and maintenance personnel.
5. The necessity for security of communications—interception, disruption, jamming, vandalism, and sabotage.
6. Traffic studies, which can predict the volume ten years hence and thus the prospective revenues.
7. Whether or not the circuit can be considered permanent in nature or a temporary expedient.

In general, it cannot be said categorically that tropo is either cheaper or more expensive over the long-haul period than any other mode of communication. It all depends on the conditions surrounding any individual project. These conditions vary widely with each project, particularly with respect to construction and logistic problems, not to mention the grade of equipment employed. Any average figure for cost per tropo channel mile that might be derived from financial records of completed projects would be meaningless, and thus should not be used even for budgetary purposes for future projects. It is to be hoped that with the advent of solid-state tropo equipment, which will replace the tube types, the initial cost of tropo radio equipment may eventually be reduced considerably. However, this cost reduction would have no overriding effect on total system cost.

It is practically impossible to compare tropo channel mile costs with those of other types of communications, since no firm basis for comparison exists. For a given length of span or circuit and where, for topographical reasons, tropo is the only tool that will do the job, there is no comparison if modern reliable multichannel communications are required. Again, where a long span of multichannel service is desired in an undeveloped area where intermediate channel drops are unnecessary, tropo may be the cheapest solution.

It is reiterated that in any decision involving the use of tropo versus some other mode of communication, every factor must be thoroughly considered. It may well be that any higher initial cost of tropo radio equipment may be overshadowed by the higher costs of construction, maintenance, and security for other modes. The future development of lower-cost solid-state tropo equipment for commercial and common-carrier use, will go far toward furnishing a more definitive answer to the cost problem. It is believed possible that with the small size of solid-state tropo now being developed for commercial use (having a less stringent reliability factor than that involved in militarized equipment) there will be not only a reduction in initial capital outlay for equipment, but also a resultant sizable reduction in construction, maintenance, and power supply costs.

Current equipment trends

Very soon all tropo equipment purchased for future systems will be solid-state except for the power amplifier; and even in the case of the latter, future developments may be able to offer more efficient klystrons or other power tube types. Otherwise the evolution from the older tube types to solid-state is virtually complete. This change, of course, results in new equipment approximately 80 percent smaller in volume and weight, with commensurate reduction in the size of building or shelters and station main power requirements, along with improved and cheaper maintenance and logistics.

Frequency utilization in the UHF spectrum during the past ten years has risen sharply, particularly because of many line-of-sight installations, radars of all types, space satellites, and other defense and scientific endeavors. Considering propagational reliability, the size of radio equipment and antennas, and the state of development in power tubes available, 900 MHz was long considered the optimum frequency for most areas of the world. But in view of the increased use of this band in many regions, equipment was developed to use the 2000-MHz band, although in some cases it was employed to obtain better antenna gain using parabolic reflectors of standard sizes—such as 12, 15, 18, 20, 30, and 60 feet in diameter—than could be obtained on 900 MHz despite the better propagational characteristics of the latter. Even so, there exist some areas where 2000 MHz is unacceptable because of interference problems. Tropo equipment is now available in the 4000- and 5000-MHz bands; and although it can use smaller antenna structures, it lacks the propagational advantages of equipment on the lower frequencies. Higher frequencies restrict its use mainly to shorter spans.

Since the advent of solid-state tropo equipment, and even before, the trend has been toward the modularization of the different units of the equipment. This practice tends to simplify the maintenance and operations problems, as well as problems of field modifications that may prove desirable at a later date. Coupled with the trend toward modularization is the greater and greater use of fault indicator circuitry and lights to display to the operator or technician precisely which circuit or circuits in the equipment have failed.

For some years efforts were directed toward increasing the bandwidth or the number of voice channels that could be used over a tropo circuit. For a conservatively designed system, with moderate lengths of spans, equipment is now available to provide for 300 channels. The present consensus appears to be that the tropo propagation mechanism will not support a greater number of channels, at least at the present state of the art. There are occasions when a much greater channel capacity is needed, and it remains to be seen whether our development engineers will be able to solve this problem.

Until the present, the utilization of tropo by the military has generally been in the form of fixed stations or in the form of mobile- or air-transportable units used as fixed stations and considered to be strategical in nature. Now there is a growing need for small tactical equipment—easily transportable by military planes and helicopters—housed in the standard military types of shelters, where space is at a premium. Although several types of such equipment have been or are under development, none appear to be the answer to all military requirements for this class of equipment.

In addition to compromises in channel capacity, weight, reliability, output of power, and order of diversity, one of the most grievous problems for the tactical set involves antennas. The size of the parabolic reflectors can be reduced somewhat with the use of frequencies in the higher UHF or SHF ranges. However, still to be solved is the problem of devising an antenna system that is collapsible, light in weight, easy to stow or pack, and that can be set up in the field in less than 30 minutes, including orientation of the dish to the proper azimuth. The necessity for the concentration of the RF signals in a narrow beam has so far precluded the successful use of other types of antennas. A number of development efforts have been undertaken to provide a tactical tropo antenna meeting the foregoing requirements, but none are considered completely satisfactory. No doubt further efforts are in order, and will be made, to solve this problem, but some new and radical approach is needed.

Another important development concerns the application of digital forms of multiplexing and modulation on tropo systems. Generally, these are variants of time-division techniques, with PCM being most prominent. They appear to be most promising to the military who are concerned with secure communications. In addition, the use of digital techniques allows operation with lower signal-to-noise ratios and therefore, on some systems, PCM can provide a better grade of service than FM. However, the propagational characteristics of tropo make the advantages of PCM obtained on the cable and line-of-sight systems more difficult to realize.

With increasing demands for tactical military tropo, and concomitant requirements for smaller size and weight of equipment, further efforts in development appear to be in order. Tropo equipment can now be considered to be nearly solid-state, the chief exception being in the power amplifiers. This situation leads to efforts in:

1. Reduction in size of klystron tubes, traveling-wave tubes, or other power-tube types and their necessary power supplies.

2. Development of solid-state devices to a level where they can replace the power tubes in the power amplifier

stage or modulator–exciter output stage, if used in the latter.

3. Reduction in size of packaged units, consonant with required heat dissipation.

4. Development of microminiature integrated circuits where these can be used to replace present solid-state circuits.

At the present time, development of these integrated circuits has reached a point where certain substitutions are possible. Any or all of the foregoing advances could lead to material reductions in size, weight, and cost.

Reference to Table I shows that the world is becoming increasingly aware of tropo for use in commercial or common-carrier systems. It is, therefore, desirable to develop commercial or common-carrier tropo equipment, based to a great extent on CCIR and CCITT specifications, and with a voice-channel capacity from 4 to 200 channels for this type of service. It may be necessary to divide this approach into two or more separate developments because of the wide variances in channel requirements and power outputs. Any such development should, of course, be confined to solid-state equipment, and should be capable of installation in a modest-size trailer or shelter when the requirement may be of a temporary nature, such as a mining exploration project, far removed from the headquarters in overall control but where ordinary commercial communication facilities are not available.

To summarize the present trends in the current development of tropo equipment, we can list the following:

1. Completion of transformation of all tropo equipment from tube types to solid-state, though still retaining power tubes in power amplifiers.

2. Further efforts to decrease size, weight, and main power requirements for solid-state equipment. These efforts include development of relatively high output

I. Tropo systems or projects completed from 1955 to July 1, 1965, plus additional systems, already announced publicly, scheduled to be completed at later dates

Arbitrary System Designation	Project or System Name, Year Completed, and Location	Prime Contractor or System Design Agency and System Use	No. of Tropo Scatter Spans	Total Route of Tropo Scatter in System, statute miles	Estimated Channel Miles	Lengths of Minimum and Maximum Spans, statute miles	Station Types: Fixed (F), Transport (T), or Mobile (M)	System Frequency Bands, MHz	Present Voice Channel Capacity	Ultimate Voice Channel Capacity	Tropo-Radio Manufacturers	Multiplex Manufacturers	Antenna Manufacturers	Reflector Diameters, feet	Transmitter Power Outputs, kW	Receiver Noise Figures, dB	Order of Diversity: Quad (Q), and/or Dual (D), Single (S), Triple (T), None (N)	Arbitrary System Designation
A	Nutley–Southampton, 1955, New York	ITT Exp[1]	1	91	Var		F	900	Var	Var	ITT	ITT	KENN	28	Var	8	D	A
B	Texas Towers prototype, 1956, Massachusetts[a]	PAGE Exp[1]	2	210	15 120	80 130	F	900	72	72	REL NRC	LENK	KENN	30	10	6	T	B
C	AN TRC-43 (XW-1), 1957, Verona, N.Y.–Baltimore	WEST Exp[1]	1	268			F	900	Var		WEST	WE	PROD	28	10	9	N	C
D	Denhelder–Domburg, NPTT, 1957, Netherlands[b]	NPTT Exp[1]	1	90	90		F	900	1		PHTE			16	80W	9	S	D
E	Kinmount–Sugarpoint, 1958, Ontario[a]	CWES Exp[1]	2	270	6 480	100 170	T	5000	24	120	CWES	STRC	PMS	16 12	2	9.5	D	E
F	Lannion–Conches, 1958, France	CFTH/CNET Exp[1]	1	200	12 000		F	2000	60		CGDT CFTH		CGDT	40	10	10	D	F
G	West Beckham–Domburg, 1958, U.K.–Netherlands[b]	NPTT/GPO Exp[1]	1	112	112		F	900	1		PHTE			16	100W	9	S	G
H	South England Tropo, 1958–59, England[a]	MARC Exp[1]	1	206	4 944		F	900	24		MARC			30	10		D	H
J	AN/FRC-53/54, 1959, Northern New York	WEST Exp[1]	1	171	4 104		F	2000	24	72	WEST		KENN PROD	28	10 1	10	D	J
K	Paris–Nancay, 1959, France	CGE Exp[1]	1	103	9 682		F	900	94					33	2	8	D	K
L	AN/GPA-35 (XW-1), 1960, Florida	WEST Exp[1]	2	244	976	121 123	F	5000	4		WEST	LENK	KENN		2	9	Q	L
M	AN/TRC-66, 1960, Carolinas	WEST Exp[1]	3	209	10 032	56 93	T	5000	48		WEST	LENK	WEST	14	1	4	Q	M
N	Radas Experimental, 1962, Florida[a]	MART Exp[1]	1	100	4 800		T	5000	48		CWES	MART		12			D	N
O	Radas Experimental, 1962, Florida[c]	MART Exp[1]	1	210	10 080		T	5000	48		CWES	MART		12			D	O
1	Polevault, 1955, Canada[d]	BTCC Def[2]	9	1609	57 924	130 228	F	900	36	36	REL	STC	KENN	60	10	9	D	1
2	Anchorage–Kodiak, 1955, Alaska[a]	PAGE CC[3]	2	190	1 140	65 125	F	100	6	12	REL WEST	WE	PAGE	30	1	4	D	2

Arbitrary System Designation	Project or System Name, Year Completed, and Location	Prime Contractor or System Design Agency and System Use	No. of Tropo Scatter Spans	Total Route of Tropo Scatter in System, statute miles	Estimated Channel Miles	Lengths of Minimum and Maximum Spans, statute miles	Station Types: Fixed (F), Transport (T), or Mobile (M)	System Frequency Bands, MHz	Present Voice Channel Capacity	Ultimate Voice Channel Capacity	Tropo-Radio Manufacturers	Multiplex Manufacturers	Antenna Manufacturers	Reflector Diameters, feet	Transmitter Power Outputs, kW	Receiver Noise Figures, dB	Order of Diversity: Quad (Q), and/or Dual (D), Single (S), Triple (T), None (N)	Arbitrary System Designation
3	AST-101, 1957, No specific location	CRC Def[2]					T	900	4 12	12	CRC	CRC	CRC	15	1	8	D	3
4	Sardinia–Minorca, 1957, Mediterranean[c]	ITT CC[3]	1	240	1 440		F	800	6	36	ITT	ITT		66	0.5		D	4
5	Florida–Cuba, 1957, S. Florida–Havana[f]	ATT/ITT CC[3]	1	185	8 695		F	750	36[n]	240[n]	ITT	WE	BKC	60	10	9.5	DQ	5
6	Mid-Can South, 1957, Eastern Canada	BTCC Def[2]	4	695	21 420	95 180	F	900	36	36	NOEC	GECL	DBCL	60	1 10	10	D	6
7	Senorita, 1957, Spain– Morocco	ITT Def[2]	2	483	17 388	218 265	F	900	36	48	ITT REL	LENK	KENN	60	10		Q	7
8	Puerto Rico—Dominican Tropo, 1957, Caribbean[e]	ITT/GTE CC[3]	1	240	1 440		F	900	6	36	ITT	LENK	KENN	28	1		D	8
9	Bougie–Ben Mansour, 1957, Algeria	CFTH CC[3]	1	50	1 200		F	450	24	24	CFTH	CFTH	CFTH	18	1	6	D	9
10	Quebec–Labrador, 1958 Eastern Canada	BTCC Comb[4]	5	570	68 400	80 150	F	900	48 192	48 252	NOEC	NOEC	DBCL AC	28, 45 60	1 10	7 2.5	Q	10
11	TST-101, 1958, No specific location	CRC Def[2]					T	900	4, 12 24	24	CRC	CRC	CRC	15	1	8	D	11
12	Dew Line, 1958, Alaska– Northern Canada[g]	WE Def[2]	29	2695	64 680	60 120	F	900	24	24	CRC	LENK	BKC	30 60	1	8	D	12
13	Northern Telephone Co., 1958, Kenora–Redlake, Can.	CRC CC[3]	1	91	364		TF	900	4	12	CRC	CRC	CRC	15	1	8	D	13
14	Venezuela Scatter System, 1958, Venezuela[h]	CRC GCmr[5,6]	2	344	8 256	110 234	TF	900	24	24	CRC	PHTE CRC	TAC KENN	28	1	8	Q	14
15	White Alice, 1958, Alaska	WE Comb[4]	23	3135	413 820	70 170	F	900	132	132	REL	WE	BKC	30 60	1 10	8	DQ	15
16	Polevault, "Dog-Legs," 1958, Lab.-Newfoundland	RCA Def[2]	6	440	1 760	46 95	F	900	4	4	REL		KENN	28	1	8	D	16
17	Iceland Tropo, 1958, Iceland[h]	RCA Def[2]	3	650	2 600	167 270	F	900	4	4	REL		KENN	28 60	10	8	DQ	17
18	Texas Towers, 1958, East Coast, U.S.[h]	RCA Def[2]	3	310	22 320	81 124	F	2000	72	72	REL	LENK	KENN	28	10	8	Q	18
19	Hot Line, 1958, Norway[h]	ITT Def[2]	3	525	18 900	100 235	F	900 2000	36		ITT NRC			28 60	10		Q	19
20	Constantine–Tebessa, 1958, Algeria	CFTH CC[3]	1	66	1 584		F	450	24	24	CFTH	CFTH	CFTH	18	1		D	20
21	AST-102, 1959, No specific location	CRC Def[2]					T	2000	4,12 24	24	CRC	CRC	CRC	15	1	9.5	D	21
22	Libya Tropo, 1959, Libya	HEPC CC[3]	2	340	4 760	120 220	F	900	14	72	REL	GECL	KENN	28 60	1 10	7	Q	22
23	Orinoco Mining Co., 1959, Venezuela[h]	CRC Cmr[6]	1	324	1 620		F	900	5	24	CRC	CRC	TAC	28	1	8	D	23
24	Blue Grass, 1959, Alaska	WE Comb[4]	8	1395	100 440	80 410	F	900	72	72	REL	WE	BKC	20,60 120	1,10 50	2 8	Q	24
25	Berlin–Torphaus, 1959, West Germany	SIEM CC[3]	1	118	14 160		F	2000	120	132	SIEM	SIEM	SIEM	33	1	8	D	25
26	AEC, 1959, Eniwetok– Kwajalein	HAN G[5]	1	350	8 400		F	900	24	24	LNP	LENK						26
27	Florida–Bahama Tropo, 1960, S.E. U.S.–Bahamas[e]	ATT/ITT CC[3]	1	186	6 696		F	2000	36	72	ITT	WE	BKC KENN	30	10	10	DQ	27
28	Gulf Missile Range, 1959, Gulf of Mexico	PHIL Def[2]	1	180	4 320		F	2000		72	PHIL			60	10		Q	28
29	Medea–Laghouat, 1960, Algeria	CFTH CC[3]	3	181	4 344	25 91	F	450	24	24	CFTH	LTT	CFTH	18 30	100W 1W	6	D	29
30	Dew Drop, 1960, Thule– Cape Dyer[i]	GE Def[2]	1	625	7 500		F	400	1, 12 24	24	GE	LENK	KENN	120	3 10	2	Q	30

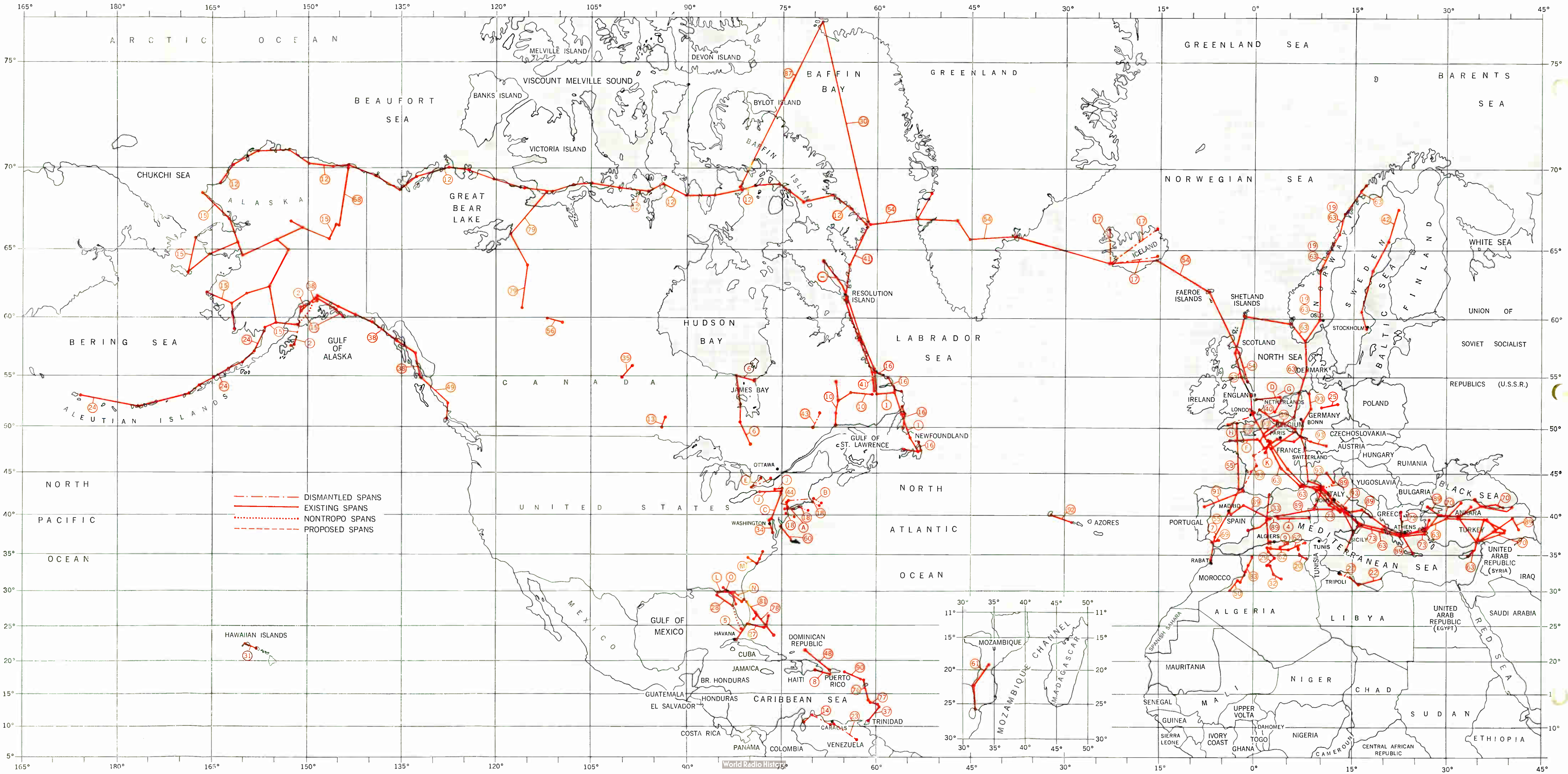
Arbitrary System Designation	Project or System Name, Year Completed, and Location	Prime Contractor or System Design Agency and System Use	No. of Tropo Scatter Spans	Total Route of Tropo Scatter in System, statute miles	Estimated Channel Miles	Lengths of Minimum and Maximum Spans, statute miles	Station Types: Fixed (F), Transport (T), or Mobile (M)	System Frequency Bands, MHz	Present Voice Channel Capacity	Ultimate Voice Channel Capacity	Tropo-Radio Manufacturers	Multiplex Manufacturers	Antenna Manufacturers	Reflector Diameters, feet	Transmitter Power Outputs, kW	Receiver Noise Figures, dB	Order of Diversity: Quad (Q), and/or Dual (D), Single (S), Triple (T), None (N)	Arbitrary System Designation
31	Pacscat, 1960, Hawaii	PAGE Def[2]	1	114	2 736		F	900	24	72	REL	GE	TAC	18	1	8	Q	31
32	Laghouat–Ouargla, 1960, Algeria	CFTH CC[3]	2	203	4 872	101 102	F	450	24	24	CFTH	LTT	CFTH	30	1	6	D	32
33	France–Algeria TV, 1960, France–Algeria[j]	TRT G[6]	2	402		185 217	F	4000	Video	Video	TRT		TRT	18	500W	9	D	33
34	East Coast Relay, 1960, Maryland,	CRC Def[2]	1	68	8 160		F	2000	120	132	CRC	CRC	AC	28	1	8	Q	34
35	Snow Lake–Thompson, 1960, Manitoba, Canada	RCA CC[3]	1	120	3 600		F	450	30	48	RCA	LENK	KENN	28	1	6	D	35
36	ADCOM II, 1960, Eastern Canada	CWES Def[2]	1				T	5000	24	48	CWES	LENK	AC	10		9.5	Q	36
37	Trinidad–Barbados, 1961, S.E. Caribbean	MARC CC[3]	1	205	2 460		F	750	12	60	MARC	ATE	MARC	30	1	10	Q	37
38	BMEWS–Alaska, 1961 Alaska	WE Comb[4]	7	855	164 160	100 180	F	900	192	240	REL	WE	BKC	60	1 10	2	Q	38
39	Schonfeld–Flobecq, 1961, Germany–Belgium	SIEM Def[2]	1	118	7 080		F	2000	60	132	SIEM	SIEM	KENN	28	1	8	D	39
40	Flobecq–Martlesham Heath, 1961, Belgium–England	SIEM Def[2]	1	141	8 460		F	900	60		REL	SIEM	KENN	28	1	8	Q	40
41	Polevault Rebuild, 1961, Canada	WE Comb[4]	5	955	78 840	150 228	F	900	72 108	132	REL	WE	BKC VICK	30, 60 120	1 10	2	Q	41
42	R.S.A.F., 1961, Sweden	CWES Def[2]	4	610	29 280	200 100	F	5000	48	120	CWES	LENK	PMS	16 10	2	9	Q	42
43	Manicougam, 1961, Quebec[a]	CWES CC[3]	1	115	2 760		T	5000	24	120	CWES	STRC	PMS	16 12	2	9.5	D	43
44	Ft. Wadsworth, 1961, S.E. New York	ITT Def[2]	1	45	540		T	2000	12	24	ITT	GE	AC	15	1	8	D	44
45	Jasdaf, 1961, Japan	NEC Def[2]	13	1708	61 488	91 145	F	1800	12 48		NEC			33	100W			45
46	Makpo–Cheju, 1962, Korea	MARC CC[3]	1	98	1 176		F	750	12	60	MARC	ATE	MARC	30	1	10	Q	46
47	Indocom, 1962, Indonesia	PHIL Comb[4]	1	260	6 240		F	900	24	24	PHIL	CRC	AC	28	1	8	Q	47
48	Atlantic Missile Range, 1962, Caribbean	PAGE Comb[4]	1	382	9 168		T'	900	24	72	REL	GE	KENN	60	10	2	Q	48
49	B.C. Telephone Co., 1962, British Columbia–Alaska	LENK Comb[4]	2	345	16 560	155 190	F	900	48	252	REL	LENK	WEBR	60	10	2.5	Q	49
50	Colomb Bechar Hammaguir, 1962, Algeria	CFTH Def[2]	1	71	2 556		F	500		36	CFTH	CIT	CFTH	30	1	6	D	50
51	Philippine Tropo, 1962, Philippines	BEND Def[2]	3	425	5 100	105 185	F	900	12	12	REL	LENK	ASI	28	1	8	Q	51
52	Back Porch, 1962, South Vietnam	PAGE DefG[2,5]	5	820	59 050	106 224	T	900	72	72	REL	LENK	ASI	60	10	2	Q	52
53	Taiwan–Philippines, 1962, West Pacific[k]	ITT Def[2]	1	310	3 720		F	450	12	24	ITT		KENN	60	10		Q	53
54	Dew East Nars, 1962, North Atlantic	WE Comb[4]	10	2420	203 280	120 450	F	900	84	84	REL	WE	Var	30, 60 120	10 50	2	Q	54
55	U.K.–Spain–Morocco, 1962, Western Europe	PAGE Def[2]	4	1230	29 520	209 515	F	900 400	24	24 72	REL RAD	SIEM	BKC	60 120	10 50	2 8	Q	55
56	Alberta Govt. Telephone, 1963, Alberta, Canada	CWES CC[3]	1	90	2 160		F	5000	24	120	CWES	LENK	AC	28	5	4.5	Q	56
57	Okinawa–Miyako Jima, 1963, Okinawa	PAGE Def[2]	1	176	2 112		F	2000	12	72	REL	GE	TAC	30	10W	8	Q	57
58	Navy Philippine Tropo, 1962, Philippines	USN Def[2]	1	45	540		F	900	24	24	CRC	LENK		30	10	9	Q	58
59	Thin Route Tropo, 1962–63, Various	GE DefCmr[2,6]					FT	400	1	1	GE	GE	TAC AC	28	500W	6	DQ	59
60	Condor I, 1963, Atlantic Coast	USN/REL Def[2]	1	Var.	4 800	Var. 250	FM	400	24	72	REL	LENK	AC	18 30	10	3.5	D	60
61	Lourenco Marques Beira, 1963, Mozambique	CFTH CC[3]	2	521	21 882	248 273	F	900	42	120	CFTH	SAT	CEI	60	1	2.5	Q	61
62	Alger–Bone, 1963, Algeria	CGE/CFTH CC[3]	2	224	18 816	81 143	F	900	84	84	CGE CFTH	CIT	CFTH	30	1.5	8	Q	62
63	Ace High, 1963, Europe	SHAPE Def[2]	40	6750	243 000	105 250	F	900 400	36	60	REL PYE	STC	ITE KRUP	33 66	1 10	8	Q	63
64	Okinawa–Japan, 1963, Japan	NEC Def[2]	9	1705	40 920	80 400	F	2000	24	60	NEC	NEC		60	10	3.5		64
65	Pacific Missile Range, 1963, California	BEND Def[2]	1				F	5000	3		CRC	CRC FEC	CRC	10	1			65
66	Bangkok–Saigon, 1963, S.E. Asia[a]	PHIL Def[2]	1	450	10 800		F	2000	24	60	PHIL	GE	KENN	60	10	8	Q	66
67	Okinawa–Taiwan, 1963, Western Pacific	CRC Def[2]	2	601	14 424	192 409	F	900	24	24	CRC	GE	BKC AC	60 28	10	2	Q	67
68	Yukon–Barter Island, 1963, Alaska	WE Comb[4]	1	230	8 280		F	900	36	72	REL	WE	BKC	120	50	2	Q	68
69	Condor II, 1963, Spain	USN/REL Def[2]	1	58	1 392		M	2000	24	48	REL	LENK	AC	30	1	8	Q	69
70	Turkey Tropo, 1963, Turkey	PAGE Def[2]	8	1179	28 296	78 268	FM	900	24	24	REL	SIEM	ASI	60 28	1 10	3	Q	70
71	AU/MRC/98, 1963, No specific location	BEND Def[2]	16				TM	900	24	252	REL	LENK	ITE	28	10	2	Q	71
72	Rokaf, 1964, South Korea[e]	ITT Def[2]	3	445	32 046	110 175	F	5000	24 120	240	ITT	ITT	KENN	6, 15 30, 40	1		DQ	72
73	Big Rally II, 1964, Southern Europe	ITT Def[2]	6	1200	28 800	150 300	T	400	24	24	REL	MOTO	ASI	30	1 10	2	Q	73
74	AN/TRC-80, 1961–64, Various	CRC Def[2]					T	5000	1	1	CRC	CRC	BIRD	8	1	9	D	74
75	AN/TRC 90, 90A, 90B, 90C, 1962–64, Various	CRC Def[2]					T	5000	1,12 24	24 48	CRC	CRC NORC	CRC KENN	10, 15 29	1	5.5 9	DQ	75
76	Antigua–St. Lucia, 1964, Caribbean	MARC CC[3]	1	220	10 560		F	900	48	72	MARC REL	TMC	KENN	30	1	2	Q	76
77	St. Lucia–Barbados, 1964, Caribbean	MARC CC[3]	1	112	8 064		F	2000	72	72	REL	TMC	KENN	20	5W	8	Q	77
78	Bahama Telecom, 1964, Bahama Islands	PAGE CC[3]	4	380	9 120	65 125	F	2000	24	60	FEC SIER	LENK FEC	ASI AC		1	5		78
79	CNT–Northwest Tropo, 1964, N.W. Canada	PAGE CC[3]	3	640	16 640	190 240	F	2000	26	72	NOEC	LENK	DBCL	60	1	2.5	Q	79
80	Bangkok–Ubon, 1964, S.E. Asia	PHIL Def[2]	2	310	7 440	155 155	F	2000	24	60	PHIL	GE	KENN	60	10	8	Q	80
81	AMR, 1964, Cape Kennedy–Grand Bahama Island	PAGE NASA	1	170	12 240		M	900	72	72	REL	LENK	ASI	60	10	2.5	Q	81
82	Wet Wash, 1964, South Vietnam	PAGE Def[2]	1	180	12 960		M	900	72	240	REL	LENK	ITE	30	10	2	Q	82
83	Saida–Colomb Bechar, 1964, Algeria	CFTH CC[3]	3	270	9 720	57 123	F	500	36	60	CFTH		CFTH	30	1	6	D	83
84	Phil. Cable Extension, 1965, Philippines	GTE CC[4]	1	70	9 240		F	2000	132	240	REL	LENK		30	1	2.5	Q	84
85	Philippine Telecom, 1965, Philippines	ITT CC[3]	3	415	29 880	100 180	F	900	72	240	REL	ITT	ROHR		1 10	8	Q	85
86	Phil–Taiwan Improvement, 1965, Philippines	PAGE Def[2]	1	310	7 440		F	600	24	48	ITT FEC	CRC	KENN BKC	60 120	10	2	Q	86
87	Thule–Fox, 1965, Canada–Greenland	PHIL Def[2]	1	590	7 080		F	400	12	72	REL	LENK	DBCL	120	100	2.5	Q	87
88	LL-176, 1965, United States	ATT CC[3]	4		54 000		F	2000	132	300	REL	WE					Q	88
89	EMT, 1965, South Europe	ITT Def[2]	18	3267	235 224	68 282	F	2600	72	120	REL	RAY	BKC	30 60	1 10	2.5	Q	89
90	Antigua–Tortola, 1965, Caribbean	MARC CC[3]	1	220	17 600		F	2000		300	REL				1	2.5	Q	90
91	France–Portugal, 1966, West Europe	CFTH CC[3]	1	293	10 548		F	900	36	120	CFTH		CEI	90	10	2.5	Q	91
92	Fayal–Flores, 1966, Azores	CFTH CC[3]	1	170	2 040		F	900	12	12	CFTH		CEI	30	150W	2.5	D	92
93	ET-A, 1966, West Europe[e]	ITT GDef[2,5]	18	2080	114 930	50 240	FT	5000	24, 36 60, 72	120*	ITT	GE	KENN	6, 15 30, 40	1 10		DQ	93
94	AN/TRC-97, 1966 Various[m]	RCA Def[2]					T	5000	24	24	RCA			8	1	4.5	D	94

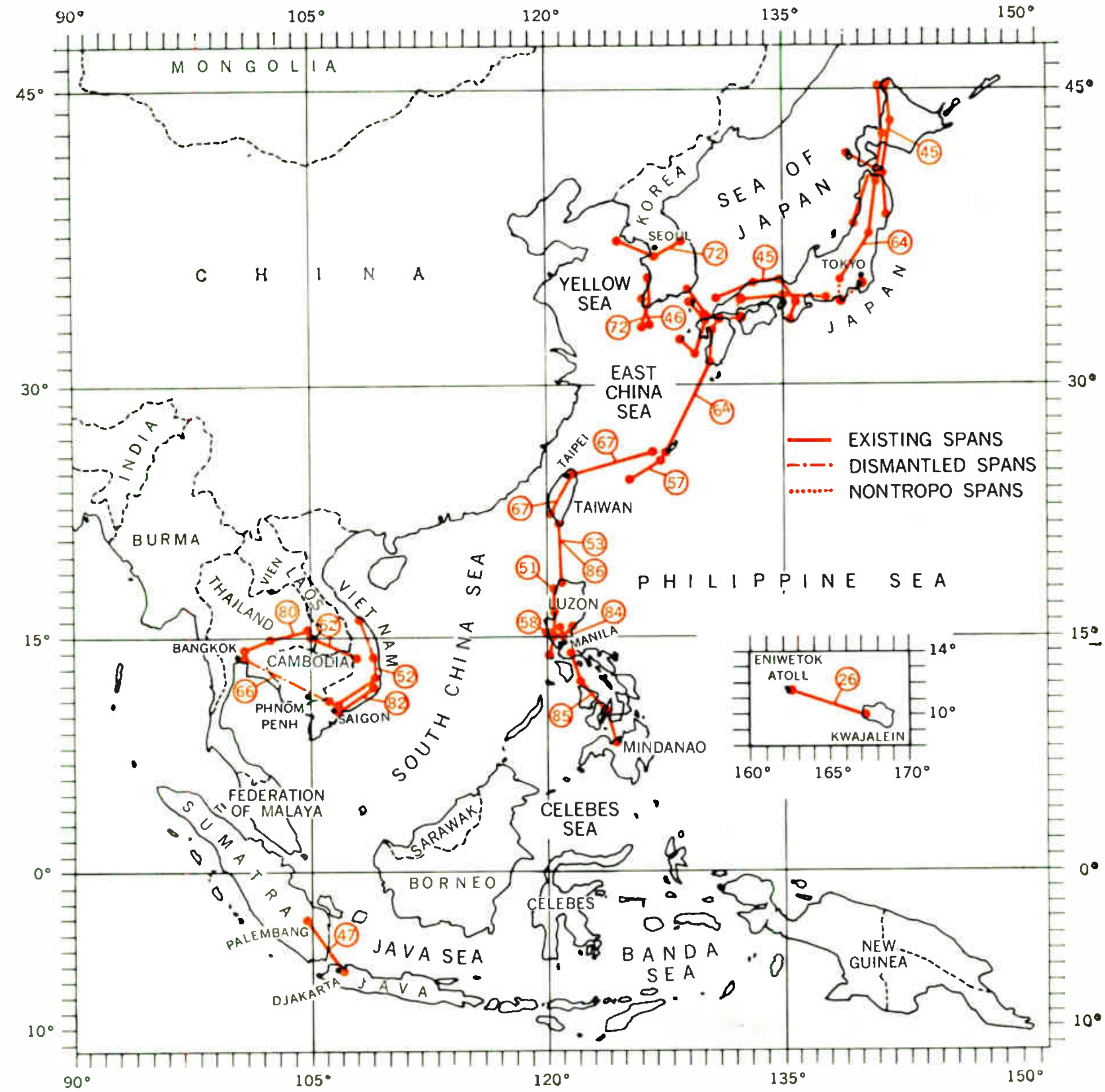
1 Experimental
2 Defense
3 Common carrier
4 Combination
5 Government
6 Commercial
a Dismantled
b Unidirectional
c Radas tests
d World's first system designed for operational traffic
e Predetection combining
f When operating in dual diversity, capable of handling 1 TV channel and 120 voice channels. Predetection combining.
g Under retrofit
h Now part of Ace High (see System No. 63)
i The only system utilizing amplitude modulation
j TV service—predetection combining
k Predetection combining (being upgraded)
m Phase-lock threshold ext.
n and/or video

* Also present capacity on some spans

Key to companies and agencies listed in abbreviated form in this table

AC	Andrew Corp.
AEC	Atomic Energy Commission
ASI	Antenna Systems, Inc.
ATE	Automatic Telephone & Electric Co. (England)
ATT	American Telephone & Telegraph Co.
BEND	Bendix Corp.
BIRD	Bird Air Structures
BKC	Blaw-Knox Co.
BTCC	Bell Telephone Co. of Canada
CEI	Compania Elettronica Italiana, S.p.A.
CFTH	Compagnie Francaise Thomson-Houston
CGDT	Compagnie General de Telegraphie Sans Fil
CGE	Compagnie Generale d'Electricité
CIT	Compagnie Industrielle des Telecommunications
CNET	Compagnie Nationale Etudes de Telecommunications
CNT	Canadian National Telecommunications
CRC	Collins Radio Co.
CWES	Canadian Westinghouse Co. Ltd.
DBCL	Dominion Bridge Co., Ltd.
FEC	Farinon Electric Co.
GE	General Electric Co.
GECL	General Electric Co., Ltd. (U.K.)
GPO	General Post Office (U.K.)
GTE	General Telephone & Electronics Corp.
HAN	Holmes and Narver
HEPC	Hycon-Eastern Co. and Page Communication Engineers, Inc.
ITE	I-T-E Circuit Breaker Co.
ITT	International Telephone & Telegraph Corp.
KENN	D. S. Kennedy Co.
KRUP	Friedrich Krupp, GMBH
LENK	Lenkurt Electric Co., Inc.
LINC	Lincoln Lab.
LNP	Lincoln Lab., National Radio Co., Page Communication Engineers, Inc.
LTT	Lignes Telephoniques et Telegraphiques
MARC	The Marconi Co., Ltd.
MART	The Martin Marietta Co.
MOTO	Motorola Inc.
NASA	National Aeronautical and Space Administration
NEC	Nippon Electric Co., Inc.
NOEC	Northern Electric Co., Ltd.
NORC	Northern Radio Co.
NPTT	Netherlands Postal, Telephone & Telegraph Bureau
NRC	National Radio Co.
PAGE	Page Communication Engineers, Inc.
PHIL	Philco Corp.
PHTE	N. V. Philips Telecommunicate Industrie
PMS	Precision Metal Spinning, Ltd.
PROD	Prodelin, Inc.
PYE	Pye Telecommunications, Ltd.
RAD	Radiation, Inc.
RAY	Raytheon Co.
RCA	Radio Corp. of America
REL	Radio Engineering Laboratories
ROHR	Rohr Corp.
SAT	Société Anonyme de Telecommunications
SHAPE	Supreme Headquarters, Allied Powers, Europe
SIEM	Siemens & Halske, A.G.
SIER	Sierra Electronic Div., Philco Corp.
STC	Standard Telephones & Cables, Ltd.
STRC	Stromberg-Carlson Corp.
TAC	Technical Appliance Corp.
TMC	Telephone Manufacturing Co., Ltd.
TRT	Telecommunications Radio-Electriques et Telephoniques
USN	U.S. Navy
VICK	Canadian Vickers Co. Ltd.
WE	Western Electric Co.
WEBR	Western Bridge Co., Ltd.
WEST	Westinghouse Electric Corp.


ARCTIC OCEAN
BEAUFORT SEA
MELVILLE ISLAND
DEVON ISLAND
VISCOUNT MELVILLE SOUND
BANKS ISLAND
VICTORIA ISLAND
BYLOT ISLAND
BAFFIN BAY
BAFFIN ISLAND
GREENLAND
GREENLAND SEA
BARENTS SEA
CHUKCHI SEA
ALASKA
GREAT BEAR LAKE
NORWEGIAN SEA
WHITE SEA
ICELAND
FAEROE ISLANDS
SHETLAND ISLANDS
NORWAY
SWEDEN
FINLAND
BALTIC SEA
OSLO
STOCKHOLM
UNION OF SOVIET SOCIALIST REPUBLICS (U.S.S.R.)
BERING SEA
GULF OF ALASKA
HUDSON BAY
RESOLUTION ISLAND
LABRADOR SEA
SCOTLAND
NORTH SEA
DENMARK
ALEUTIAN ISLANDS
CANADA
JAMES BAY
IRELAND
ENGLAND
LONDON
NETHERLANDS
GERMANY
BONN
POLAND
BELGIUM
PARIS
CZECHOSLOVAKIA
NEWFOUNDLAND
GULF OF ST. LAWRENCE
AUSTRIA
FRANCE
SWITZERLAND
HUNGARY
RUMANIA
OTTAWA
NORTH
YUGOSLAVIA
BLACK SEA
BULGARIA
DISMANTLED SPANS
EXISTING SPANS
NONTROPO SPANS
PROPOSED SPANS
UNITED STATES
NORTH ATLANTIC OCEAN
ITALY
ROME
MADRID
SPAIN
PORTUGAL
GREECE
ANKARA
ATHENS
TURKEY
WASHINGTON
AZORES
ALGIERS
MEDITERRANEAN SEA
SICILY
TUNIS
TUNISIA
NORTH PACIFIC OCEAN
RABAT
MOROCCO
TRIPOLI
UNITED ARAB REPUBLIC (SYRIA)
IRAQ
MEXICO
GULF OF MEXICO
HAVANA
ALGERIA
LIBYA
UNITED ARAB REPUBLIC (EGYPT)
SAUDI ARABIA
RED SEA
SPANISH SAHARA
HAWAIIAN ISLANDS
CUBA
DOMINICAN REPUBLIC
MOZAMBIQUE
MOZAMBIQUE CHANNEL
MADAGASCAR
MAURITANIA
JAMAICA
HAITI
PUERTO RICO
BR. HONDURAS
NIGER
MALI
GUATEMALA
HONDURAS
CARIBBEAN SEA
CHAD
SENEGAL
EL SALVADOR
UPPER VOLTA
SUDAN
GUINEA
DAHOMEY
NIGERIA
TRINIDAD
CARACAS
COSTA RICA
SIERRA LEONE
IVORY COAST
TOGO
GHANA
CAMEROUN
CENTRAL AFRICAN REPUBLIC
ETHIOPIA
PANAMA
COLOMBIA
VENEZUELA



Map 1 (foldout). Tropo systems or projects throughout the world (exclusive of Western Pacific area).

Map 2 (above). Tropo systems or projects in the Western Pacific area.

powers utilizing only solid-state devices, thus eliminating power tubes.

3. Additional development to reduce size and increase efficiency of klystron tubes. Utilization of traveling-wave tubes in power amplifiers where economical.

4. Investigations of various new types of power tubes.

5. Development of cheaper and lighter antennas, particularly for transportable and tactical tropo equipment, where weight, stowage, and ease of assembly in the field are paramount.

6. Investigation of alternatives to parabolic antennas.

7. Advances in further modularization of solid-state equipment, including use of integrated circuitry, to reduce size and weight and increase accessibility.

8. Further development of automatic fault indicator and performance monitoring equipment to facilitate maintenance and reduce maintenance personnel.

9. Development of a lightweight, easily transportable tropo terminal which better meets the tactical requirements of the military, and which can be set up in the field in a minimum of time. It must be easily transportable by all types of helicopters.

10. Development of a reliable, relatively cheap type or types of tropo equipment, conforming more or less to CCIR and CCITT recommendations, and capable of incorporation into national communications systems without degradation of the system.

11. Further studies and investigations of the tropo propagation mechanism to determine its exact nature and behavior, and thus afford designers a stronger base for development of improved equipment and perhaps better bandwidth for increased channel capacity.

The future of tropo

In the years ahead, what will be the position of tropo in the world of modern communications? Many persons may say that it will be more or less eliminated from the picture with the development of communication satellites. At the beginning of this article, tropo was termed a "gap filler" between radio facilities that provide short-range and long-range communications. This is a fairly accurate description. Probably tropo will be complementary to satellite systems rather than competitive with them. At this moment, several nations are reluctant to exercise their rights of satellite access for a number of reasons—stemming principally from economic considerations and a lack of operational control. Whatever satellite ground stations the various nations may establish, these communication centers must have feeder circuits to handle the domestic relays. Although various modes of effecting these feeder systems are possible, tropo scatter may be used in many cases where other means are impracticable and where reliability is required.

For commercial or common-carrier purposes, there appear to be many nations that are beginning to realize the value of tropo in places where the nature of the terrain presents obstacles and, in addition, where there is no need for circuit drop requirements at intermediate points between stations. As this is written, aside from the systems being installed (see Table I and maps), a number of others are under consideration or in the process of procurement. It can be assumed that during the next ten years there will be a steady, if limited, demand for common-carrier tropo systems. Where planning is done properly, these systems will provide for expansion of channels as requirements increase with economic progress and the population explosion.

The continued and additional employment of new and advanced types of tropo for military purposes will, of course, depend upon the international situation. As pointed out previously, the accent is now on development and procurement of light tactical types for utilization in jungle, guerrilla, or amphibious warfare, with some additions or upgrading of existing or planned strategic systems.

We can take steps to produce better tropo equipment by proceeding with the developments listed in the preceding section. To these might be added improved low-noise receivers, preamplifiers, and threshold extension devices.

There is no doubt that tropo is here to stay, and will occupy its own niche in the array of various types of communications systems. Under certain conditions its employment is now indispensable; the future may well add to these conditions to augment its value to later use in both worldwide and domestic communications. Its first 11 years of development have been impressive and fruitful. The further enhancement of tropo will depend upon the effort and vision of our engineers.

A number of firms, at considerable inconvenience, supplied valuable data and comment on various aspects of tropo scatter communications, including the equipment involved. The author is indebted to these companies, listed below, and extends to them his sincere appreciation for their generous assistance and cooperation.

American Telephone & Telegraph Company
Bell Telephone Company of Canada
Bendix Corporation
Collins Radio Company
Compagnie Francaise Thompson Houston
Compagnie Générale d'Electricite
General Electric Company
General Telephone & Electronics Corporation
International Aeradio Limited
International Telephone and Telegraph Corporation
Marconi Company Limited
Martin Company
N. V. Philips Gloeilampenfabrieken
Page Communication Engineers, Inc.
Pye Limited
Radio Corporation of America
Siemens & Halske, A.G.
Telecommunications Radioelectriques et Telephoniques
Western Electric Company
Westinghouse Electric Corporation

The author also desires to express his warm thanks and appreciation to certain members of the staff at Radio Engineering Laboratories who assisted in the compilation of data and in the preparation of illustrations and manuscript.

REFERENCES

1. Marconi, Guglielmo, "Radio communications by means of very short electric waves," *Proc. Roy. Inst. Gt. Brit.*, vol. 22, pp. 509–544, 1933; reprinted in *IRE Trans. on Antennas and Propagation*, vol. AP-5, pp. 90–99, Jan. 1957.

2. Clavier, A. G., and Altovsky, V., "Beyond-the-horizon 300-megacycle propagation tests in 1941," *Elec. Commun.*, vol. 33, June 1956.

3. Terman, F. E., *Radio Engineers' Handbook*, 1st ed. New York: McGraw-Hill, 1943, pp. 754–755.

4. Bullington, K., "Characteristics of beyond-the-horizon radio transmission," *Proc. IRE*, vol. 43, pp. 1175–1180, Oct. 1955.

5. Norton, K. A., "Propagation during World War II," *Proc. IRE*, vol. 50, pp. 698–704, May 1962.

6. Manning, L. A., "Radio propagation following World War II," *Proc. IRE*, vol. 50, pp. 709–711, May 1962.

7. Hull, R. E., "E. H. Armstrong participation in early tropo propagation studies," Unpublished memorandum to W. E. Yost, Jr., Dec. 16, 1964, Radio Engineering Laboratories.

8. Rice, P. L., Longley, A. G., Norton, K. A., and Barsis, A. P., "Transmission loss predictions for tropospheric communication circuits," Tech. Note 101, vols. 1 and 2, National Bureau of Standards, May 7, 1965.

9. Carroll, T. J., Lebatich, J. L., and Vergara, W. C., "VHF air–ground propagation far beyond the horizon and a new meteorological parameter," Bendix Radio Rept., Feb. 1961.

10. Carroll, T. J., "Mode theory as proof of partial internal reflections from the normal air," International Scientific Radio Union Program, May 1961, pp. 52–53.

11. Carroll, T. J., "Interpretation of Tukizi's calculations of twilight propagation," International Scientific Radio Union Program, Oct. 1961, p. 34.

12. Armand, N. A., Vvdenskii, B. A., *et al.*, "Long-range tropospheric of ultrashort radio waves," *Radiotekhn. i Elektron. (U.S.S.R.)*, vol. 6, June 1961, pp. 867–885.

13. Bullington, K., "Reflections from an exponential atmosphere," *Bell System Tech. J.*, vol. 42, pp. 2849–2868, Nov. 1963.

14. Medhurst, R. G., and Small, G. F., "An extended analysis of echo distortion in the FM transmission of frequency-division multiplex," *Proc. IEE (London)*, vol. 103, pt. B, pp. 190–198, Mar. 1956.

15. Booker, H. G., and de Bettencourt, J. T., "Theory of radio transmission by tropospheric scattering using very narrow beams," *Proc. IRE*, vol. 43, pp. 281–290, Mar. 1955.

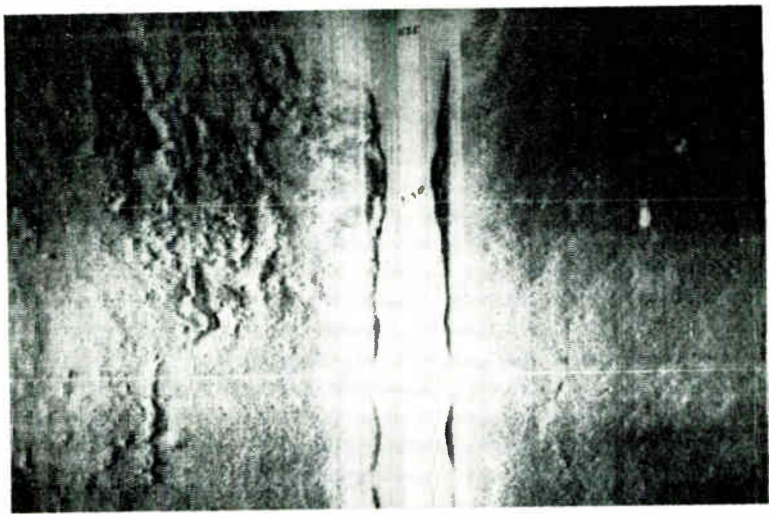
16. Chisholm, J. H., Rainville, L. P., Roche, J. F., and Root, H. G., "Measurements of the bandwidth of radio waves propagated by the troposphere beyond the horizon," *IRE Trans. on Antennas and Propagation*, vol. AP-6, pp. 377–378, Oct. 1958.

17. Von Baeyer, H. J., "Selective fading effects on UHF tropospheric scatter paths," *Proc. IRE*, vol. 47, pp. 2021–2022, Nov. 1959.

18. Shaft, P. D., "Information bandwidth of tropospheric scatter systems," *IRE Trans. on Communications Systems*, vol. CS-9, pp. 280–287, Sept. 1961.

19. Beach, C. D., and Trecker, J. M., "A method for predicting interchannel modulation due to multipath propagation in FM and PM tropospheric radio systems," *Bell System Tech. J.*, vol. 42, pp. 1–36, Jan. 1963.

20. Prescott, H., "A method for predicting the statistical distribution of bandwidth in tropospheric scatter links," *Proc. 7th Nat'l Conv. on Military Electronics*, Washington, D.C., 1963, pp. 281–285.



Seismic and sonic instrumentation for 'ocean engineering'

Seismic and sonic techniques are discussed for offshore petroleum prospecting, the search for sunken vessels, accurate charting of the ocean floor, and the prediction and warning of tsunami—*the giant tidal waves produced by submarine earthquakes and volcanic eruptions*

Gordon D. Friedlander *Staff Writer*

Since the successful launching of Sputnik in 1957, much of our attention has been directed skyward in viewing the wonders of our rapidly evolving technology for the conquest of outer space. More recently, our gaze has been diverted in the opposite direction in appreciation of the remarkable progress that is being made in the conquest of another frontier—inner space—the deep and heretofore impenetrable reaches below the oceanic vastness that constitutes 77 percent of the earth's surface area. In a series of three articles, the diverse areas of ocean engineering—the application of the oceanographic sciences—will be discussed in the significance of their broad implications for the future benefit of mankind.

Oceanography and ocean engineering—definition of terms

The field of oceanography, in its range of expanded activities during recent years, has necessitated an enlargement of its definitions and a compartmentation of its specialties. We may say generally that "ocean engineering" provides the tools for the application and exploitation of the principles of oceanography (the basic scientific knowledge itself) for human advantage and betterment.

Oceanography itself may be subdivided into four categories:

Biological oceanography—the study of animal and plant life within the sea, and the acquisition of knowledge pertaining to the habits and migration patterns of schools of food fish and large cetaceous mammals.

Chemical oceanography—the study of the chemical and mineral composition of the sea itself and the mineral and petroleum deposits beneath the ocean floor.

Geological oceanography—the study and mapping of the topographical features of the ocean bottom and the substrate beneath the ocean floor; the study of submarine crustal faults and the formation of submarine volcanoes.

(Above) Photographic reproduction of a bottom record, at a depth of more than 8000 feet, obtained from a towed sonar apparatus during the search for the lost nuclear-powered submarine U.S.S. Thresher. The white hairlines indicate the extent of successive 5-foot strips.

Physical oceanography—the study of ocean currents, water temperature and salinity, how water moves, the generation of waves, and the phenomena of the ocean–air interface.

Development of sonic devices

The basic fact in natural science that sound travels about 4½ times faster in water—and up to 18 times faster in rock layers—than it does in air has proved to be the key for unlocking many of the secrets of the ocean environment and the strata of the earth's crust immediately beneath the ocean floor.

During World War I, primitive "hydrophones" were developed for use in antisubmarine warfare. Operators on board destroyers or patrol craft could hear the engines and propellers of submerged vessels by means of these instruments. By turning such a listening device toward the origin of the sound, a rough approximation could be made of the relative bearing of an enemy submarine with respect to the monitoring ship.

The echo sounder. A later electronic application of the principle of sound transmission through water, and its backward reflection after striking a denser medium or the ocean floor, led to the development of the echo sounder, or Fathometer.® These devices have been used by vessels for many years for the avoidance of shoal water, in piloting and navigation, and as an aid in fixing a surface position when a good bathymetric chart of an area is available.

An echo sounder is really a form of sonar, which, by measuring the time difference between the transmission of an electronically generated sound wave and its return after striking the sea bottom, will give the necessary data for depth determination. To achieve an accurate reading, an average velocity of sound must be determined with respect to existing water temperature and salinity. The device consists of a transducer, located near the keel of the ship, which serves as both the transmitter and receiver of the acoustical signal. An oscillator, receiver, and amplifier generate and receive the electrical impulses to and from the transducer. A recorder is calibrated for direct reading in meters, fathoms, or feet. The frequency usually employed is in the low ultrasonic range (20 000–30 000 Hz). The readout may be a trace-type graphic display, or a direct dial that gives the instantaneous depth of water under the vessel's keel.

Instruments for hydrographic surveying and bottom searches

For the past 30 years, hydrographers have generally utilized the wide-cone echo sounder[1] for both large- and small-scale oceanographic bottom surveys. Although this equipment has done a fairly accurate plotting job for small-scale charting, it has been insufficiently accurate for medium- and large-scale mapping because of the relatively large area of coverage that is plotted as one point in a single sounding sweep. For example, if a survey unit were working in water in the range of one-mile depth, the wide-sweep echo sounder would pick up the shallowest point in a circle of one-mile diameter and record that depth as the center, or position of the echo sounder's transducer. In small-scale charting this error may not be significant, but at larger scales its magnitude increases.

Today, advanced oceanographic technology and equipment permit the observation of detailed topographic features of the ocean floor at great depths. By means of an ocean-bottom-scanning sonar, developed by Westinghouse, operating on the analogous principle of an optically viewed panorama that is illuminated by side lighting, a visible pattern of contrasting highlights and shadows caused by the irregular topography of the ocean bottom is acoustically reproduced.

Fig. 1. Isometric diagram of the sweep pattern of a towed, side-looking sonar for ocean-bottom sounding.

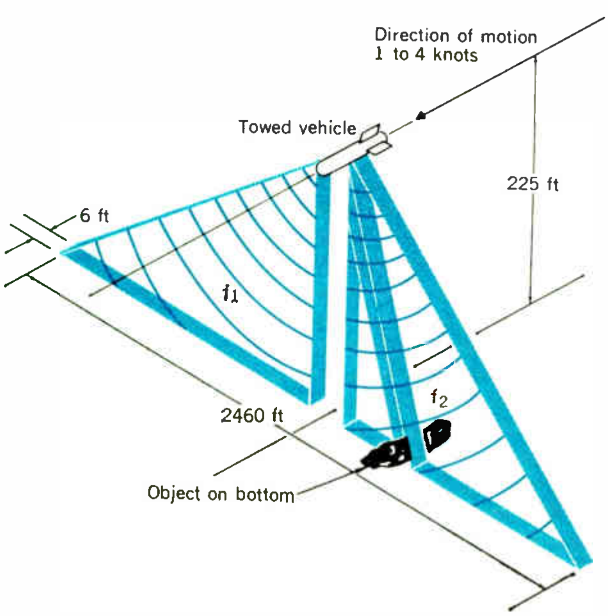
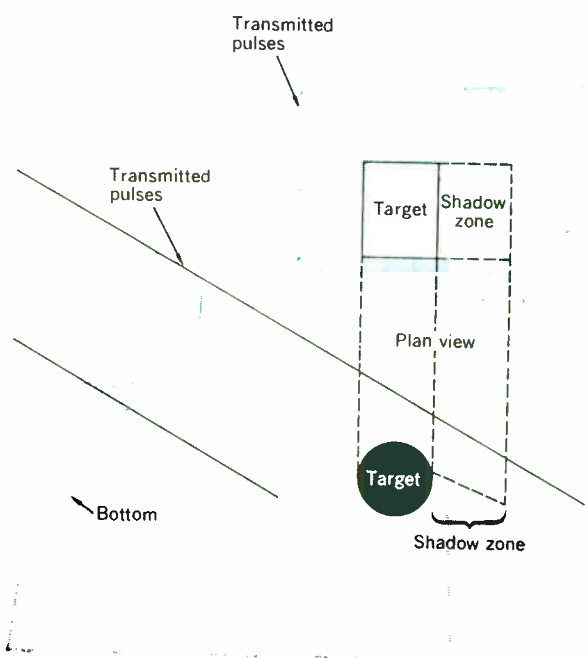


Fig. 2. Line-transmitting transducers on both sides of the underwater sonar vehicle transmit pulses of high-frequency acoustic energy to the sea bottom. A topographic irregularity or sunken object will vary the pulse pattern in the path of the sonar sweep.



In meeting the need for instrumentation that falls between the capabilities of long-range low-resolution sonar and short-range high-resolution optical techniques such as cameras or television, the side-looking sonar (Fig. 1) can adapt itself, through a large variety of equipment configurations, to a versatile range of oceanographic tasks.

Following the loss of the nuclear-powered submarine *U.S.S. Thresher* in April 1963, in 8000-foot-deep waters, Westinghouse developed and built a towed, unmanned submersible vehicle containing a side-looking sonar, that was capable—as shown in Fig. 1—of recording in one shot a path area of six-foot length, extending 1230 feet (2460 feet overall) from either side of the craft. When the vehicle was held at an altitude of 225 feet above the sea floor, many striking features of the bottom topography (see title illustration, page 101) were observed in sequential 5-foot strips. Note the "blind spot" in the center of the recording which corresponds to the area directly beneath the vehicle in Fig. 1.

Essentially, the ocean-bottom-scanning sonar comprises a towed underwater vehicle that contains two sonar devices, each of which has a sonar transmitter with transducer, a receiving transducer, and preamplifier. Also on board is the necessary equipment for accurate depth control. The remaining elements of the sonar system are installed in the surface vessel.

As the fan-shaped sonar beam energy pulses progress outward, they strike the sea bottom at an angle (Fig. 2). Each pulse scans the bottom over the interval of a given time period. Although the pulse energy is backscattered evenly in a smooth bottom condition, the pattern will vary if there is a topographic irregularity or a sunken object in the path of the sonar sweep. When this occurs, as illustrated in Fig. 2, there will be an intensification of the return energy signal that is followed by a "shadow zone." The return signal is translated and transcribed onto a recording chart whose paper speed trace is proportional to the speed of the survey ship (1 to 4 knots). This figure shows how line-transmitting transducers, which may be focused or unfocused, depending upon their application, can limit the isonified bottom area to a narrow strip. Optimum resolution in the fore and aft direction is obtained by focusing down to the narrowest possible strip area.

The precision, side-looking sonar allows a detailed inspection of the ocean floor down to extreme depths with a resolution that may vary from one inch to several feet. A selection of operational parameters for a specific application is necessary, however, since resolution performance varies inversely with the compatible range, or mapping width. The lateral resolution is determined by the pulse width, which is made as short as bandwidth considerations permit at the selected operating frequency.

The Fig. 3 block diagram shows how synchronous pulses, originating at the shipboard console, actuate the transmitters in the towed vehicle by means of the connecting cable circuit. Here, high-frequency short-duration pulses are radiated downward by the transducers. Display sweeps at the receiver are synchronized with the transmitted pulse. Signals that are received by the hydrophones are amplified in the preamplifiers and then transmitted back to the shipboard receiver unit. Time-varied gain (TVG) control circuits remove the predictable amplitude variations caused by irregularities in the topographical geometry of the sea floor. Further amplification occurs and a sunken object is represented at the display as a voltage whose amplitude is essentially independent of range. Additional automatic gain control (AGC) is also included in the receiver circuits to avoid excessive saturation of the display that may be caused by unpredictable variations in the received signal level.

The console display represents the received signals from each transmitted pulse as a line of varying intensity or density. The process, repeated sequentially as the sonar vehicle proceeds, converts the signals from successive transmissions into a stored image with the three-dimensional quality of light and shadow, outlining objects in a manner that will permit their identification. The selected

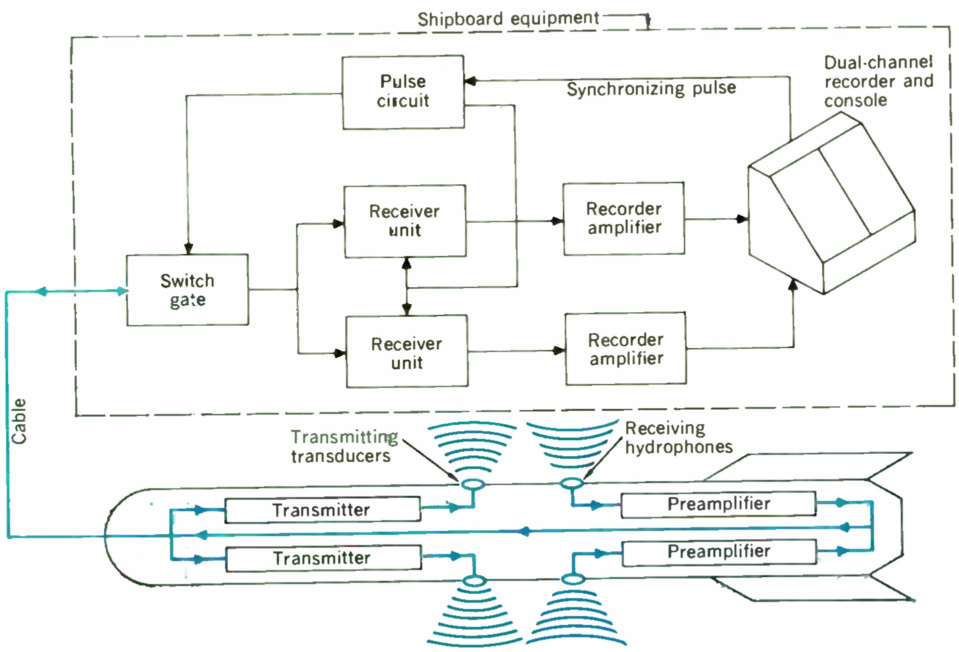


Fig. 3. Block diagram of shipboard and towed vehicle sonar gear, indicating how synchronous pulses originate at the console to trigger the outboard transmitters, from which short-duration HF pulses are radiated downward by the transducers.

Fig. 4. Diagram, based on data from a recent dive made by Cousteau's "diving saucer," showing the distribution and numbers of myctophids and physonect siphonophores observed during the downward migration of scattering layers. The inverse time scale (calibrated in half-hour Greenwich time increments) is greatly compressed. Depth course of the saucer was established by its upward-directed echo sounder "pinging" on the sea–air interface.

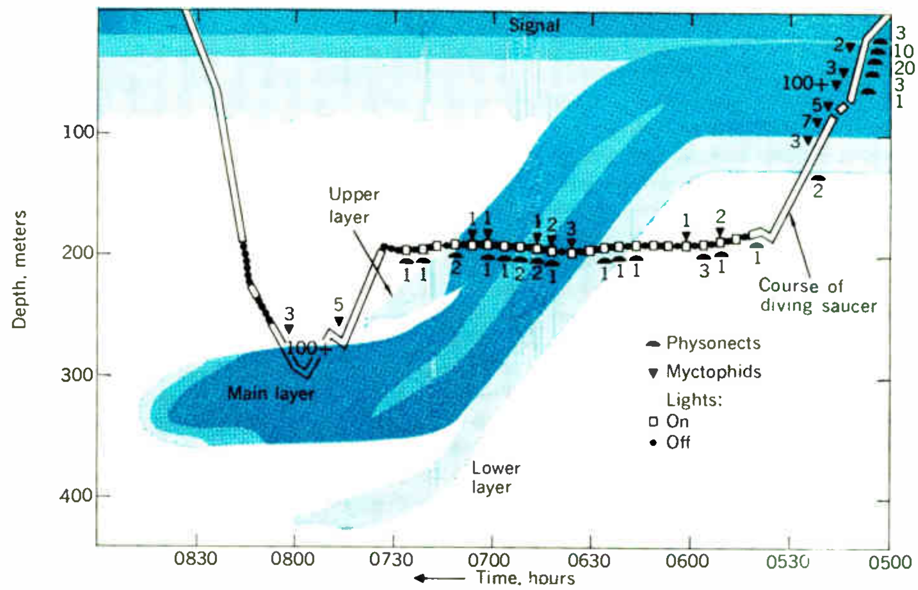
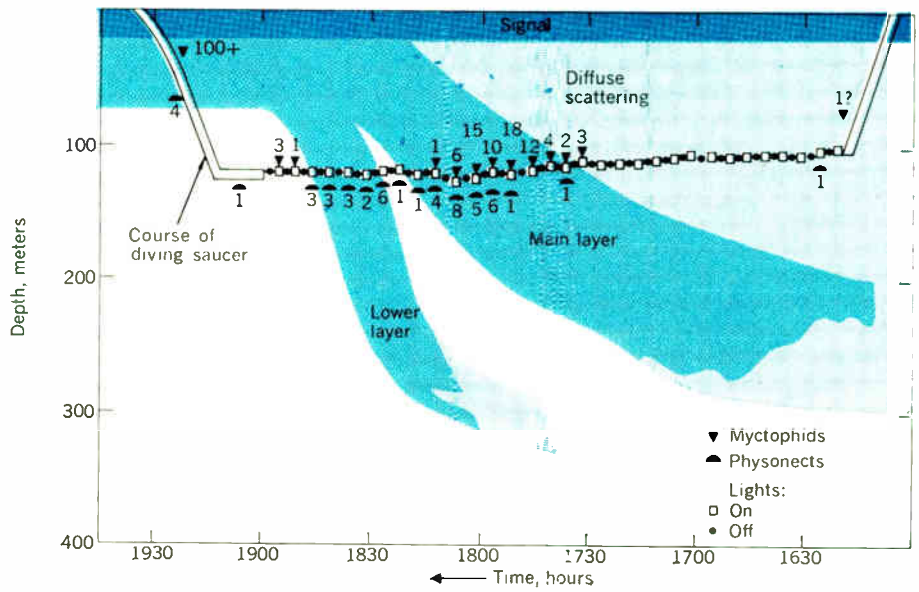


Fig. 5. Diagram indicating the distribution and numbers of myctophids and physonect siphonophores observed from the diving saucer during an upward migration of the scattering layers. All of the other parameters are the same as shown in Fig. 4.



display may be either a paper recorder or a cathode-ray tube picture simulator.

Recent hull-mounted configurations of this device were incorporated into the U.S. Navy's bathyscaphe *Trieste II* for use in further search operations for *Thresher* and for oceanographic research down to depths of 20 000 feet.

Narrow-beam echo sounder. By means of newly developed, sophisticated sonar and echo-sounding apparatus, including several effective narrow-beam configurations that are comparable to medium-scale stereo pairs, the task of plotting accurate bottom topography has been greatly simplified. There is presently available a tested conversion of the standard 60-degree-cone-width echo sounder that reduces the cone to 6 degrees to permit a large reduction of the previous plotting errors. By the use of this instrument, the diameter of the single-point plotted area will be reduced to 10 percent of the water depth. Thus, if the plot is taken in water of an average depth of 1000 feet, the plotted area will be 100 square feet.

The scattering layers

The term *scattering layers* in oceanography refers to a phenomenon, caused by dense layers of minute marine animals or organisms, that has the effect of deflecting and dispersing sound waves to produce false echoes. Recordings by sonic sounding devices indicate that the scattering organisms, in medium-depth waters (up to 200 fathoms), are arranged in approximately horizontal layers in the water. In such waters, the layers are usually well above the bottom.

Deep scattering layers. In deep water, one or more well-defined scattering layers are generally present. They are readily detected by sonar equipment that is capable of scanning the sound spectrum from 1–60 kHz. Each layer, depending upon its composition and distribution of myctophid fish and physonect siphonophores,[2] will demonstrate maximum scattering effect at different frequencies.

Generally, the deep water layers "migrate" vertically in apparent response to changes in natural or artificial illumination. Figures 4 and 5 indicate graphic cross sections of the distribution and numbers of myctophids and physonect siphonophores observed during a downward migration of scattering layers in the course of recent dives made by Captain Jacques-Yves Cousteau's "diving saucer" in water approximately 1300 meters deep off the coast of Baja California. The most pronounced downward migration pattern occurs during the diurnal cycles. The layers rise at night, sometimes to the surface, and may descend to great depths during the daylight hours. The common range of daytime depths is 100–400 fathoms (183–732 meters). The upward migration is modified by bright moonlight at night and the downward drift has been observed to be modified by heavy local cloud cover during daylight hours.

The scattering layers were discovered during World War II, when the parallel movements of ubiquitous stratified zones of sonic reverberation, producing false echoes, were observed during the pursuit of enemy submarines by surface vessels. Since 1945, intensive efforts have been undertaken to identify and classify the marine creatures and organisms responsible for the phenomenon, and to establish regional oceanic patterns of areas subject to scatterers.

Although the scattering layers are a disruptive factor and a general nuisance in the accurate charting of bottom topography and in the search for sunken objects, they have, nevertheless, introduced a valuable detection technique for plotting the migration patterns of large schools of food fish.

Thermal refractions. There are basically three variables—temperature, salinity, and density—that affect the transmission and reflection of sonic waves in the water medium. The thermal gradients in deep water generally descend quite rapidly at depths beyond 25 fathoms. And, as is the case in air, the colder and denser layers tend to fall in a convection pattern. In these strata the velocity of sound is retarded.

We have all observed the refraction and deflection of light when it strikes a denser medium such as glass. An analogous effect occurs when sonic waves strike the denser layers of cold water. Thus the sonar transmission and return signals are distorted or "bent" in this situation and an error factor is introduced into the ranging accuracy and the display readout of the graphic sonar plot. The complexity of the thermal currents in the ocean are further compounded by temperature inversion phenomena that are analogously comparable to the inversion conditions and stagnation common to the atmosphere.

Other performance factors. Numerous factors influence the performance efficiency of an active sonar system. Principal among these, because of certain unpredictable variables, is the propagation loss in the sea between the transmitting transducer and the target. This loss is a function of the sound frequency and the refracting thermal gradients already discussed.

Another important performance element is the *target strength*. In order to detect an echo from a target, the reflecting power, or target strength, must be greater than the other extraneous signals, caused either by sonar self-noise or reverberation, that are received. Self-noise is noise that is produced by the motion of the vessel carrying the sonic gear, and the major portion of this noise is generated by cavitation at high speeds and by water turbulence in the vicinity of the transducer. Self-noise is

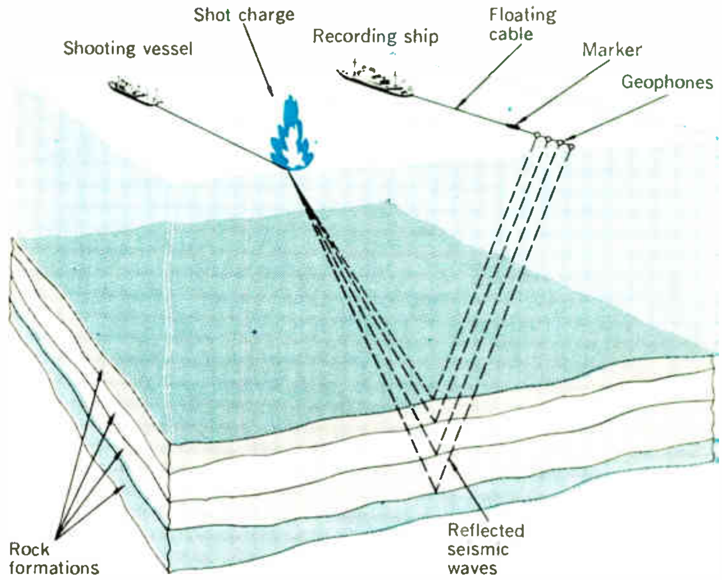


Fig. 6. Diagram showing the "reflection shooting" technique used for the analysis of shock waves from submarine rock strata that may contain oil deposits.

sharply augmented by an increase in the speed of the equipment-carrying vessel, particularly above the critical velocity in which laminar water flow breaks into the turbulence produced by propeller cavitation and bow wave action.

Reverberation is the combination of all echoes returned to an active sonar system by the sea itself. Such noises include those produced by the sea–air and sea–bottom interfaces, entrained air bubbles, the scattering layers, and powerful ocean current interactions. An experienced sonar operator can aurally detect reverberation interference by the characteristic "quavering ring" that occurs at the instant that the sound pulse is transmitted.

Sonic Doppler effect. By means of the familiar Doppler effect, in which there is a change in frequency of a sonar echo, the directional movement of a target may be determined. Since the sound scatterers that are responsible for reverberation are of relatively fixed frequency, a down- or up-shift in frequency, caused by the Doppler shift, can generally be detected in the presence of reverberation.

As of now, neither science nor electronic technology has been able to overcome the unpredictable barriers to accurate sonar and echo sounding; however, far-reaching investigation and data accumulation, aided by the computer and advanced instrumentation, are removing any of the baffling and frustrating aspects of these natural phenomena.

Origins of offshore seismic instrumentation

For many years scientists have been able to detect and record, through the use of an ultrasensitive instrument called the seismograph, ground-transmitted seismic disturbances of a terrestrial origin that occur thousands of miles away. And we have known for more than a century that seismic disturbances and submarine volcanic eruptions occur along predictable areas of the ocean floor to produce gigantic tidal waves that travel at phenomenal speed to smash inland along the coastal margins of continents and islands many thousands of miles distant from the point of origin of the disturbance.

About 25 years ago, extensive R & D work was undertaken to design and construct accurate "listening" instrumentation that could detect and record the intensity and geographical location of these submarine disturbances. But with our active involvement in World War II, research effort was accelerated in the military application of these devices for the detection and location of enemy surface vessels and submarines.

Passive sonar. The forerunner of many of our present-day, sophisticated seismic and listening devices is to be found in the concept of *passive sonar*. Essentially, this encompasses an underwater acoustic system that achieves its detection capability by merely listening for noise radiated from a possible target or point of origin. A passive system consists of a highly directional and trainable array of transducers, electronic amplifiers, and a display system, which is usually aural in shipborne configurations. We are all familiar with the characteristic "pinging" sound of a sonar contact in military usage. This is actually the "aural display" of sonar frequencies that lie within the audible range. These are the optimum frequencies in reference to the spectrum of sound wavelengths emitted by the propulsion machinery and impulse beats produced by the turning of a ship's screws.

The effectiveness of passive sonar in all its applications, however, depends upon the magnitude of the radiated sound from a target, the propagation loss between the target and the receiver, and the background noise from scatterers, self-noise, and reverberation.

A pioneering effort. It is interesting that toward the conclusion of World War II, the pioneering scientific effort in recording seismic disturbances of oceanic origin was concurrent with the applied engineering attempts to utilize this principle in the offshore exploration for petroleum deposits, and a reliable passive sonar system for the detection of enemy submarines during wartime.

One of the earliest offshore seismic exploratory operations was conducted in 1944 in the Gulf of Mexico. Six land-type geophones were bottom-planted from a small open boat. A separate, two-conductor cable led from each geophone to a recording vessel anchored in open waters. The geophones comprised a mass of known weight, supported by a compression spring that was constrained to move vertically. An electromagnetic transducer transformed the shock-wave-produced motion of the spring and weight into varying voltages that were a function of the degree of mechanical motion. Special amplifiers were designed for small size, large gain, and freedom from background noise. They incorporated bandpass filters, with adjustable low- and high-frequency cutoff. Automatic gain control provided usable signals from the arrival of the initial amplitude waves until the signal strength dropped below the level of the background noise of the sea. Unfortunately, however, land-type geophones, although they respond well to the vibratory motions of the earth, react readily to wave agitation, and are, therefore, practically useless in most marine applications.

In calm weather, data recording from eight different locational areas in the course of a day was considered to be a notable achievement, but on other occasions, three-foot-high seas halted geophone emplacement.

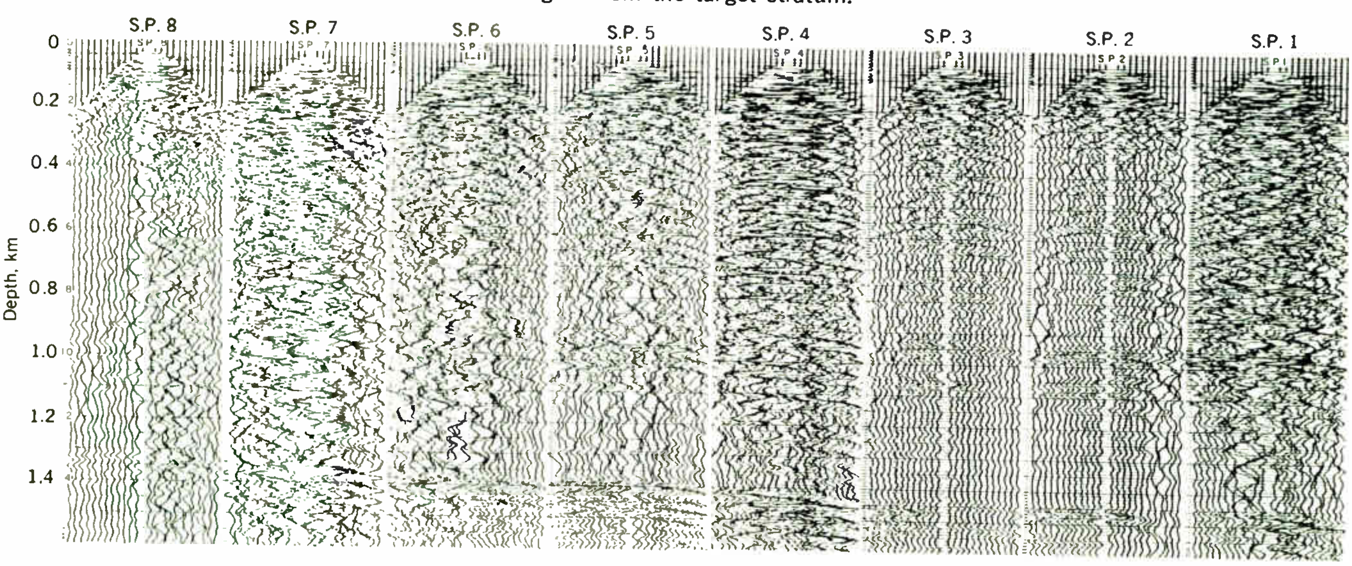
Submarine seismic reflections pass through the earth–water interface to travel upward through the water, producing acoustic pressure waves in this medium. Accurate marine seismic detectors must respond as much as possible to these acoustic waves and as little as possible to the motion that results from wave action or cable vibration.

Present-day seismic technology

During the past 22 years, tremendous advances have been made in the improvement of instrumentation and techniques. The volume of data production and recording quality have greatly increased, particularly since the introduction of computer systems and the digital "filtering" of extraneous noise from seismograms.

Today's search for oil. The worldwide demand for petroleum and natural gas, since 1945, has multiplied

Fig. 7. A typical "raw" field seismogram in which noise clutter effectively masks the signal from the target stratum.



at an incredible rate. The exhaustion of many oil fields throughout the United States and abroad has necessitated the exploration for and recovery of these vital fossil fuels from the marine deposits that exist under the waters of the continental shelves. Because of the tremendous costs of "wildcatting" for these underwater deposits by conventional methods, new techniques for eliminating much of the risk and wasted investment of a dry boring had to be found. Thus a special application of the science of seismology was developed to record the structure of the undersea rock strata for the revelation of the oil-bearing shale and limestone layers.

Here, and in subsequent sections of this article, the reader is invited to don his mental SCUBA gear for imaginary dives into the "wild blue under" in the vicarious search for the riches of the sea.

Seismic prospecting on land is a relatively simple procedure because of the stationary and stable terrestrial "platform" upon which operations are conducted. But searching for a submarine oil field that may be only a mile wide and under 100 fathoms of water, silt, and rock is comparable to hunting for the needle in a haystack.

Many major petroleum companies engage geophysical research firms to do the actual offshore prospecting. As shown in Fig. 6, these marine technicians conduct the survey from special-purpose vessels that are equipped with very precise navigational instrumentation, sonic gear, cable-detection systems, and recording devices.

Reflection shooting. One method of conducting these surveys, in which accurate profiles of the rock structure below the continental shelf are charted, is by *reflection shooting*. In this technique (Fig. 6), the two survey craft—the "shooting" vessel and the recording ship—steam ahead on parallel courses at 4- to 6-knot speeds. During the course of a run, the shooting vessel maintains a constant position of about 2000 feet ahead and 250 feet to port or starboard of its companion craft.

The recording vessel tows a submerged cable-detector streamer, almost a mile long, containing the sensitive geophones that pick up the reflected shock waves. To generate these required seismic waves, the shooting vessel detonates charges of ammonium nitrate that are packed in special canisters in quantities up to 50 pounds. These explosives are set off singly, at timed intervals of one to two minutes, at depths ranging from five to ten feet beneath the sea.

The shock waves generated by the explosions pass downward through the various media at known velocities (about 5000 ft/s in water, 20 000 ft/s through granite, and 18 000 ft/s through potentially oil-bearing limestone). The reflected seismic waves return to the surface of the water and are picked up, over a period of several seconds, by the sensitive tape-recording instruments aboard the second vessel. By measuring the time taken for the waves to travel from the point of explosion to the various rock layers and back to the instrumentation, it is eventually possible for seismic technicians to determine the composition of the rock structure for a depth of two to three miles below the sea's floor. A new and revolutionary process, which will be discussed later in this article, now makes it possible to interpret "raw" seismic data by recording the signals in digital form that is suitable for the computer "cleansing" of all forms of signal distortion.

Figure 7 illustrates a complicated pattern of oscillating lines, reproduced on photographic paper, that represents a raw seismogram of the geological strata beneath the sea bottom. Obviously, the "noise clutter" in this diagram effectively masks the desired signal from the target stratum, thereby making it almost impossible to interpret the substrata conditions.

Refraction shooting. The second technique, *refraction shooting*, in which two vessels are also involved, is used for mapping a layer whose elastic velocity is higher than the strata above it. In this method, large charges of explosives are detonated at eight-minute intervals as the recording ship moves at a constant speed from a zero datum to a point several miles distant from the shooting vessel (which remains stationary). By measuring the refraction to which the sound waves are subjected as they pass through media of varying densities, a ready cross-check of data gathered by reflection shooting is made available. Hence, both shooting techniques are generally used during a geophysical prospecting survey.

In both methods the explosive charge is dropped overside by the shooting crew. After it has floated beyond a possible danger point, the recording cycle for each shot is started on the shooting boat by a technician who depresses, and holds down, two electronic switches until the charge is detonated by radio relay from the recording craft. If the technician should feel, at any time prior to firing, that it is unsafe to proceed with the shot, he may stop the cycle in a "fail-safe" action by releasing the switches to break the circuit.

Each shot is carefully pinpointed by a marine surveyor in the chart room of the recording vessel. At the instant the charge is fired, he must note all positional readings and the exact time so that the precise location of the shot can be plotted on the bathymetric and hydrographic charts.

Digital filtering

That faithful servant of science and technology, the digital computer, has succeeded in providing what geophysicists have long felt would be invaluable assistance to the petroleum prospector in finding:

1. Stratigraphic traps (natural reservoirs in which oil or gas may be confined because of changes in the porosity and permeability of the rock strata rather than as a result of their structural and geological configuration).
2. Oil reservoirs that exist within very complex geological and mineral deposits.
3. Stratigraphic traps that may be situated close to the sea bottom–water interface.

In filtering analog data, the original waveform may be distorted but digits cannot be deformed. Thus the theory behind this unique digital application is that the extreme versatility of digital processing in speed and accuracy would permit the filtering of seismic signals, either by the elimination or attenuation of all types of ambient noise and interference, without distorting the prime signal.

Functional operation. In essence, analog signals representing the amplified output of the geophones, strung along the cable-detector streamer in the water, are digitized by analog–digital converters, and the digital output is recorded on magnetic tape. Next, the tapes are fed into digital computers that process the data by filtering and cross-correlation before it is reconverted into analog form for ultimate presentation.

The most important element in the optimum use of digital equipment has been the development of efficient software or programs (see Fig. 8) for the elimination of unwanted noise. Such programs instruct the computer how to apply the various sequential processes for the removal of unwanted data, which include:

1. Instrumentation effects that interfere with the recognition of a signal peculiar to a known geological structure.
2. Detonation noise from the seismic explosive charge.
3. Air–water interface effects that may obscure the prime signal.
4. False echoes and reverberations.

As shown in Fig. 8, several filtering processes, depending upon the variety of noises encountered, may be necessary to produce a "clean" final seismogram.

Sequential filtering—a field example. Figure 9 illustrates the end product in the processing of the raw seismogram (shown in Fig. 7). In Fig. 9, the cluttering noise from the raw seismic data, gathered in an experimental line shot between two existing petroleum deposits, was removed by digital filtering. The problem in this example was to trace and map the uptilt termination of the Miocene sand layer, which was about 50 feet thick at the 4000–5000-ft levels. The distance separating each of the eight shot points (S.P. 1–S.P. 8) in the line was about 1300 feet.

The field experiment and the study of existing geophysical data indicated obstacles to prospecting such as sharp sonic velocity contrasts and surface (sea bottom–water interface) and near-surface materials that varied from shale to sand to gravel. Further, the individual problems created a seismic interference pattern that predominated over the target signals that were sought.

Note that the original seismogram, Fig. 7, and the final filtered recording, Fig. 9, are analog reconversions of the digital data.

The original recording was so cluttered with noise that nothing definite could be determined about the characteristics of the subsurface strata. However, a noticeable improvement of the recording was apparent after digital processing through an intermediate series of data-enhancement procedures. But a noise clutter and a deficiency in line resolution still persisted.

The seismogram quality was further improved after a pass through a digital velocity-filtering process in which extraneous noise was removed. Nevertheless, the noise-free data still afforded inadequate line resolution to reveal the elusive sand bed changes that are indicative of a stratigraphic trap.

The final deconvolution pass in the filtering process,

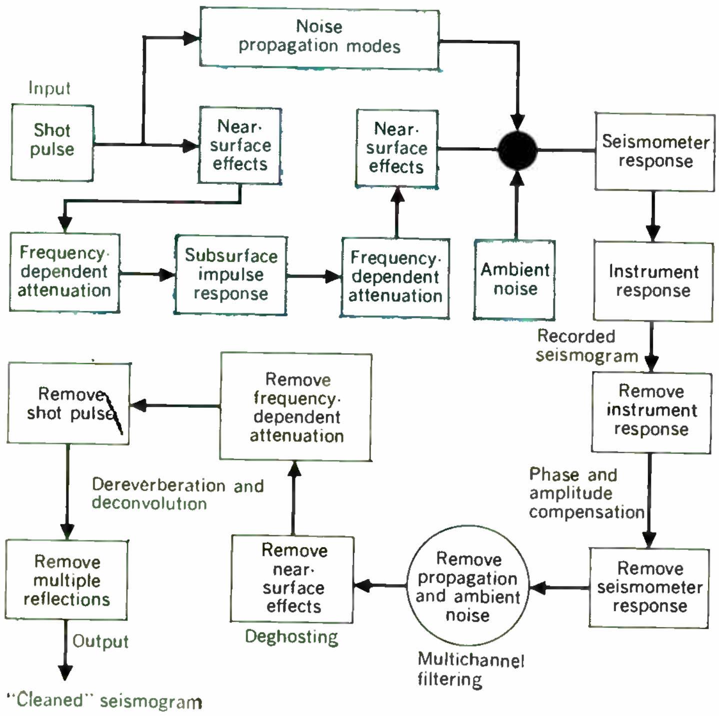


Fig. 8. Block diagram showing a typical computer program for the elimination of unwanted noise in a raw seismogram.

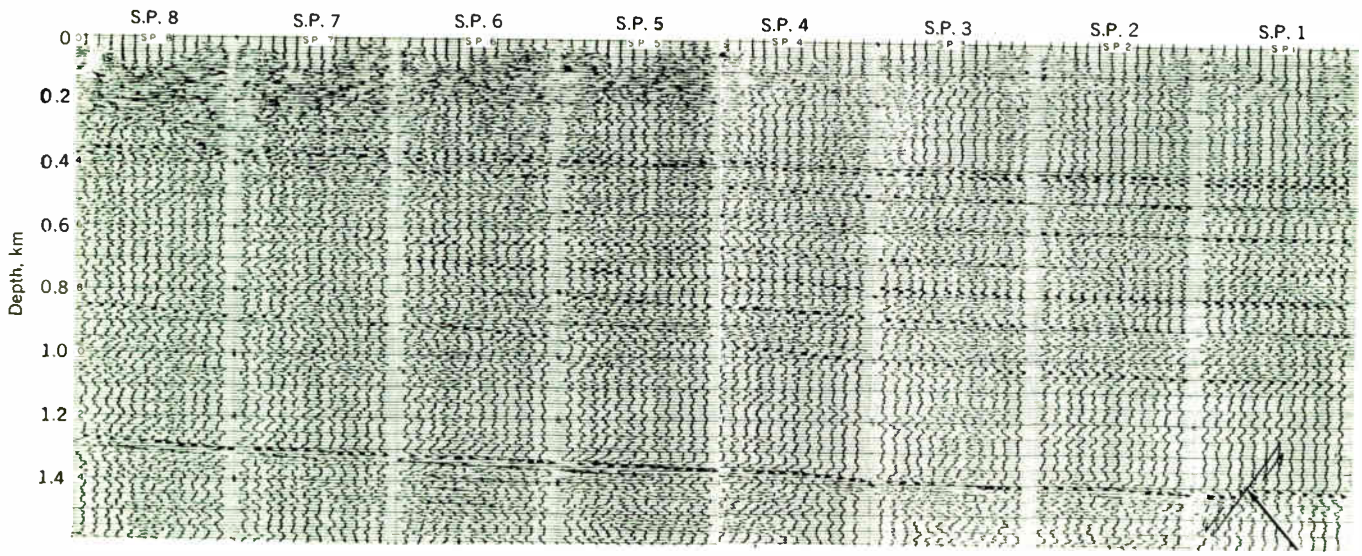


Fig. 9. The final "clean" seismogram is obtained by a set of data-enhancement techniques in the digital-filtering process that include velocity filtering, and elimination of seismic shot pulse and subsurface reverberation. Note minor geological fault indicated by arrow under S.P. 1.

however, produces the sharp resolution shown in Fig. 9. Here, the effects of the seismic explosion pulse and the substrata reverberations have been removed, and black, triangular dots have been added to stress the sand "pinchouts." In this final form, it is possible to trace the sand layer to its point of fadeout. Note the minor rock fault, indicated by the arrow under S.P. 1.

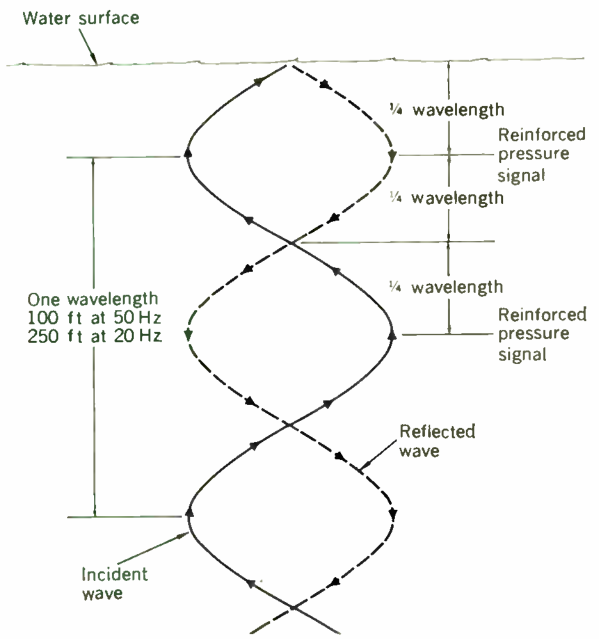
The cable-detector systems

One of the most important electronic components in the array of modern seismic and sonic instrumentation is the cable-detector system. The seismic marine cable may often be a mile in length; it must electrically connect each geophone, or detector, to the receiving instruments, and it must physically support and transport the detectors at the proper depth in the line of progress through the water.

Waves, passing over detectors, cause pressure variations, at frequencies under 2 Hz, on the sensitive devices. Thus detector systems that respond the least to wave-generated noise are the most useful and economical for use in bad-weather operations.

Acoustic pressure waves (reflections) travel upward toward the surface, where, because of the great density differential at the air–water interface, they are reflected downward (Fig. 10). Since the acoustic waves are in the denser medium of the interface, a phase reversal occurs. Therefore, it is apparent that acoustic pressures are at a minimum at the water surface, and that surface-reflected downward pressures will interfere with upward-traveling pressure waves except when the detectors are at an optimum depth. Thus pressure-actuated detectors must be at about one-quarter wavelength—or odd multiples thereof—below the surface for best reception. Record quality improves dramatically when phones are set at depths ranging from 40 to 60 feet beneath the surface.

Fig. 10. Diagram of acoustic pressure waves, indicating the nodes and points of maximum amplitude occurring during the upward and downward course of the incident and reflected wave patterns.



Cable buoyancy, length, and strength. Since the cable-detector systems are towed behind a recording craft, they should have about the same density as seawater to achieve a neutral buoyancy (neither sinking to the bottom nor rising to the surface). The seismic cables are available in two general types to attain this necessary buoyancy: those fabricated with a low-density foamed-plastic sheath, and those using light oil as a filling material within a tube. The former cable configuration carries its detectors external to the sheath and connected by molded takeouts, while in the latter system, the detectors are contained within the streamlined covering.

The drag involved in towing the foamed-plastic sheath cable is 2–3 times greater than that required for the oil-filled type. The majority of 24-channel marine cables range from 5000 to 8000 feet in length. Towing the smooth, oil-filled streamer—whose ultimate breaking strength is about 8000 pounds in tension—requires 2000–3500 pounds of traction at a 6-knot speed. The foamed-plastic sheath cable, of 16 000 pounds ultimate strength, may require 6000 pounds of tractive effort at the same towing speed.

Surface-reference system. One of the earliest cable streamers, of the foamed-plastic sheath type, was the surface-reference system [Fig. 11(A)] that was first used extensively in 1956. It may be described as a configuration of buoyant cables bearing detectors whose weight, when the streamer is not being towed, causes the cable in the vicinity of the detectors to sink to a depth that is suitable for recording. Since this cable system tolerates widespread density variations, it is particularly adaptable for use in traversing estuaries whose salinity may range from full salt to fresh water. Also, the system may be towed through waters as shallow as ten feet in depth.

The principal disadvantages, however, are

1. The cable must be intermittently stopped during its progress through the water for periods of 30–60 seconds to permit the detectors to sink to the proper recording level.

2. Those portions of a surface-reference cable that remain on the surface of the water are subject to wave agitation; hence the noise level in the system is dependent, to some extent, upon sea and weather conditions.

Bottom-reference system. Bottom-reference cables [Fig. 11(B)] are those that maintain their detectors at an approximately constant level above the sea bottom by means of a combination of cable buoyancy and localized "sinker" weights that are in contact with the sea floor during towing and recording operations. Since the cables are completely submerged, they are subject to very little noise interference and, therefore, the configuration is practically unaffected by sea surface and weather conditions. In addition, the system is virtually free from current-induced drift, or leeway errors, from the charted line of exploration.

Rapid seismic charting can be achieved with bottom-reference cables because they attain a terminal "quiescence" within 10 to 15 seconds from the time that forward towing progress ceases. Further, the record quality is significantly improved when the geophones are maintained at a level a few feet above the bottom.

The practical depth limit for efficient operation appears to be about 500 feet. The main disadvantage of the system is the large towing force required to move the weighted configuration.

Submerged-buoyant system. "Submerged-buoyant" denotes a complex marine cable system that operates by towing (preferably) an oil-filled cable sheath, with integrally contained detectors, at a constant depth that is controlled by surface buoys and special, streamlined wire lines [Fig. 11(C)] spaced at predetermined intervals. The salient advantage of the technique is that it offers a low noise level combined with good record quality under almost all sea surface conditions. This may be offset, however, by the fact that the detector cable must be at the same density as the surrounding water. Thus, under conditions in which water salinity varies over the course of a run (as will happen in the vicinity of river mouths), an inordinate amount of operational time will be required in adjusting the cable density.

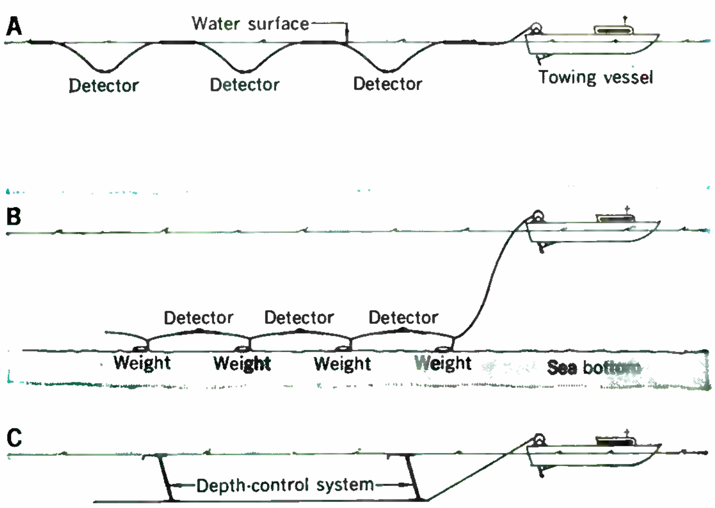
Present-day marine geophones. Marine geophone technology has advanced a long way in its evolution from the rather primitive land device that was adapted for sea use back in 1944. Today's pressure-responsive detectors of seismic acoustic waves fall into three general categories: *piezoelectric*, *electrodynamic*, and *magnetostrictive*.

Piezoelectric devices generate electric signals, in proportion to the strength of the seismic acoustical reflections, by means of the pressure excitation of barium titanate or lead zirconate crystals that are stressed in bending, torsion, or ring compression by the acoustical waves. These detectors are inherently high-impedance, capacitive devices that require transformer coupling to the transmission line leading to the recording instruments.

Electrodynamic pressure-sensing detectors usually involve diaphragms that drive coils in a magnetic field. Two such units, arranged back to back (astatically), reduce motion sensitivity to an acceptable level. Essentially, these are devices adapted from land-type geophones.

Magnetostriction devices utilize the characteristic of certain materials to vary their magnetic permeability under physical stress. Such materials, permanently magnetized and equipped with a flux, linking a winding—such as a toroid around a cylinder—constitute effective pressure transducers. The response is proportional to frequency and requires integrating filters to yield a conventional data display.

Fig. 11. A—Profile indicating a typical surface-reference cable-detector configuration, first used extensively in 1956. B—Bottom-reference cables maintain detectors at a constant level above the sea bottom by a combination of cable buoyancy and bottom-contact weights. C—Submerged-buoyant cable system maintains a constant level of submergence by means of a complex system of buoys and a depth-control system.



Some new techniques for seismic charting

Modern oceanographic survey vessels utilize many of the basic principles of seismic prospecting in gathering comprehensive data concerning the topography and geology of the ocean floor. But some of the larger craft, such as the U.S. Naval Oceanographic Office's 2500-ton *Silas Bent*, are equipped with the latest and most sophisticated instrumentation for recording bottom-tracking bathymetry; gravity, magnetic, and seismic data; and surface temperature while the ship is under way. Anchored on station, the vessel's electronic gear can record acoustical velocities; ambient light at various depth levels; and water currents, depths, salinity, and temperatures. In addition, digital computer modules, complete with stored logic, a random-access memory, and analog–digital–analog converters, provide an almost on-line data interpretation capability while the vessel is in under way mode at speeds up to 15 knots.

The instrument package, in brief. The electronic instrumentation and sensors for recording pressure, temperature, ambient light, current, and salinity will be discussed in greater depth (no pun intended) in a subsequent article in connection with the *Sealab II* experiment.

Included in the survey ship's instrumentation are

1. *Irradiance meters* (for ambient light)—precision hydrophotometers that measure directly the attenuation coefficient of a self-generated light beam, passing through water over a one-meter path length. The basic configuration contains a depth transducer and sensors.

2. *Vibrotron*(s)® (for recording water pressure down to 1000-meter depth)—vibrating wire transducers. A digital reading may be obtained by means of a frequency counter.

3. *Savonius current meter*—a rotor-type instrument for measuring the speed and direction of subsurface water flow, particularly at a low-speed threshold (less than 0.05 knot). This meter has a pulse output suitable for long-distance transmission.

4. *Roberts current meter*—an impeller-driven device that aligns itself with the current by means of a vertical fin to sense both speed and direction of flow. Unlike the Savonius rotor, this meter has no upper speed limit.

5. *Bathythermograph*—a readout or recording instrument, which, when towed, gives a continuous profile measurement of sea temperature vs. pressure (depth). A three-conductor, watertight cable, with an internal strain member, connects the readout with the sensor housing. The temperature sensor may be either a platinum resistance thermometer or a thermistor temperature probe. The depth sensor is a pressure potentiometer type.

Salinity is usually gauged by an electrical conductivity cell, using a precise standard of seawater salinity as a control reference.

Arcing, sparking, and air guns. The marine surveyor's present arsenal of seismic "weaponry" has rendered the conventional explosive charges, used in petroleum prospecting, somewhat obsolescent. In sea-bottom mapping, the oceanographer generally uses a "sparker"—a high-voltage static charge released across a spark gap—to produce the necessary acoustical waves. This device, which is much safer and easier to use than explosives, is especially adaptable for seismic profiling (Fig. 12). Depending upon acoustical frequency and sparking intervals, penetration data can be obtained for depths up to two kilometers. The Fig. 12 profile, taken off the coast of northern California, clearly shows the acoustical discontinuity within sedimentary rock layers. A section of the famous San Andreas fault, responsible for most of California's earthquakes, may be clearly seen at the right-hand end of the profile.

One of the limitations of the sparker, however, is the time interval required to build up a static charge in the capacitors. To overcome this disadvantage, a new technique, employing an "air gun," has been introduced to speed up acoustical mapping in geophysical surveys. Jets of highly compressed air (1500–2000 psi), released at frequent intervals, form bubbles that burst with explosive force upon reaching the surface of the water to generate acoustical waves that simulate the detonation of conventional ammonium nitrate charges. The air gun permits recordings to be run at 8–9 knots.

Tidal waves and 'distant early warning'

The giant tidal waves (called "tsunami" by the Japanese) produced by submarine earthquakes and volcanic eruptions, and common to the Pacific Ocean, have long been a source of danger, death, and destruction to people and property in the oceanic archipelagos and the coastal areas of the Asian and North American continents.

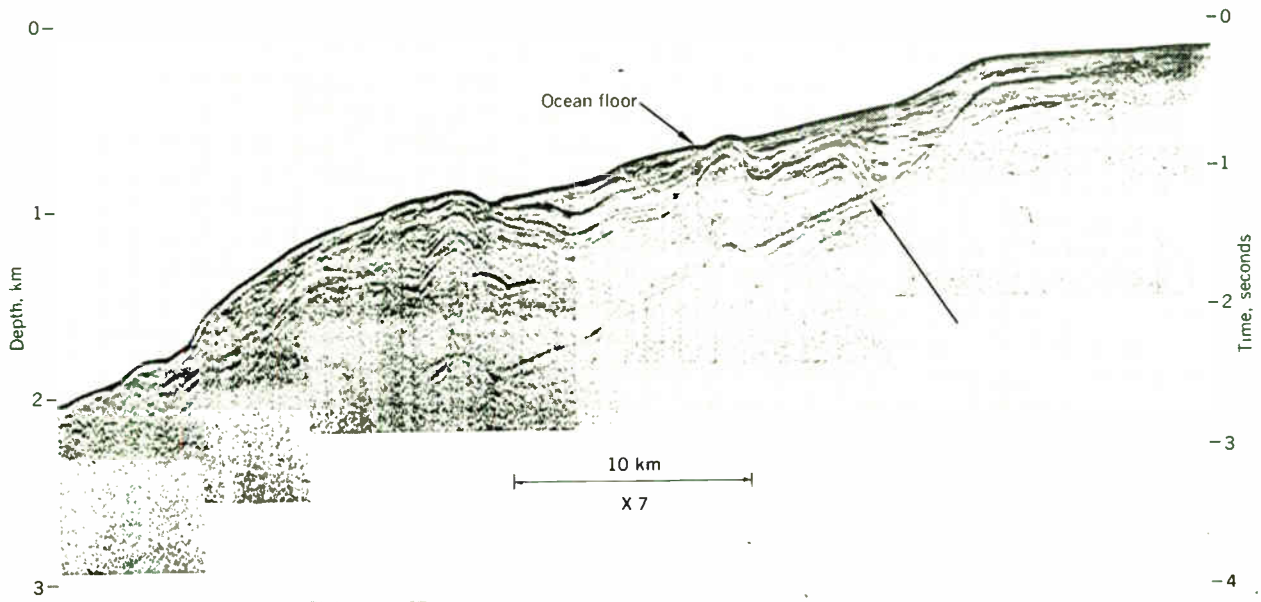
Mechanics of the tsunami. The tsunami is a long, surface gravity wave (often extending radially over an arc length of more than 2000 miles at a distance of 2000 miles from its source) formed by an impulsive vertical displacement of the sea floor. This dislocation is generally associated with a shallow-focus—less than 30 kilometers—earthquake, with a submarine epicenter, whose intensity is more than Force 7 on the Gutenberg–Richter scale (a logarithmic scale, calibrated from Force 0 through 10). From the studies of earthquake accelerations on land, plus precise triangulation of seismic waves from similar submarine earthquakes, the conclusion has been reached that the generating source, as defined by the initial tremblor and subsequent aftershocks, is usually elliptical in form.

The effect of a seismic shock may be a rapidly sequential—but discontinuous—series of multidirectional accelerative forces in the horizontal plane, plus a vertical component. The vertical motion causes a local elevation or depression of the sea surface, which disperses rapidly into a series of oscillatory waves that are analogous to the radiating wave train produced by dropping a stone into a shallow pool.

Most of our knowledge of tidal wave behavior is largely empirical; theoretical analyses by mathematical and physical models, and computer studies, have met with only qualified success. Physical model experimentation, however, shows that wave heights near the epicenter are directly proportional to the intensity and amplitude of the seismic disturbance. Also, the wavelength is related to the horizontal length of the line of vertical displacement (which may be more than 100 miles) along a geological fault plane. Thus the initial sea surface disturbance may contain numerous frequency components, including that which duplicates the resultant natural frequency of the seismic shock.

The insidious aspect of the tsunami is that those components which have long wavelengths in comparison to the ocean depth (where depth is less than one half of the

Fig. 12. Seismic profile, taken off the coast of northern California, shows the acoustical discontinuity within sedimentary rock layers. Arrow denotes the San Andreas fault, the landward end of which is responsible for most of California's earthquakes. Note that the horizontal scale of the profile is exaggerated by a seven-time reduction.



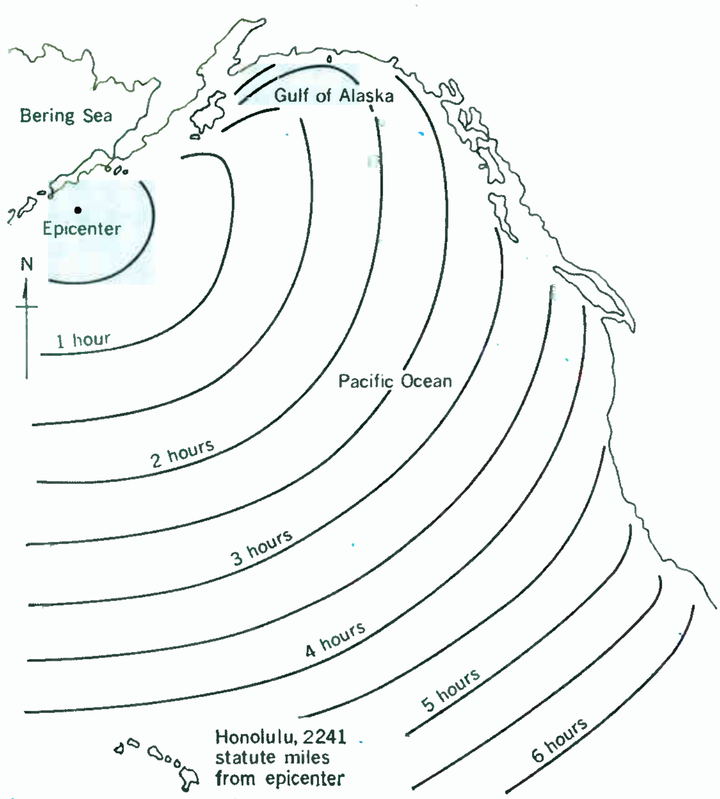
wavelength) will radiate, in the Pacific Ocean, at the phenomenal speeds of 400–500 knots (see Fig. 13). All shorter wave components will travel more slowly.

At an advanced stage in its radial evolution, the wave train consists of numerous individual crests, whose wavelengths diminish with time and with increasing distance from the front of the train at any instant of time, at a fixed point of observation or recording station. The amplitudes of the individual wave crests are modulated and damped by a slow pulsation that divides the train into groups. The maximum amplitude of each grouping diminishes slowly with distance from the front, and in direct proportion to the increase in distance from the source of the disturbance.

The highest individual crest in the wave train near the origin of the disturbance will be the first wave. Thenceforth, its order in the wave train will retrogress slowly until, at a distance of several thousand miles from the origin, it may be the sixth to pass the recording station.

Although the maximum initial tidal wave crest may be 70–80 feet in height, the highest wave at a radius of several hundred miles from the epicenter, in open ocean, will be only of the order of a few feet in height. When a tsunami approaches a large island or continental coastline, however, the cumulative effects of refraction, resonance, funneling through embayments and lagoons, etc., can

Fig. 13. Diagram showing the advance of the tidal wave of April 1, 1946, produced by an earthquake whose epicenter was southeast of Unimak Island in the Aleutian group. Note that within 4½ hours from its origin, the effects of this seismic phenomenon were felt all the way from the Oregon coast to Hawaii.



produce a huge local augmentation of wave height to carry the tsunami to watermarks that are 40–50 feet above the level of mean high water.

The trouble spots of the Pacific. There are known areas throughout the Pacific Ocean, such as along the Aleutian chain of islands strung out from the Alaskan coast, the Japanese and Philippine archipelagos, Chile, Indonesia, etc., that are subject to severe seismic or volcanic disturbances, the epicenters of which may be beneath the sea. Recently, huge tsunamis, generated by violent earthquakes in Chile, struck the Japanese islands (more than 10 000 miles distant) with devastating force, thereby illustrating dramatically the ability of this phenomenon to register its physical effect almost halfway around the world.

A possible warning network. To predict and to provide adequate warning of the approach of dangerous tsunamis, traveling at virtually subsonic speeds, an elaborate network of seismic detection devices, transmission relay buoys or vessels, and recording stations situated at strategic points in the oceanic islands is proposed.

An ocean-bottom seismograph system could be used in submarine earthquake zones for the continuous monitoring of seismic disturbances along the sea floor. The recorded analog input signals could be digitized and transmitted from permanently moored, unmanned surface buoys to distant relay ships. The surface buoys could also record and transmit data as to unusual wave agitation. Thus, for example, referring back to Fig. 13, Hawaii would have about a 4½-hour advance warning of a tidal wave that was generated in the Aleutians.

With sound and theory...

In this discussion of ocean engineering, we have attempted to delineate a broad-brush, "in depth" seascape of contemporary technology and the seismic and sonic instrumentation used in oceanographic operations. Yet, the range of theoretical and practical applications is broadening at an incredible pace.

For example, the detailed charting of every square mile of the ocean floor and its fantastic "mountain ranges" is proceeding intensively. Oceanographic survey ships constantly crisscross the seven seas, while the "ping" of the sonar and the visual displays accumulate data recorded along the continental shelves, in the abysses, and in the Stygian trenches that lie six miles below the surface.

...To volcanic fury

A current theory is that the birth of submarine volcanos may be predicted along the path of the mid-Atlantic ridge of that ocean's submarine mountain range. The eruption of a submarine volcano and the subsequent formation of the new volcanic island of Surtsey, off the coast of Iceland, has provided a measure of validity to this fascinating theory.

Figures 4 and 5 are reproduced by permission of Eric G. Barham, U.S. Navy Electronics Laboratory, San Diego, Calif. The illustrations originally appeared in *Science*.[2]

REFERENCES

1. Willett, C. F., "The manned submersible—ideal mapping tool," Internal Rept., Westinghouse Electric Corp., Underseas Div., Baltimore, Md.

2. Barham, E. G., "Deep scattering layer migration and composition: observations from a diving saucer," *Science*, vol. 151, no. 3716, pp. 1399–1402, Mar. 18, 1966.

A dc transformer

Like the left-handed monkey wrench, the concept of a dc transformer has always been something of an inside joke among engineers. However, by utilizing superconductors a device has now been created in which a direct current actually can be transformed

I. Giaever *General Electric Company*

Although conventional transformers are ac, a device that may be termed a dc transformer has been constructed by using superconductors. To provide an understanding of how such a transformer would operate, some of the properties of type I and type II superconductors are reviewed. Since the dc transformer under discussion is constructed from thin superconducting films, the main emphasis is on these structures; the concept of flux motion is also explained. The result of the work described is a device in which a direct current or voltage can be transformed, and in which it is possible to extract power from the secondary circuit.

Ordinary transformers operate on alternating current, and the concept of a transformer that can operate on direct current has often been referred to facetiously by electrical engineers. However, it is possible to construct a device that may be called a dc transformer by utilizing the unique magnetic and electrical properties of some superconductors. In fact, two entirely different types of dc transformers can be created with superconductors.

By simply winding the secondary coil in a conventional transformer with superconducting wire and using only a superconducting load, we can, in a sense, produce a dc transformer. The current in the secondary circuit will always follow that in the primary circuit, even when the primary current is a direct current. This is attributable to the fact that a superconductor has zero resistance below the superconducting transition temperature. However, as soon as we attempt to take power out of the secondary circuit—for example, by loading it with a resistor—it ceases to work on direct currents. Therefore, this device can be considered a dc transformer only in a trivial sense.

Type I superconductors

Superconductors are usually divided into two classes, referred to as type I and type II. A type I superconductor is the "classical" superconductor, which is characterized by a zero resistance and by the fact that the magnetic field inside the superconductor itself is always zero.[1] This type includes lead, tin, and indium.

To illustrate, let us consider an experiment in which a toroid or doughnut-shaped ring of tin is placed in a bath

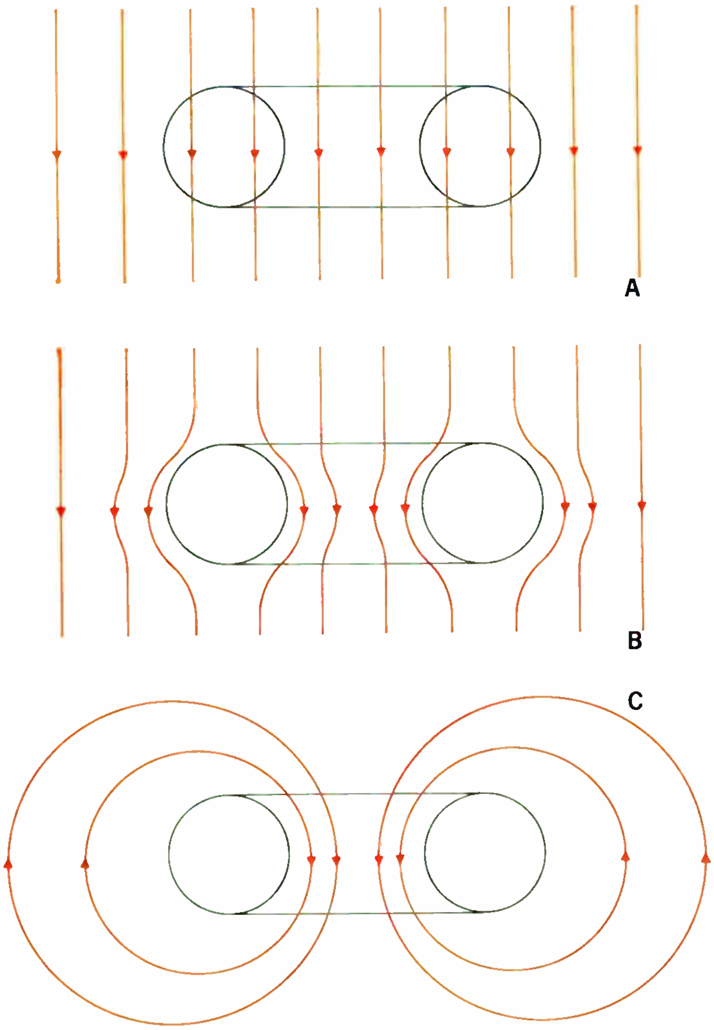


Fig. 1. (A) Toroid-shaped ring of tin in a small magnetic field, with the tin in the normal state. (B) The toroid in the superconducting state. The magnetic field has been excluded from the interior of the ring. (C) If the applied field is removed, a current is induced in the toroid such that the flux (fluxoid) contained in the hole is unchanged.

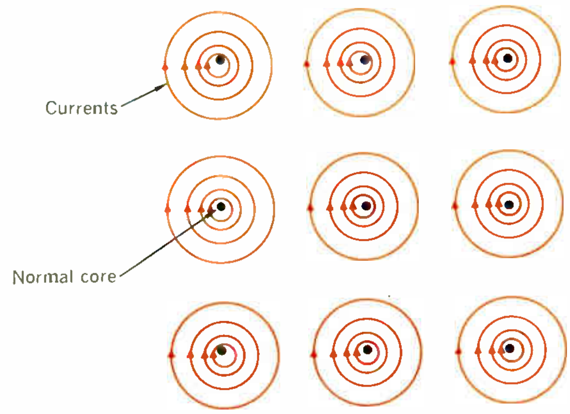


Fig. 2. Type II superconductor in an intermediate homogeneous magnetic field applied perpendicular to the plane of the paper. The magnetic field will penetrate in certain spots and a current will be induced that keeps the magnetic field "bundled up." In a very pure type II superconductor, the magnetic field will penetrate in a regular array.

of liquid helium, which has a boiling point of 4.2°K. Let us imagine that a small uniform magnetic field is applied parallel to the axis of the toroid, as shown in Fig. 1. At 4.2°K, the tin is still in its normal state and the magnetic field will be practically undisturbed by the presence of the ring; Fig. 1(A). Now let us slowly cool the assembly by letting the liquid helium evaporate under reduced pressure. Tin becomes superconducting at approximately 3.7°K, and if we look at the ring at 3.5°K, for example, we will find that the magnetic field has been expelled from the interior of the tin; Fig. 1(B). This occurs because currents are induced on the surface of the tin to a depth of about $\lambda \sim 500$ Å, called the penetration depth. These currents always flow in such a way as to make the field inside the superconductor zero.

Another peculiar phenomenon that arises is that the total currents and total fields always adjust themselves in such a manner that the following integral around any closed contour or loop contained inside a superconductor is satisfied:

$$\int \mathbf{B}\cdot d\mathbf{S} + \mu\lambda^2 \oint \mathbf{j}\cdot d\mathbf{l} = n\frac{h}{2e} = n\phi_o \qquad (1)$$

The first term equals the amount of flux contained within the chosen closed contour as B is the magnetic flux density and $d\mathbf{S}$ is an infinitesimal area element. The second term is the line integral of the current density j along the closed contour where $d\mathbf{l}$ is an infinitesimal line element, λ is the penetration depth, $\mu = 4\pi \times 10^{-7}$ in MKS units, n is simply an integer, and $\phi_o = h/2e \sim 2 \times 10^{-15}$ weber is the flux quantum where h is Planck's constant and e is the electronic charge. If the chosen contour is a simply connected structure, i.e., if the interior is superconducting, $n = 0$. However, if we integrate around a multiply connected structure such as our toroid, n can assume any integral value.

Because currents only flow close to the surface of type I superconductors, we can almost always choose a contour such that the second term on the left in Eq. (1) is zero. For example, if we integrate along a line in the interior of the toroid close to the center of its cross section, the current is effectively zero, and we talk (loosely) about "flux quantization"; that is, the flux contained within a macroscopically large superconducting ring is quantized in multiple units of ϕ_o. Any other path contained within the superconducting ring and which encloses the hole will give exactly the same answer for the sum of the integrals; however, the second term involving the current is not necessarily zero, and for this reason, London,[2] who first suggested this concept, spoke of the quantization of the "fluxoid," i.e., of the sum of the two integrals.

If the external magnetic field surrounding the superconducting toroid is removed, a current will be induced that flows around the ring, keeping the fluxoid constant; Fig. 1(C). Thus, flux contained in a macroscopic superconductor is not only quantized, it is also conserved. As long as the chosen contour of integration is kept within superconducting material, the fluxoid contained within the contour is conserved.

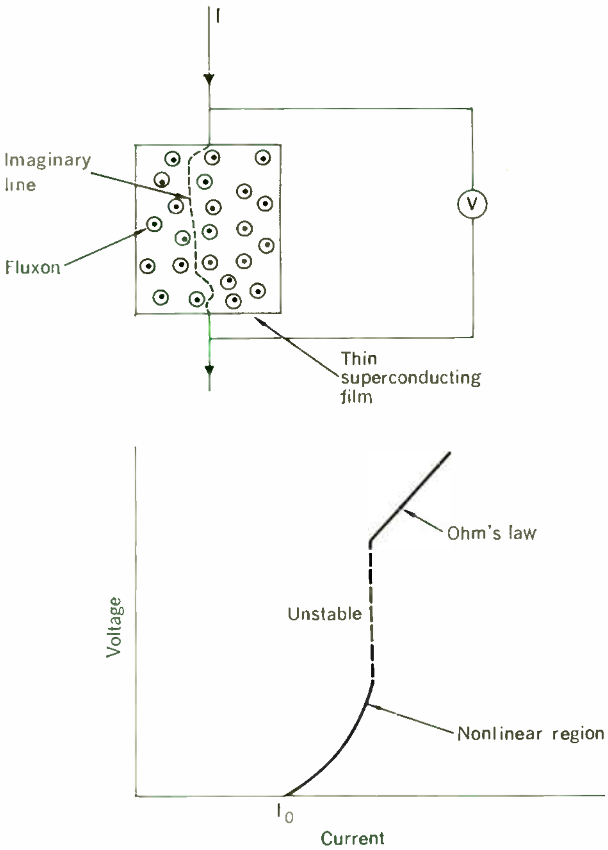


Fig. 3. Simple arrangement for measuring the current–voltage characteristic of a superconducting film placed in a small applied field perpendicular to the paper. At first, the transport current I avoids the normal cores, and the current flows without loss. At a current I_0, because of the Lorentz force between the flux spots and the transport current, the flux spots will move across the film, causing a voltage to appear. Because the flux spots move, the transport current cannot avoid the normal cores and in an approximate sense this accounts for the I^2R loss. Finally, at a higher current, the film abruptly enters the normal state and Ohm's law is obeyed.

Type II superconductors

Type II superconductors include alloys such as BiPb, PbIn, Nb_3Sn, and also possibly pure niobium. Just as with a type I superconductor, type II will exclude all the magnetic field from its interior when the applied magnetic field is low. If the applied field is increased (until the so-called critical field is reached), the type I superconductor reverts to the normal state through a first-order phase transition. Similarly, in a sufficiently high magnetic field, the type II superconductor also becomes normal; however, there is an intermediate range of fields in which the type II superconductor exists in a never-never land; that is, in a sense this superconductor is neither normal nor superconducting.

In the intermediate field range the flux penetrates the type II superconductor but only in certain spots, as shown in Fig. 2. It is possible to think of the center of each spot as a core of normal material surrounded by circulating current. In a sense, each of the spots, which we shall call "fluxons," looks like the toroid previously described, and the fluxoid of each spot is quantized and conserved. However, each fluxoid now will contain only one flux quantum—i.e., $n = 1$ in Eq. (1). The magnetic structure is often referred to as the "Abrikosov[3] structure," for the person who first visualized this strange behavior in a superconductor.

Thin superconducting films

Since the dc transformer to be discussed here is constructed from thin superconducting films, let us concentrate on such structures. Thin, evaporated, superconducting films tend to behave as type II superconductors regardless of the material they contain, partly because of their demagnetization factor and partly because evaporated films usually contain a large amount of impurities. Figure 3 shows schematically a simple experiment measuring the resistivity of a thin superconducting film in a small magnetic field applied perpendicular to the film. A typical result is also shown. When a small current is passed along the film, it flows without resistance. At a certain current I_0, however, a voltage will appear and, as seen from the graph, the relationship between the current and the voltage above I_0 is nonlinear. At a somewhat higher current, the film will go abruptly into the normal state, presumably because of heating effects, and the current–voltage characteristic follows Ohm's law.

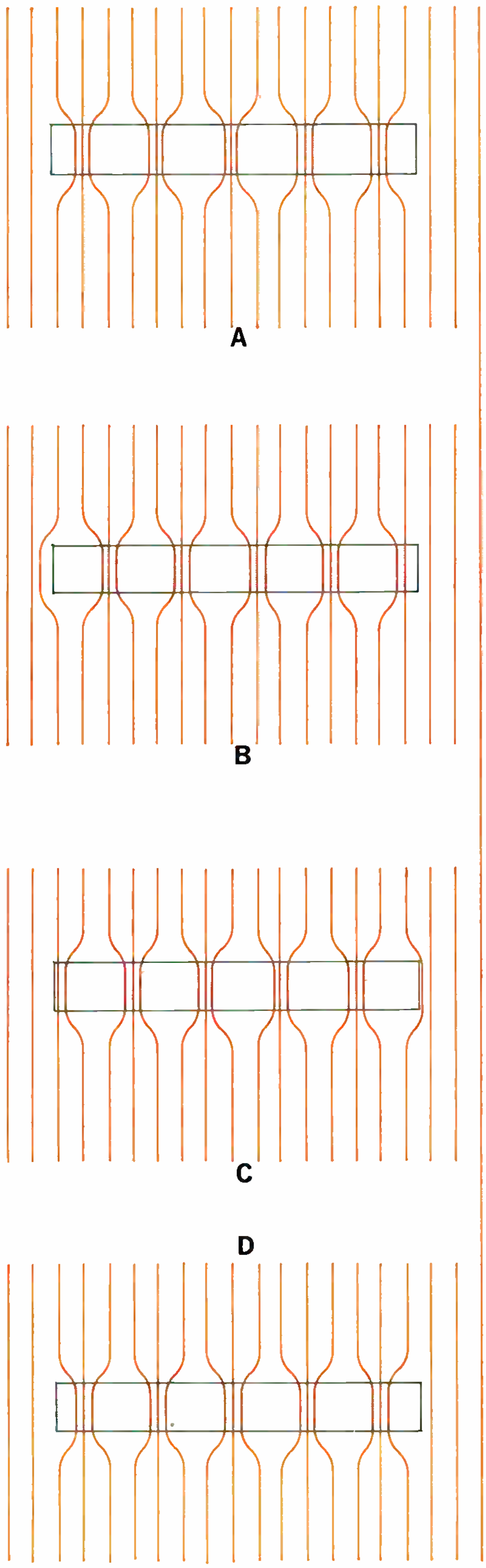
We shall be concerned only with the nonlinear region where a voltage first appears. If Fig. 3 is regarded as a "snapshot" of the superconducting film, then it seems strange that a voltage should appear across the film because an imaginary line can be drawn there without crossing any resistive material. However, we must remember that in general a voltage consists of both a resistive and an inductive term. Since the resistive voltage is zero, the answer must be that flux is cutting our imaginary line so that we obtain an inductive-like voltage. This result is achieved by letting the flux pattern in the film change as a function of time; specifically, the flux pattern will move across the film from one side to the other. Thus, from time to time, the rather arbitrary line we drew across the film will also be crossed by resistive material. This model of a moving flux pattern was first suggested by Anderson.[4] (A similar model had previously been suggested by Gorter[5] for type I superconductors in the intermediate state.) The model may seem very artificial and unphysical, as it literally moves flux across the film from one side to the other; however, if we remember that the flux lines have meaning only inside the superconductor then everything works out satisfactorily. Figure 4 is a schematic flux plot of the film with the flux lines in various positions. Since flux lines have no particular meaning in a uniform field, as can be understood from the illustration, the flux pattern within the film may change or move uniformly, leaving the magnetic field a short distance away from the film completely unaffected. The force on the flux lines is provided by the transport current flowing in the film.

To sum up, we regard the voltage appearing along a stationary superconductor in the type II state as being caused by moving flux lines, an almost exact analogy to the voltage appearing along an ordinary bar that moves in a stationary magnetic field.

Unipolar generator

The motion of flux inside superconductors is an interesting subject. The following describes a unipolar generator

Fig. 4. Schematic flux plot of the cross section of a superconducting film in a uniform field. As seen here, it is possible to let the flux move within the film without disturbing or changing the field a short distance away from the film. The important fact is that flux lines have no particular significance in a uniform field, whereas concentration of magnetic field is well defined inside the superconductor.



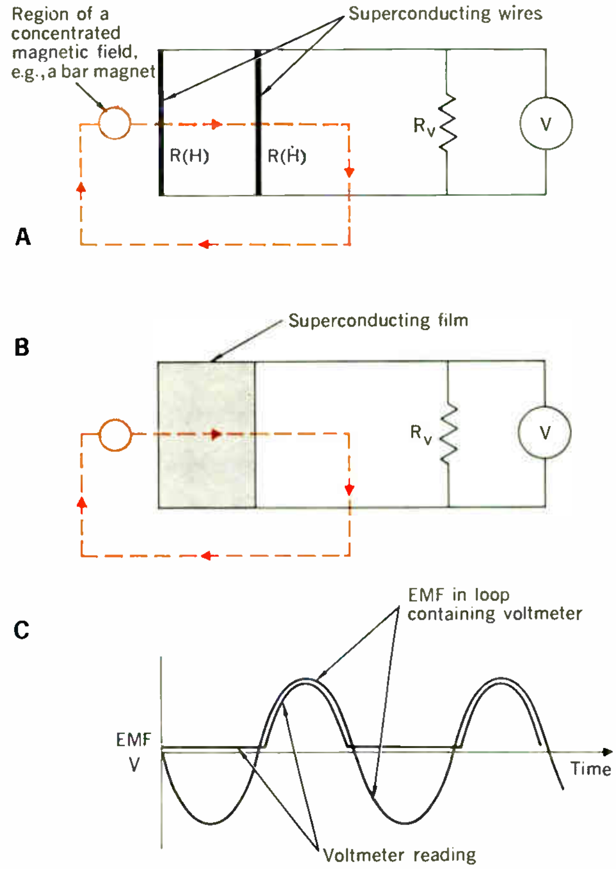


Fig. 5. (A) Possible unipolar generator. When the magnetic flux region crosses the first superconducting wire, it goes into the normal state and an EMF is also induced. No current flows at this time through the voltmeter as it is effectively short-circuited by the second superconducting wire. When the magnet crosses the second superconducting wire, the first wire short-circuits the voltmeter. Only when the flux region leaves the loop containing the voltmeter is a voltage observed. (B) The superconducting wires have been replaced by a flat superconducting plate. (C) The EMF in the loop containing the voltmeter and the voltage registered on the voltmeter as a function of time.

using superconductors, which is similar to one proposed by Volger.[6]

Figure 5 shows two superconducting wires connected in parallel across a voltmeter. Imagine a flux region due to a bar magnet that can move parallel to, but spaced from, the plane of the wires in the manner illustrated, such that the magnetic flux threading the loop changes periodically. Let us designate the resistance of the superconductor $R(H)$ and the resistance of the voltmeter R_v. We assume that the magnetic field is large enough to destroy superconductivity in the wires when the magnet passes over them. Thus, the resistance $R(H)$ depends upon the magnetic field H. When the flux region is moved across the first superconductor, the total current that flows through that conductor is

$$I = \frac{EMF}{R(H) + \dfrac{R(0) \cdot R_v}{R(0) + R_v}} = \frac{EMF}{R(H)} \qquad (2)$$

and the current through the voltmeter is simply:

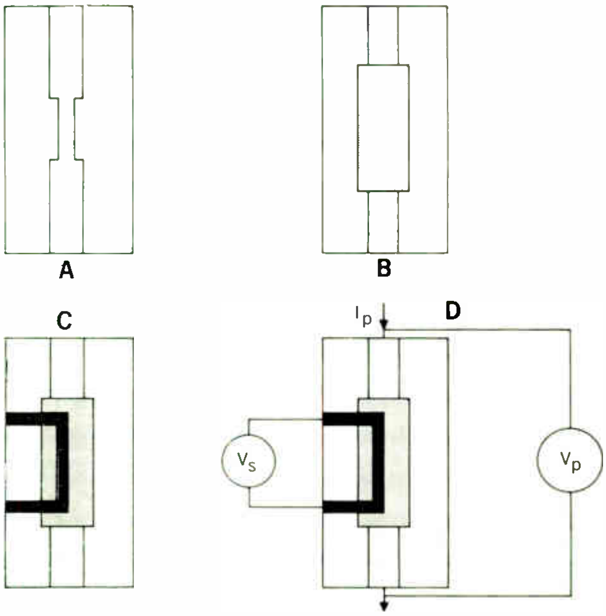


Fig. 6. Sample preparation. (A) Sn film deposited on a microscope glass slide. (B) Thin insulated layer of SiO vacuum-deposited over the primary Sn film. (C) Second layer of Sn evaporated over the first two layers. (D) Current and voltage connections to the primary film, and voltage connections to the secondary film.

$$I_v = \frac{R(0)}{R(0) + R(H)} I = 0 \tag{3}$$

Finally, the voltage registered by the voltmeter is

$$V = R_v I_v = 0 \tag{4}$$

[Since the resistance $R(0) = 0$, the voltage registered by the voltmeter is zero.] The same argument applies when the flux region crosses the second superconductor because now the first superconductor short-circuits the voltmeter.

When the flux region moves out of the loop containing the voltmeter, the current through the voltmeter is

$$I_v = \frac{EMF}{R_v + \dfrac{R(0) \cdot R(0)}{R(0) + R(0)}} = \frac{EMF}{R_v} \tag{5}$$

and therefore the voltage registered is equal to the induced electromotive force. We see then that the voltmeter on the average will register a direct voltage as the two superconducting wires effectively act as a perfect commutator (inductance effects are ignored). A similar effect is produced if the two separate wires are replaced by a flat plate as indicated in Fig. 5, as we may think of the plate in this instance as a number of wires placed in parallel.

The dc transformer

Let us consider the construction of a dc transformer.[7-9] The simplest procedure is shown in Fig. 6. First, a tin film a few thousand angstroms thick is evaporated onto a microscope glass slide. This film will be referred to as the primary. Next, a thin film of silicon oxide, approximately

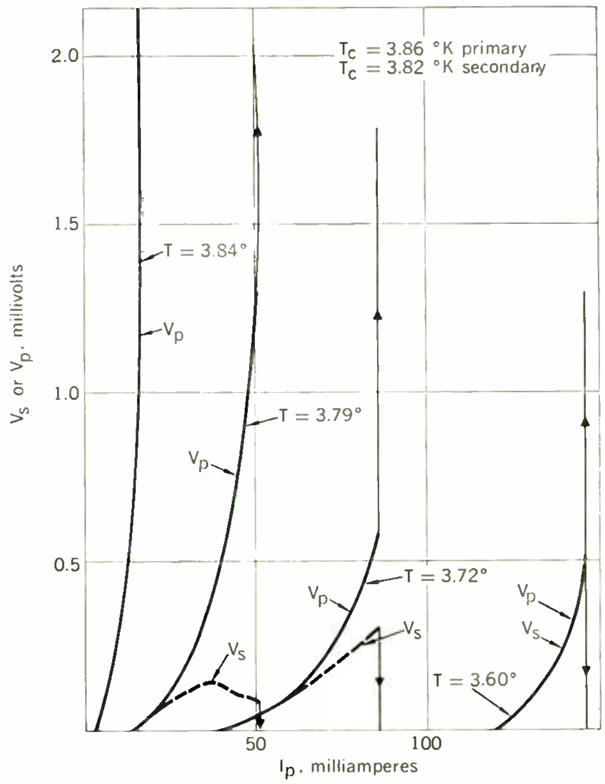
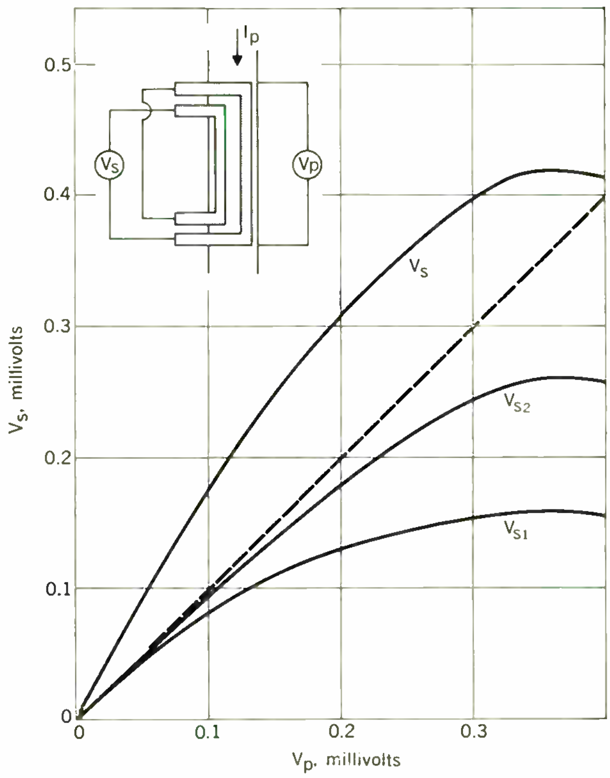


Fig. 7. Primary voltage V_p and secondary voltage V_s as a function of the primary current I_p at various temperatures. Note that the two films have different superconducting transition temperatures T_c, and that both films have to be superconducting for the transformer action to take place.

Fig. 8. Experimental arrangements of a transformer with a 2:1 turn ratio. Primary voltage V_p is plotted as a function of the secondary voltage, measured in each secondary film separately, V_{s1} and V_{s2}, and coupled in a series, V_s.



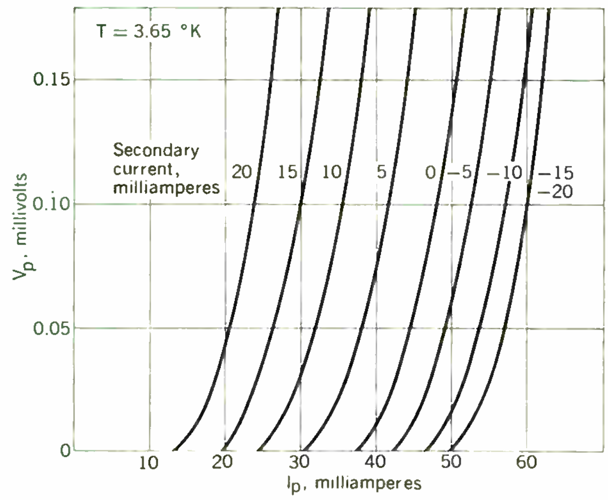


Fig. 9. Current-voltage characteristic of the primary film as a function of secondary current. The secondary current is taken as negative when it flows antiparallel to the primary current.

100–200 Å thick, is evaporated over the primary film to form a thin insulating layer. Finally, a second tin film is evaporated on the top. This film will be referred to as the secondary. The secondary film is made as thin as possible—of the order of 500–1000 Å—and slightly narrower than the primary film.

The experiment consists of passing a current down the primary film I_p and measuring both the primary voltage V_p and the secondary voltage V_s. Because of the narrow section in the center of the primary film, the voltage drop in the primary is completely paralleled by the secondary (the wide ends remain superconducting). The current and voltage connections are shown in Fig. 6. As illustrated in Fig. 7, as soon as both films are superconducting, a voltage appears in the top film.

Why does this voltage appear? The secondary film acts like the unipolar generator we described. The fluxons in the primary film behave as very small magnets. A current impressed on the primary film at one point will cause the fluxons to move across the film, and a voltage then appears along the primary film. Because the spacing between the films is so small, the secondary film senses this flux motion. Since the secondary film acts as a nearly perfect commutator, the same voltage appears across the secondary film.

If we want to transform the voltage, all we need do is to imitate a regular transformer. Figure 8 shows the results of putting two secondary films in series on top of one primary film. The device can also be used as a rectifier, as illustrated in Fig. 9, where the primary current–voltage characteristic is plotted as a function of secondary current.

The experimental results shown here are taken essentially without any applied magnetic field. The fluxons are caused entirely by the self-field from the transport current. However, basically the same effects are observed in small applied magnetic fields.

It should not be assumed that this dc transformer will be of great commercial value. At present, to my knowledge, superconducting electronics are not utilized commercially and the device probably does not have sufficient advantages to induce people to work with the liquid-helium temperatures that are necessary. However, it may find its way into commercial applications in which such low temperatures already are being used for other reasons.

Flux motion

The concept that the measured voltage in a type II superconductor is due to motion of flux is not so strange as it may first seem. Figure 1(C) shows a superconducting toroid that contains a certain amount of flux, as the result of a current induced in it. If the toroid is made of a type II material, flux lines will leak into the interior of the superconductor and annihilate each other in the middle of the ring. The net consequence will be that the flux enclosed in the ring will diminish and we then talk about flux leakage. If the toroid were made of a normal material, the flux enclosed in the ring would also diminish, but in this case we are inclined to refer to the resistance of the normal ring. Yet we can just as well take the point of view that flux leaks across the normal toroid and the two cases can be treated similarly. If we make the rings very large and insert a voltage source, we may still retain the concept of moving flux. In that case, the voltage source would really be a source of flux, whereas the resistive part would act as a flux sink. This picture is very satisfactory if we look at the Poynting vector in our system. The Poynting vector would indicate that energy flows into the fields from the battery, and then from the fields into the resistor.[10]

We must use extreme caution when we try to develop this and similar physical pictures. When such pictures are used to visualize a situation, it pays to check that the end result is consistent with Maxwell's equations, which are always valid. It has been shown that the simple application of the flux flow concept made in this article is indeed consistent with Maxwell's equations.

The author wishes to thank Drs. W. G. Johnston and M. D. Fiske for their helpful suggestions and discussions.

REFERENCES

1. Lynton, E. A., "Basic characteristics" in *Superconductivity*. New York: Wiley, chap. 1, 1962.

2. London, F., "The superconducting ring. Conservation law of the flux through a ring" in *Superfluids*. New York: Wiley, chap. 6, 1954.

3. Abrikosov, A. A., "On the magnetic properties of superconductors of the second type," *Soviet Phys. JETP*, vol. 5, p. 1174, 1957.

4. Anderson, P. W., "Theory of flux creep in hard superconductors," *Phys. Rev. Letters*, vol. 9, pp. 309–311, Oct. 1, 1962.

5. Gorter, C. J., "On the possibility of a dynamic variety of the intermediate superconductive state," *Physica*, vol. 23, pp. 45–56, Jan. 1957.

6. Volger, J., and Admiral, P. S., "A dynamo for generating a persistent current in a superconducting circuit," *Phys. Letters*, vol. 2, p. 256, 1962.

7. Giaever, I., "Magnetic coupling between two adjacent type-II semiconductors," *Phys. Rev. Letters*, vol. 15, pp. 825–827, Nov. 22, 1965.

8. Solomon, P. R., "Voltage associated with the coupled motion of flux in Type-I superconductors," *Phys. Rev. Letters*, vol. 16, pp. 50–52, Jan. 10, 1966.

9. Giaever, I., "Flux pinning and flux flow resistivity in magnetically coupled superconducting films," *Phys. Rev. Letters*, vol. 16, chap. 27-5, p. 460, 1966.

10. Feynman, R. P., Leighton, R. B., and Sands, M., "Examples of energy flow" in *The Feynman Lectures of Physics*. Reading, Mass.: Addison-Wesley.

A new theory of nerve conduction

Early nerve conduction theories created almost as many questions as they explained. Now semiconductor theory takes a crack at the mystery that has dazzled men since Galvani first noticed it in 1786—animal electricity

Ling Y. Wei *University of Waterloo*

This article offers a fresh view of the problem of nerve conduction through the "looking glass" of semiconductor physics. It sees the nerve axon as being something like a combination of an ionic transistor and a distributed active transmission line, a view that allows the interpretation of a number of important observations of nerve behavior.

Nerve conduction, though at the basis of all of man's activities, did not come to his notice until Galvani's discovery. One evening in September 1786, Luigi Galvani, professor of anatomy at the University of Bologna, discovered that a frog's leg twitched when he touched it by accident simultaneously with copper and iron. This phenomenon aroused his profound interest and curiosity in bioelectric processes. In 1791, he published an article, "De viribus electricitatis in motu musculari," in the Proceedings of the Bologna Academy, theorizing "animal electricity," claiming to have proved its existence, and thereby igniting one of the volcanic controversies in the history of science.

Alessandro Volta, a contemporary of Galvani, vehemently repudiated his theory and argued that the electricity of Galvani's experiments was produced by the reaction of dissimilar metals with the surrounding medium. To prove his thesis, Volta fabricated the famous Voltaic pile, which consisted of two stacks of paired zinc and silver disks packed between layers of brine-soaked pasteboard. When the two columns were connected at the top, current flowed between them, arising from the chemical reaction of the paired metals. This brilliant achievement really started modern electrical science. Unfortunately, Volta's scientific feat and apparent defeat of Galvani suppressed and delayed man's thinking and understanding of the nerve process by more than 100 years. It was not until the 20th century that man began to appreciate the meaning and value of Galvani's animal electricity.

Early theories

The first serious attempt to explore the mechanism of nerves was made in 1902 by Bernstein, who boldly postulated the so-called ionic hypothesis.[1] To comprehend Bernstein's hypothesis, one needs to know something about nerve axons. A nerve axon consists of a membrane in tubular form, residing in an external solution and containing an "axoplasma." The membrane is about 50–100 angstroms thick. Both the axoplasma and the external solution contain sodium, potassium, and chloride ions as well as small amounts of other ions. In the squid axon, the ion ratios are: $(Na)_i/(Na)_o = 1:9$, $(K)_i/(K)_o = 10:1$, $(Cl)_i/(Cl)_o = 1:14$, and $(Na:K:Cl)_i = 5:40:4$ where the subscript i indicates "inside" and o "outside." In the resting state, the membrane is polarized with outside positive; this membrane potential between the axoplasma and the external solution is about 60 mV. When the nerve is stimulated by a depolarizing potential (negative outside), there occurs the well-known all-or-none response.

Bernstein postulated that: (1) in the resting state, the membrane is permeable only to potassium ions but completely impermeable to ions of all other kinds; (2) when the nerve is stimulated, the membrane suffers "breakdown" and suddenly becomes equally permeable to all the types of ions present in the axon. Item (1) explains how the high concentrations of Na and Cl ions can be maintained in the external solution. Since the resting membrane is permeable only to K ions, they can diffuse across the membrane from the inside to the outside; the result is the building up of an electric potential (diffusion potential), which eventually makes the drift of K ions balance out their diffusion and thus brings the system into equilibrium. Potassium equilibrium potential (E_K) is given by the Nernst relation $E_K = (RT/F) \log (K)_i/(K)_o = 75$ mV, where R is the gas constant, T is the absolute temperature, F is Faraday's constant, and where the outside is positive. This relation is considered to be in good agreement with the observed membrane potential both in magnitude and in polarity. Item (2) explains the all-or-none response. However, the membrane breakdown would render the ion concentrations of each kind equal inside and outside the membrane and thus would reduce the membrane potential to zero during stimulation. This interpretation could not be tested in Bernstein's time.

For some 40 years, Bernstein's hypothesis enjoyed wide acceptance and popularity among neurophysiologists; but by 1940 it was beset by some disturbing observations. Boyle and Conway first observed that the resting membrane of muscle fibers was permeable to chloride ions as well as to potassium.[2] This finding, though a shock to the hypothesis, does not materially affect Bernstein's predicted result on the membrane potential because $(C)_o/(C)_i$ does not differ very much from $(K)_i/(K)_o$ on a logarithmic scale. What toppled the hypothesis was the discovery that the action potential (the membrane potential during stimulation) reverses its direction and makes the inside positive by as much as 40 mV relative to the outside.[3]

Although by the late forties, the hypothesis had become largely a historical artifact, there has survived as a legacy the following elements: (1) the selective membrane permeability to ions; (2) the change of membrane permeability with membrane potential; and (3) the dependence of the resting membrane potential on the ion concentrations by the Nernst relation. Bernstein is indeed regarded by many as the originator of the "membrane theory," which has dominated neurophysiology for over 60 years.

In 1951, Hodgkin and Huxley advanced the sodium theory, a revision of Bernstein's hypothesis, that accounted for most of the experimental findings about nerve processes up to that time.[4] They had studied ion transport in squid giant axons using microelectrodes and the voltage clamp technique. Their most significant result was to show how the action potential, and sodium and potassium currents, changed with time. These are shown in Fig. 1. Based on their experimental data, Hodgkin and Huxley devised a theory with these main points: (1) in the resting state, the membrane is moderately permeable to the potassium and chloride ions but is relatively impermeable to sodium and to the internal anions; (2) it is assumed that any sodium that leaks into the cell is pumped out by a secretory process; (3) during the excitation of the nerve the membrane momentarily becomes much more permeable to sodium ions than to any other ions, thus starting a regenerating cycle: depolarization of membrane → increase in sodium permeability → entry of sodium (provided $E > E_{Na}$) → further depolarization of membrane; (4) the inward flow of sodium ions continues until the inside potential builds up to such a high *positive* level as to repel the sodium ions; (5) in a short while (one or two milliseconds) after the peak of inward Na rush, the "sodium gate" (permeability) of the membrane is automatically switched off and the "potassium gate" is automatically switched on—the large outward potassium current repolarizes the membrane and thus returns the membrane potential to normal; (6) the local loop of sodium current depolarizes the neighboring region of the membrane so that the impulse propagates down the line.

Needless to say, a theory of such a splendid design could and should account for a large number of facts—and it does. As the name implies, the theory relies completely on the presence of sodium ions in the external solution and on the proper functions of the sodium gate and sodium pump in the membrane. An inquiring mind might question the physical grounds of these sodium devices (the gate and pump) but neurophysiologists largely accepted the theory and came to regard it as a classic.

In 1961, this situation took a sharp turn. About that time, an NIH (National Institutes of Health, Bethesda, Md.) group under Tasaki and the Cambridge group under Hodgkin were able to perfuse squid axon (by injecting an artificial solution into its axoplasma) and to study resting and action potentials in the perfused condition. This work was soon followed by the Tokyo group under Narahasi. The results obtained by these workers were so exciting and confusing that a conference was called in January 1965 in Miami to discuss the "newer properties of perfused squid axons."[5] It should be mentioned here that the sodium theory had been formulated at a time when reliable data were available from the study of *normal* squid

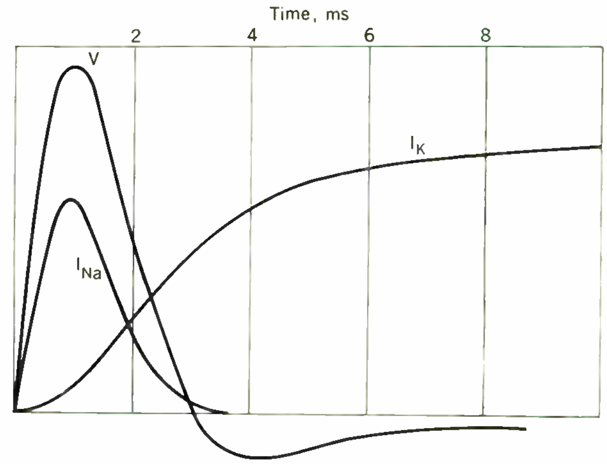


Fig. 1. How action potential (V), inward sodium current (I_{Na}) and outward potassium current (I_K) vary with time.

axon. Now, the experiments on *perfused* squid axon yielded many new facts incompatible with the sodium theory. According to Tasaki, the Na permeability of the resting (excitable) axonal membrane of squid was found to be *roughly equal* to the K permeability.[6] Tasaki, Singer, and Watenabe also demonstrated that under intracellular perfusion with favorable salt solutions, squid giant axons maintain distinct excitability in Na-free media.[7] This fact led Tasaki to conclude that a combination of a high external with a low internal concentration of Na is *not an indispensable* condition for the maintenance of excitability. In the light of such recent findings, one cannot help but consider the sodium theory as unsatisfactory.

What is wrong with the membrane theory? The answer is simple. At each stage, the theory was based only on limited experimental data. Without resorting to basic physics or chemistry, the major proponents of the theory chose to manipulate two vague terms, the membrane potential and the membrane permeability, neither of which represented functions that were well understood in biological systems. To complicate matters, the theorists bonded the two terms with reciprocal dependence. Things became even more entangled when the mysterious sodium pump and potassium pump[8,9] were brought onto the scene with their functions geared in an obscure way to the changes of the membrane potential and permeability. One can become so dazzled and puzzled by these bizarre theoretical interplays of gates and pumps that he does not think about what is really going on. The membrane theory is not a theory in the true sense but rather is an assembly of many artifices designed to fit the data at hand.

Though very little can be said about the gates and pumps of the membrane theory at this time, the action of the membrane potential can be clarified somewhat. According to Johnson, Eyring, and Polissar,[10] the total transmembrane potential difference (E) between the external solution and the axoplasma is the sum of three potentials, that is, $E = E_o + E_M + E_i$ where E_o and E_i are the potential changes at the outer and inner surfaces (or junctions) of the membrane and E_M is the potential change across the membrane proper. In the present state of the art, only E can be measured. So far, the workers in the field have taken E as E_M without giving much thought to E_o and E_i. As pointed out by Johnson and his colleagues,[10] the requirement $E_i = E_o = 0$ cannot strictly be met, since potassium would have to possess a *negative* distribution coefficient. Furthermore, they showed that the surface potentials are, in general, of the same order as the observed transmembrane potential, and that the value of E_M is apt to differ markedly from the observed value of E. With this understanding, one can see that to relate the *observed* transmembrane potential E to the potassium ion concentration in the two solutions by a simple Nernst relation is at best misleading. Unfortunately, with a few exceptions, most neurophysiologists invariably get these two things—the resting membrane potential and the potassium equilibrium potential—so confused that they spend all their efforts fabricating ways of making one fit into the other. That is why the gates and pumps were advanced. Since results from perfused squid axon almost completely invalidate the sodium theory, the time has come to take a new look at the problem.

Physical basis of axon action

To make a fresh start, we shall look at a nerve axon from a physical rather than a physiological basis. Here, physical basis means two things: the basic physics of ions and the physical structure of the nerve axon. We know that ions are electrically charged particles, and thus they should obey Coulomb's law and Poisson's equation. In an electric field, ions are acted upon by an electric force, $f = qE$, which causes them to flow in a "drift current," $I = q\mu nE$, where μ is the mobility and n is the ion concentration. Wherever there is a concentration gradient of ion species, they will diffuse down the gradient according to Fick's first law, $I_d = -qD\nabla n$, where ∇n is the concentration gradient and D is the diffusivity. In general an ionic current (J) has two components, the drift and the diffusion, given by the Nernst–Planck equation,

$$J = qn\mu E - qD\nabla n \qquad (1)$$

The distribution of ions in energy space is usually taken to follow Maxwell–Boltzmann statistics. Ions also observe Einstein's relation and the law of mass action. The ion concentration in a region where there are current flow, generation, and recombination is determined by means of the equation of continuity

$$\frac{\partial n}{\partial t} = G - R - \nabla \cdot J \qquad (2)$$

and G and R are the rates of generation and recombination, respectively, and $\nabla \cdot J$ is the divergence (or the *outward* flow rate per unit volume) of current. The J in (2) is given by Eq. (1). All the "first principles" stated here are no different from those usually applied to electrons or holes in a semiconductor. However, as everybody knows, a man's behavior is determined not only by "principles" (such as moral codes) but also by environment. Accordingly, we must examine the environment of ions—the physical structure of a nerve axon.

It is generally recognized that a junction region exists between a liquid and a solid. Thus, in a nerve axon, which consists of a membrane, an axoplasma, and an external solution, there are two junction regions separating its three main compartments. The total transmembrane potential between the inner and outer solutions is the sum of the three potentials, E_o, E_M, and E_i, where E_o and E_i are the potential differences across the outer and inner junctions, respectively, and E_M is that across the membrane per se. We must recognize the existence and the importance of the membrane junctions and the junction potentials. This is the basic difference between the new "junction theory" and the membrane theory.

In the inner and outer solutions the *active* current carriers are positively charged ions. In the jargon of semiconductor physics, these two ionic solutions may be regarded as p-type materials. Since the observed membrane potential is positive outside and negative inside, the outer junction is biased in the *forward* direction and the inner junction is biased in the *reverse* direction. Thus a nerve axon is very much like a p-n-p transistor both in its physical arrangement and in its biasing.

We now have a clear indication that the ions in a nerve axon and the electrons (or holes) in a p-n-p transistor follow the same physical principles and have similar physical environments. Therefore, we should expect them to behave in quite similar ways. In what follows, we shall test this physical theory against the findings of nerve studies. Since the theory is not based on experiments, we are in the comfortable position of judging its validity by examining how its predictions agree with the observations.

Sodium and potassium currents

One of the most significant results from Hodgkin's and Huxley's extensive studies of normal squid axon is the plot of the change in sodium and potassium currents with time. It is the basis of the sodium theory, a theory that does not interpret but merely describes the experimental facts in a neat form. We shall show that the physical theory can predict the results observed by Hodgkin and Huxley. There are two ways of doing the derivation. The first is to employ an equivalent circuit and to take the Laplace transform. The second is to resort to first principles—the equation of continuity and the Nernst–Planck equation. We shall merely indicate the method.

Our analogy of the nerve axon (as being like a p-n-p transistor with ions replacing the role of electrons and holes) may be clearer if we look on the external solution as the emitter, the membrane as base, and the axoplasma as collector. Since the external solution is usually at or referred to as the ground potential, a nerve axon has the so-called grounded-emitter or common-emitter configuration. The equivalent circuit of such a configuration can be found in any electronics book today. If one applies a stimulating (step) potential between the external solution (emitter) and the axoplasma (collector), part of the potential will automatically appear across the emitter junction. Denote this voltage as ΔV, the base voltage. By means of the Laplace transform, it is easy to calculate the base and the collector currents with the equivalent circuit (which completely specifies the transfer function). These calculated transient base and collector currents[11] bear a great resemblance to the sodium and potassium currents shown in Fig. 1. Furthermore, the calculated maximum values of these currents (I_b and I_c) are proportional to the base voltage ΔV, and hence to the stimulating potential, just as Hodgkin and Huxley observed for the I_{Na} and I_K currents. In the circuit point of view, the fast decay of I_b (or I_{Na}) and the delayed rise of I_c (or I_K) are primarily due to the action of the capacitances and resistances of the emitter and collector junctions.

If one is not familiar with transistor circuits or has little trust in such an analogy, he may try to solve the two basic equations—the equation of continuity and the Nernst–Planck equation for the ionic currents in a nerve axon—by assuming a constant field in the membrane along with appropriate membrane boundary conditions. Such a procedure has been taken by Johnston to obtain the transient response of a drift transistor for which a constant field was assumed in the base region.[12] The results are not essentially different from those obtained from circuit analysis. It is to be understood that the two basic equations employed describe the actions of four fundamental physical processes—generation, recombination, drift, and diffusion—and involve nothing mysterious. Neither the circuit analysis nor the solution of the two basic equations requires a special role played by sodium ions or any prescribed changes of the membrane permeability to ions. Thus, either approach should be valid for a nerve axon, normal or perfused. We have at long last reached the stage when we can begin to feel the shattering of the artificial restraints imposed on neural theory!

Excitation and inhibition

A nerve axon is not merely like an ionic transistor; it is also a transmission line. To be more exact, it is a distributed active line. The transistor part provides gain while the line part suffers loss. The criterion for nerve excitability is then

$$G_u \geqslant L_u \tag{3}$$

where G_u and L_u are the gain and loss per unit length respectively. If

$$G_u < L_u \tag{4}$$

the axon will be inhibited or show only local response.

As discussed, the transistor-like part of the nerve axon is in a grounded-emitter configuration. The gain G_u is proportional to the current multiplication factor α_{cb},

$$G_u = K\alpha_{cb} \tag{5}$$

where
$$\alpha_{cb} = \left.\frac{\partial i_c}{\partial i_b}\right|_{V_c = \text{constant}} \tag{6}$$

From (3), (4), and (5), it can be seen immediately that the nerve excitability depends decisively on α_{cb}. Here is another test for our theory: to compare alpha theory predictions and the observations on nerve excitability.

The standard procedure for calculating α_{cb} is to solve the two basic equations, the equation of continuity and the Nernst–Planck equation. However, there are several important factors to consider. The first is the recombination of ions at the interface between the external solution and the membrane. The second is the effect of the field in the membrane. The third is the membrane conductivity modulation by the stimulating current. And the fourth is the volume recombination in the membrane. Webster has considered all these effects in a transistor and arrived at an expression for α_{cb},[13]

$$\frac{1}{\alpha_{cb}} \simeq \frac{swA_s}{D_pA} g(Z) + \left[\frac{\sigma_b w}{\sigma_e L_e} + \frac{1}{2}\left(\frac{w}{L_b}\right)^2\right](1 + Z) \tag{7}$$

$$Z = w\mu_e I_E/D_pA\sigma_b \tag{8}$$

where the terms, when translated for the axon, are

s = surface recombination velocity
A = area of the membrane junction per unit length
A_s = effective area for ion recombination at the interface between the membrane and the external solution
w = thickness of the membrane
D_p = diffusion constant of sodium ions (or of the majority active cations in the external solution)
L_e = diffusion length of anions in the external solution
L_b = diffusion length of sodium ions in the membrane = $(D_p\tau_p)^{1/2}$; D_p and τ_p are their diffusion constant and lifetime
σ_e = conductivity of external solution
σ_b = conductivity of membrane
μ_e = mobility of anions in the membrane
I_E = emitter current or the inward injecting current
$g(Z)$ = a function that decreases sharply near $Z = 0$ and very gradually after $Z = 5$.

The first term on the right-hand side of Eq. (7) represents ion recombination at the interface of the membrane and external solution; the term $(\sigma_b w/\sigma_e L_e)$, inside the bracket, represents the "injection efficiency" or the ratio of non-sodium to sodium inward flux; and the term $(w/L_b)^2$ represents the ion recombination within the bulk of the membrane. The multiplying factors $g(Z)$ and $(1 + Z)$ arise from the effect of the field in the membrane. Since $g(Z)$ decreases with Z, which is proportional to the emitter current (I_E), the α_{cb} as given by (7) will rise with I_E, reach a maximum value, and then decrease with I_E. The plot of the calculated α_{cb} versus I_E is shown in Fig. 2.

The most prominent consequence of Eq. (7) and Fig. 2 is that the nerve response would be all-or-none because at very weak stimulus (I_E), α_{cb} is too small and so $G_u \angle L_u$. This means either no response or only a local response; after the peak, α_{cb} decreases with I_E and thus the output stays nearly constant. After a certain large value of I_E, α_{cb} may drop to such a low value that again $G_u \angle L_u$, indicating that the nerve would be inhibited. This prediction was substantiated by the so-called Wedensky's paradox—a small cathodal current initiates a conducted impulse whereas a larger current fails to do so.[14]

The α_{cb} given by Eq. (7) predicts that it should decrease with decreasing conductivity (σ_e) of the external solution. There is ample evidence that the reduction of the sodium concentration in the external solution decreases the action potential and hence the ability of nerves to conduct impulses.[4,15–18] However, (7) does *not explicitly* contain σ_c, the conductivity of the axoplasma. This implies that change of ion concentrations in the axoplasma, if not too drastic, would have little effect on action potentials. Interestingly enough, this happens to be the most important result coming out of the recent experiments on *perfused* squid axons,[5,19,20] and it is the very fact that had confused and puzzled many neurophysiologists whose traditional thinking had always banked on the intimacy of a relation between the membrane potential and ion concentrations. Here is a theory to show them otherwise. The apparent alpha independence of the ion concentrations in the axoplasma is the result of the biasing potentials at the outer and inner junctions. As explained before, the outer junction is biased in the forward direction and the inner junction is biased in the reverse direction. Thus the flow of outer ions inward is much easier than that of inner ions outward. Only when the outer ions reach the inner junction and cause a temporary change in the potential at that junction can the inner ions see and respond to the stimula-

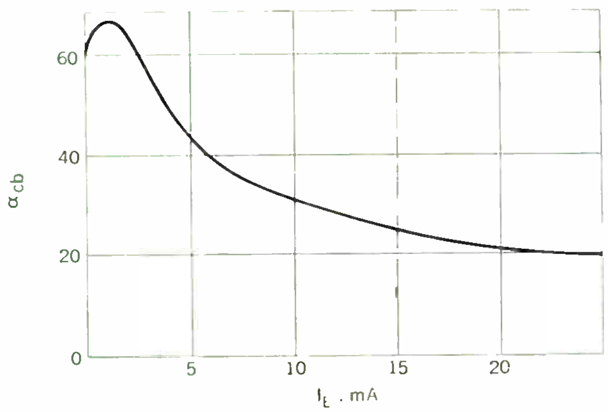


Fig. 2. Current amplification factor α_{cb} versus emitter current I_E as calculated from Eq. 7.

tion. Hence the alpha or nerve response depends decisively on how many outer ions reach the inner junction and their transit time. This final number of outer ions and their transit time undoubtedly depend on: (1) the total number of cations available in the external solution; (2) their chance of survival in the course of travel through the outer interface and the membrane; (3) their traveling speed; and (4) the distance of travel. With this understanding, one can see without much difficulty the rationale of the functional relation as given by Eq. (7).

The first term in (7) is due to surface recombination or, in other words, to the chance of survival of outer ions as they cross the interface between the external solution and the membrane. It is conceivable that certain agents may adhere to this interface and act as traps for cations. There has been some evidence of this effect. Kishimoto and Adelman have studied the effect of detergent on electrical properties of squid axon membrane.[21] A detergent is generally considered to be a surface active agent that adheres to boundaries between two different phases such as exist between living membranes and extracellular fluid. Kishimoto and Adelman observed that both anionic and cationic detergents generally decrease the Na and K conductances irreversibly. The effect of nonionic detergent is to decrease the Na conductance reversibly, leaving the K conductance almost unchanged. Curran *et al.* have studied the effect of Ca and antidiuretic hormone (ADH) on Na transport across frog skin.[22] They concluded that the primary effect of both agents is on the Na permeability of the *outward facing* membrane of the cells; Ca decreases and the hormone increases permeability of this barrier. Neither agent (assuming that it is located at the inner membrane of the cells) appears to have a direct effect on the active transport system itself. All these facts point to the importance of surface recombination in the ion transport across the membrane, as indicated by the first term in Eq. (7). The value of (7) lies in the fact that it can explicitly relate the nerve response to the physical quantities of the axon and that it is not based on experimental data but is derived from the two basic equations of physics. No existing theories can claim such a status. Not only can (7) be used to interpret a number of important observations but it also suggests some experiments to test the theory further.

In conclusion, if one recognizes the existence of the membrane junctions and the junction potentials in a nerve axon, one can easily derive the nerve behavior from semiconductor physics. This should not be surprising either to the general reader or to the specialist for the basic physics is one and the same in the two domains. It is this physics that can and should provide us with a basic understanding of nervous conduction.

REFERENCES

1. Bernstein, J., "Untersuchungen zur Thermodynamik der bioelektrischen Ströme," *Pfluegers Archiv fuer die Gesamte Physiologie*, vol. 92, pp. 521–562, 1902.

2. Boyle, P. J., and Conway, E. J., "Potassium accumulation in muscle and associated changes," *J. Physiol.*, vol. 100, pp. 1–63 1941.

3. Hodgkin, A. L., and Huxley, A. F., "Action potentials recorded from inside a nerve fibre," *Nature*, vol. 144, pp. 710–711, Oct. 21, 1939.

4. Hodgkin, A. L., "The ionic basis of electrical activity in nerve and muscle," *Biol. Rev. Cambridge Phil. Soc.*, vol. 26, pp. 339–409, 1951.

5. Cole, K. W., *et al.*, "Newer properties of perfused squid axons," *J. Gen. Physiol.*, vol. 48, no. 5, pt. 2, pp. 1–92, 1965.

6. Tasaki, I., "Permeability of squid axon membrane to various ions," *J. Gen. Physiol.*, vol. 46, pp. 755–772, 1963.

7. Tasaki, I., Singer, I., and Watanabe, A., "Excitation of internally perfused squid giant axons in sodium-free media," *Proc. Nat. Acad. Sci.*, vol. 54, pp. 763–769, Sept. 15, 1965.

8. Katz, B., "How cells communicate," *Sci. Am.*, vol. 205, pp. 209–220, Sept. 1961.

9. Eccles, J., "The synapse," *Sci. Am.*, vol. 212, pp. 56–66, Jan. 1965.

10. Johnson, F. H., Eyring, H., and Polissar, M. J., *The Kinetic Basis of Molecular Biology*. New York: Wiley, 1954.

11. Shea, R. F., *Transistor Circuit Engineering*. New York: Wiley, 1958, p. 306.

12. Johnston, R. C., "Transient response of drift transistors," *Proc. IRE*, vol. 46, pp. 830–838, May 1958.

13. Webster, W. M., "On the variation of junction-transistor current-amplification factor with emitter current," *Proc. IRE*, vol. 42, pp. 914–920, June 1954.

14. Lorento de Nó, R., and Honrubia, V., "Electrical stimulation of the internodes of single fibers of desheathed nerves," *Proc. Nat. Acad. Sci.*, vol. 52, pp. 1142–1149, Nov. 15, 1964.

15. Hodgkin, A. L., and Katz, B., "The effect of sodium ions on the electrical activity of the giant axon of the squid," *J. Physiol.*, vol. 108, pp. 37–77, 1949.

16. Nastuk, W. L., and Hodgkin, A. L., "The electrical activity of single muscle fibers," *J. Cellular Comp. Physiol.*, vol. 35, pp. 39–73, 1950.

17. Huxley, A. F., and Stampfli, R., "Evidence for saltatory conduction in peripheral mylinated nerve fibres," *J. Physiol.*, vol. 108, pp. 315–339, 1949.

18. Draper, M. H., and Weidmann, S., "Cardiac resting and action potentials recorded with an intracellular electrode," *J. Physiol.*, vol. 115, pp. 74–94, 1951.

19. Tasaki, I., Watanabe, A., and Takenaka, T., "Resting and action potential of intracellularly perfused squid giant axon," *Proc. Nat. Acad. Sci.*, vol. 48, pp. 1177–1184, July 15, 1962.

20. Baker, P. F., Hodgkin, A. L., and Shaw, T. L., "The effects of changes in internal ionic concentrations on the electrical properties of perfused giant axons," *J. Physiol. (London)*, vol. 164, pp. 355–374, 1962.

21. Kishimoto, U., and Adelman, W. J., Jr., "Effect of detergent on electrical properties of squid axon membrane," *J. Gen. Physiol.*, vol. 47, pp. 975–986, 1964.

22. Curran, P. F., and Gill, J. R., Jr., "Effect of calcium on sodium transport by frog skin," *J. Gen. Physiol.*, vol. 45, pp. 625–641, 1962.

23. Herrera, F. C., and Curran, P. F., "The effect of calcium and antidiuretic hormone on sodium transport across frog skin. I—Examination of interrelationships between Ca and hormone," *J. Gen. Physiol.*, vol. 46, pp. 999–1010, 1963.

24. Curran, P. F., Herrera, F. C., and Flanigan, W. J., "The effect of calcium and antidiuretic hormone on sodium transport across frog skin. II—Sites and mechanisms of action," *J. Gen. Physiol.*, vol. 46, pp. 1011–1027, 1963.

Motion perception using oscilloscope display

Current investigations for the purpose of providing more efficient visual recording, transmission, and display methods are based on the importance of gaining an increased understanding of the human observer's eye–brain perceptual mechanism

C. F. Hempstead *Bell Telephone Laboratories, Inc.*

A new type of stimulus generator has been designed and built to investigate some of the properties of the eye–brain perceptual mechanism and how it operates to perceive and interpret motion. Several time constants of the eye–brain have been measured, and their relations to other known time constants have been interpreted. Some time constants are associated with the brain and others with the retina, but all appear to interact with each other and with time constants of comparable magnitude that may occur in practical visual display systems. The application of this information to the design of practical systems is discussed briefly.

Methods of recording and reproducing the motions of objects in the real world for remote or delayed display have existed for at least a century. A number of inventions, such as the film motion-picture machine, television, and magnetic recording devices, have made possible almost universal availability of high-quality visual presentations. Some of these are exceedingly lifelike, although we are seldom really fooled by the illusion. In spite of these technological accomplishments, and in spite of the recent upsurge of physiological and psychological research into the visual system of many animals, our detailed understanding of how the eye–brain perceptual mechanism operates is still very limited. A large number of often-conflicting hypotheses fill the literature, and much of the technological progress being made in this field is still based on trial-and-error engineering. Widespread use and economic considerations are now building up significant pressure for more efficient media of visual recording, transmission, and display. But the engineering to accomplish this must be based on a better understanding of the basic mechanisms in the final receiver—the human observer.

In an attempt to gain a better understanding of how the eye–brain perceptual mechanism operates to perceive motion, a number of new experiments have been carried out. These have been conducted with apparatus sufficiently flexible to allow independent variation of most of the parameters affecting motion perception, yet sufficiently simple to allow testing in a reasonable length of time.

Background information

The illusion of motion; stroboscopic display and sampling. All visual motion recording and reproducing systems having general usefulness are based on a sampling technique in which the real-world objects are recorded or stored as a sequence of discrete two-dimensional images or "snapshots" containing luminance information. The reason is economic; it is much too costly to use a continuous independent storage or transmission channel for each resolvable area in an image having two spatial dimensions as well as luminance and time dimensions. Even low-definition images contain more than 10 000 spots. The property of the eye–brain mechanism that makes this sampling technique acceptable is only partly understood. If the complete picture-sampling period is sufficiently short compared with (1) the "time constant" of the eye–brain and (2) the time required for the image to move a resolvable distance, then the illusion of motion in the display is quite satisfying. It is not clear that these are also necessary conditions in the mathematical sense. But when either of these conditions is not met, the illusion of real motion in the display is partly or wholly destroyed, and the observer may be quite dissatisfied. It is the lack of understanding of visual motion perception that has prompted this investigation. A short sampling period is expensive in bandwidth required, and any "trick" that the eye–brain will tolerate and that will permit increasing the sample period will be very attractive economically.

History and existing theories. Even before the invention of the film motion-picture machine, it was known that a sequence of images displayed in rapid succession appears continuous. The phenomenological explanation is that the eye–brain perceptual mechanism has a finite temporal resolving power, just as it has a finite spatial resolving power. The latter is limited, through diffraction, by the wavelength of light, but the limiting process for temporal events is thought to be more complex, for the following reasons:

1. As light is absorbed in the retinal receptors there is a physiochemical bleaching action, which results in a time constant for recovery of initial sensitivity.
2. The messages sent from these receptors to the brain consist of sequences of electrical pulses called spikes; there is a maximum frequency at which the supposed relaxation oscillator that generates these spikes can operate, and thus another time constant is introduced.
3. Finally, there is some evidence that the brain itself operates by scanning its multitudinous inputs in an intermittent or sampled-data manner, which results in a third time constant.

One result of these time constants is that there exists a critical fusion frequency. A sequence of images displayed at a higher frequency appears continuous, whereas at lower frequencies the observer sees either a flickering image or a sequence of discrete images with little sense of continuity from one to the next. Although the critical fusion frequency depends on color, average luminance, and the waveform (in time) of the luminance, it varies only slightly among normal human observers.

This phenomenon (which is still being actively investigated) will be referred to as display flicker, to distinguish it from a more recently studied phenomenon called edge flicker.[1] The latter, associated with high contrast moving edges, is independent of display flicker in its occurrence, although the two may have a common ultimate cause. Edge flicker will be described in more detail in a subsequent section.

Numerous studies of motion rendition, particularly of apparent motion, have been made during the past 50 years. The Gestalt psychologists have been the center of a considerable controversy about motion perception, and at one time the number of theories approached the number of psychologists. Factors known to affect the perception of apparent motion include (1) the temporal duration of each image, (2) the temporal separation of successive images or the amount of temporal overlap, (3) luminance, (4) surrounding field, (5) the amount of experience and psychological "set" of the observer, and (6) the presence or absence of other clues (sensory inputs), which may include auditory, orientation, and acceleration, as well as peripheral visual stimuli. Also, the basic problem of (7) perception of (static) shape and figure is intimately connected with motion perception. For example, certain styles of sports cars "appear" to be going fast even when they are standing still.

Two factors enter into any *quantitative* perception of velocity: change in position and time interval. Thus an observer needs two scales or standards for calibration. The sensation of time intervals is notoriously variable for humans, and there are conditions under which space or distance judgment is almost as poor. Hence quantitative experiments often reduce to *A–B* comparisons (i.e., *A* is bigger or better than *B*, or vice versa) rather than attempts to establish an absolute scale.

Experimental situation. Experiments in motion perception carried out more than ten years ago tended to be qualitative and specific. Rather than designing a test to compare various theories, the experimenter often designed the test to "verify" a particular theory, often his own. The design philosophy of the physical sciences has been widely applied only recently to biological and psychological research. And only recently have the highly precise and powerful instrumental techniques of the physical sciences started to supplement the statistical approach used earlier. The past decade, however, has witnessed a rapid growth in the use of well-controlled experiments, using delicate measuring devices and, in particular, multiple stimuli and response sensors. In addition, much progress has resulted lately from the use of high-speed on-line computers with large, essentially random access memories.

Nevertheless, the eye–brain perception problem is an exceedingly complex one, and many levels of understanding will be required. Most of the early experiments were based on stimuli so simple that the situation was highly artificial. The results often were merely clues to how the visual system operates, rather than facts that could be incorporated into a larger theory. At the other extreme were the complete systems, such as theater motion pictures and television. Practical design questions needed prompt, if only qualitative, answers, and as soon as a (temporarily) satisfactory solution was found, further research slowed down until new demands for quality forced it back into activity.

Design of new experiments

Flexibility vs. complexity. Two conflicting factors require compromise in the design of an experimental setup to investigate motion rendition with a television-type display. These factors are flexibility and complexity. To be even reasonably useful the system must allow independent control over such parameters as luminance, image duration, image display rate, new information frame rate, target or object motion (position and velocity), and background. Other parameters that may usefully be varied include effective integration time in the system between the real object and the display, resolution in the image, color of the image display, and viewing conditions.

This is a rather formidable list of variables, even before the complexity of a real-time television system is introduced. Since the latter is difficult enough to keep in good adjustment even when some of the foregoing parameters—e.g., display rate and new frame rate—are held fixed, a serious compromise is in order. The system that seems to offer the best compromise has grown through several stages, but is basically a simulator. Final display for the human observer is a cathode-ray tube (CRT) oscilloscope. The trace (beam position) is driven not by a complete television system, but by one of a variety of deflection generators. These generators are in turn controlled by an electrooptic-mechanical transducer and/or an electronic stimulus generator having the complexity of a very small and simple computer. There is no camera as such in the system, and the display of motion is essentially one dimensional (horizontal). The vertical dimension is effectively held constant and merely provides a line instead of a point as the basic target.

The mechanical transducer allows direct control of the CRT trace by a motion of, say, the hand. Since many applications will involve observation of a moving person, such a transducer facilitates the simulation of motions having acceleration characteristics of humans. For other purposes, the generation of images having known velocities and accelerations is essential and, for this purpose, a function (or waveform) generator provides great flexibility.

Goals. The entire system, while containing a large degree of quantitative control, was designed to simulate a wide variety of viewing situations and give qualitative tests of various theories quickly. To this end, some effort has been devoted to the design of convenient, independent controls, so that the desired parameters can be varied quickly and easily, freeing the experimenter to concentrate on the experiment itself rather than the apparatus. It is, in fact, quite easy to experiment on oneself, and the resulting immediate feedback is most useful.

The main goal then was to be able to evaluate those new ideas or hypotheses about the eye-brain perceptual mechanism that arise in attempts to understand how we perceive and evaluate illusory motion displays of real motion. Detailed exhaustive studies on a particular phenomenon were not an object; instead, many more cursory experiments seeking interrelations between hypotheses were the goal. The more basic hypotheses are described in this section and several specific ones are introduced later.

Controls and definitions of terms. The controls that are regularly used in various experiments will be briefly described here.

1. Luminance of the target is controlled by several parameters, which are usually preset before a particular experiment and do not require further adjustment. A direct-coupled beam-unblanking circuit is normally set to give a CRT beam current as large as possible without seriously defocusing the beam. This practice is necessary because the beam is usually on for only a small fraction of the time, and the effective brightness (i.e., luminance perceived by the observer) is proportional to the product of the beam current and the total on-time during a psychological moment (defined later).

2. Image duration, or on-time, is controlled digitally, in steps of powers of two, by a single switch. On-times from 0.25 to 32 ms have been found most useful, although a much wider range could be easily obtained.

3. Image display rate is the number of complete displays per second on the CRT, each display lasting for one image duration. Some displays in fact are made up of many separate (sequential) parts, hence the term applies only to a complete display. Both the image duration and image display rate affect the luminance and brightness, so when brightness is an important parameter there is between beam current in the CRT, duration, and display rate, a three-way interaction that must be controlled. Also, the difference between the reciprocal of the display rate and the image duration is the off-time, which may be an important parameter for apparent-motion experiments.

4. New information frame rate, which will be called simply frame rate when there is no danger of confusion, is the number of times per second that a new image, or picture, or frame of information is presented. It may or may not be the same as the image display rate. In the case of frame repeating, each new frame of information may be repeated several times to increase the display rate (usually to above the critical fusion frequency) while holding the new information rate low. To avoid semantic difficulties, "n-to-one frame repeating" is defined here to mean that each new frame is displayed a *total* of n times. The latter two time parameters (display rate and new information frame rate) are controlled by a "clock," which times the entire apparatus, and by switches, which select the desired submultiple of the clock frequency. In general, the clock frequency is held constant and only discrete values are selected for the various rates. However, since the apparatus is asynchronous and completely controlled by clock pulses, continuous control is available by adjustment of the clock frequency.

5. Target (or trace, or image) position on the CRT is controlled by the usual oscilloscope position and gain controls, and by the amplitude of applied horizontal deflection voltage from a generator. Two types of stimulus generator are used. The analog generator is a so-called function generator, which can produce square, triangular, or sinusoidal waveforms with instantaneously variable frequency. This last property allows nonlinear waveforms to be generated when it is desired to study the effect of acceleration. Otherwise, target velocity will be either constant or sinusoidal, and controllable by both the amplitude and frequency of the deflection voltage. The other generator is a digitally controlled device. A sequence of precision pulses—or "words," in computer language—is integrated to obtain the desired waveform, which can be of much greater complexity than that obtainable from the function generator. In this case, target velocity will be a function of the digital pulse rate, pulse amplitude, and pulse duration, as well as depending on the gain controls. Appropriate control of the word generator allows the simulation of camera exposure or integration times, shutters in the simulated camera or display device, and frame repetition.

6. To allow the display of various background viewing conditions, and to facilitate the display of multiple targets, a dual-beam oscilloscope is used. Two phosphors,

blue and yellow-green, which combine the necessary speed of response, luminance, and resolution, are available.

Hypotheses to be tested

The experiments carried out with the apparatus described in the preceding section were chosen to test the validity of several hypotheses regarding the eye-brain perceptual mechanism, as well as to simulate various proposed television systems. These hypotheses cannot be stated explicitly and concisely, but rather are phenomenological descriptions having some generality. This status is illustrated by the following quotation from Vernon's book on visual perception[2]:

"Although the threshold for movement perception (both for least perceptible speed and for least perceptible distance covered) is actually lower in the fovea than in the peripheral retina, yet form perception is so dim and blurred in the periphery that moving objects appear with relatively greater clarity than do stationary ones. Hence the impression of movement as such, without perception of the object moving, is a characteristic experience in the peripheral vision. Such a movement 'catches our attention' with extreme rapidity and we immediately respond by turning the eyes and head in order to focus the eyes on it and see it clearly."

This characteristic trait of humans, in which we try to fixate on an object of interest (i.e., maintain its image in the central or foveal region of the retina where our resolution for detail is high) leads to another phenomenon, which greatly influences how we interpret what we see. When the object of interest is in motion, our eyes must *track* it in order to achieve fixation. The nature of the eyes' tracking mechanism has been described in considerable detail by Stark, Kupfer, and Young.[3] From a wide variety of experiments they have derived a multiple "black box" model of the visual control system. They give very convincing arguments that only part of the system operates as a continuous servo loop. The tracking mechanism of the eye seems to use a sampled data system. In their model, the brain accepts information from the eye for a short period, and then essentially rejects new information while the muscles ballistically drive the eyeball to the predicted new position to maintain tracking. The intermittency operator that controls this sampling process is supposed to be the accessory vestibular nuclei, and the sampling rate is about five per second. It is also shown by these authors that the eyes' tracking system contains an excellent position-error detector and a rather good velocity predictor, but since there is no element that detects acceleration, changes in velocity are not predicted at all.

The 200-ms period associated with the tracking mechanism of the eye is only one of several time constants that have been identified, though not really verified. One of the other time constants, which has a very reasonable theoretical basis, has been demonstrated in a rather clear-cut experiment by van den Brink,[4] who has measured the loss of visual acuity caused by the smearing on the retina of the image of a moving object. The smearing is a result of retinal summation in which all stimuli occurring within a critical time T are integrated and treated as if simultaneous. Van den Brink obtained a value for T of about 80 ms for images near but outside the fovea (6° from fixation point).

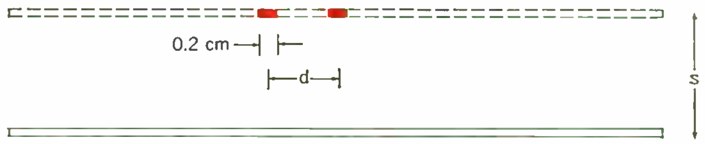


Fig. 1. Double target display on a dual-beam oscilloscope. Upper trace is masked except at spots separated by d, whereas lower trace is visible over entire path.

Another critical time of similar magnitude has been postulated by numerous investigators[5,6] from a wide variety of experiments. This is often called the psychological moment and, as usually proposed, is a deeper and more fundamental time constant associated with the perception process in the brain itself. However, the basic idea is essentially the same as for retinal summation; it is postulated that there is a unit of time, possibly of variable length, during which incoming stimuli are integrated by the brain rather than being treated separately. The stimuli may be tactile, auditory, visual, or sometimes a combination of these. With varying degrees of plausibility, values from 50 to 300 ms are reported. Dr. Paul Kolers of M.I.T. makes an important distinction, however. He believes that for those tasks requiring only the recognition of existence or nonexistence of separate stimuli, the psychological moment lies in the range of 50 to 100 ms. For tasks requiring more analysis by the brain, such as form or shape perception, the psychological moment is about 300 ms. This range of values for a psychological moment leads naturally to the idea of serial processing in the brain. The more complex concepts or percepts are built up from a sequence of elementary ones. However, the physiological source of the psychological moment, if indeed it exists, is very much an open question.

Results and interpretation

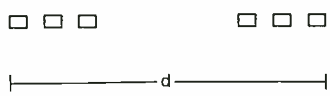
Temporal resolving power of the eye. The first experiment to measure the temporal resolving power of the eye, which should be related to a psychological moment, was suggested by Julesz.[7] He presented the luminous edges of a regular polygon first sequentially, each for a short time, and then in nonsequential order, but with the same on-time. The viewer's task was to distinguish which presentation he was viewing, as a function of time interval between flashes. The minimum time observed for resolution was about 30 ms. In Julesz' display the target appeared stationary. It was natural to ask for this study, "What would the result be with a moving target?"

To answer this question, two moving targets, each a spot of light, were presented on the oscilloscope. The path of each target is shown in Fig. 1.

The upper path, where shown dotted, is covered by a mask. Upper and lower targets were separated by s centimeters. The lower target path was 10 cm long by 0.2 cm high. The oscilloscope was driven by a function generator. The deflection circuitry of the oscilloscope was arranged so that in one case the two targets moved with the same speed, direction, and relative position (in phase), or in the other case with the same speed but in opposite directions (out of phase). The viewer's task was to distinguish whether the upper trace, which could be seen only at a few points, was in phase or out of phase with the lower trace, which was continuously exposed. Variables included the separations d and s, the target velocity v, the viewing distance, and the viewer's orientation—i.e., whether the direction of motion of the traces was in the viewer's left–right direction, or in his head–toe direction. In a preliminary set of tests, the lower reference target was masked, and the viewer's task was simply to determine the direction of motion of the upper target. In all cases, the target sweep was single, initiated by a push button, and separated by at least one second from preceding sweeps. The times observed were calculated simply from $t = d/v$.

When the lower reference beam was absent the determination of direction of sweep of the upper beam was more difficult than when the reference was present. At 50-ms separation of the two flashes, all subjects could determine the direction with at least 75 percent accuracy; at 25 ms, however, few could tell the direction. This situation is basically a static one, similar to that of Julesz, so the similar result is to be expected.

With the lower trace visible, the resolving time of the eye brain is significantly better. The length of the lower trace was usually about five times the separation d, so that even with high velocities (maximum about 200 cm/s) which the eye could not track, the time interval between end points of the sweep was great enough (50 ms) that the direction of the lower beam was obvious. When the viewer's head is in the normal position so that the sweep direction is his left–right direction, the critical time separation of the two flashes for phase determination of the upper trace is about 20 ms. Some viewers could do better with the sweep direction from head to toe. When more information was given, performance improved further. For example, through use of an upper trace mask which exposed two groups of three small rectangles as shown here



and allowing the subjects to practice, about half of the viewers were able to determine the relative phase of the two targets when the time interval corresponding to d was 10 ms. However, the target separation s had to be kept small, below about 4 cm. At larger values of s, or when viewed from farther than the normal 30 cm, temporal resolving power was slightly poorer.

It is concluded from these experiments that if a viewer is given adequate information to perform a relatively simple discrimination, the temporal resolving power of the eye-brain can approach 10 ms, but it is typically longer than this, of the order of 30 ms. The stimuli to be resolved must be separated on the retina to achieve the faster performance. If they are not, the problem becomes one of flicker fusion, for which the critical time at normal luminance levels is of the order of 30 ms (ranging from about 20 to 60 ms).

Psychological (optical) moment for simple tasks. Van den Brink measured the critical time for retinal summation at a given point on the retina, obtaining a value of about 80 ms.[4] If the stimuli fall at different points on the retina, however, physiochemical integration of the stimuli in a retinal receptor cannot take place. Nevertheless, the following experiment illustrates that the brain's handling of temporally *and* spatially separated images shows a certain type of integration effect, which can be called a psychological moment.

For this experiment, a basic digital stimulus generator was used. The target was a vertical line about 1 cm high, which was flashed on for an image duration time of typically 4 ms. (Other durations from 1 to 16 ms had negligible effect on the results.) During the on-time the target did not move; during the off-time, it was moved one unit distance to the right or left. The size of the unit distance was not important, as long as it was greater than the thickness of the line; in this case, ½ cm was used. Whether the line moved right or left between flashes was determined by a random-pulse generator, so the apparent motion of the line was one of a type of random walk. No matter what frame rate of flashing was used, it was impossible for the viewer to track the motion, and, therefore, it was pure coincidence if one image was superimposed on the preceding one. The variable in the experiment was the frame rate.

At low frame rates, below 8 flashes per second, all observers reported seeing discrete flashes, with little sense of connection between them. In the rather narrow range from 13 to 15 flashes per second, a transition occurred such that, for higher rates, observers reported seeing two lines instead of isolated flashes. When 2-to-1 frame repeating was used, so that the target flashed *twice* in a given position before moving, the transition to seeing two lines occurred around a display rate of 28 flashes per second. This is equivalent to 14 new frames per second. Finally, at new frame rates between about 30 and 60 frames per second, observers saw at least two and often three lines simultaneously, depending on the sequence of direction reversals caused by the random-pulse generator.

These frequencies all indicate a psychological moment, for this simple task of counting lines, of about 70 ms. It is tempting, though very likely misleading, to compare this time with the 80-ms retinal summation period of van den Brink. The latter could arise from either the brain or from the retina itself, but the 70-ms period measured by these flashes probably comes from the brain.

Frame repeating and dependence on observer tracking. Under two rather commonly satisfied conditions, the observer's interpretation of what he sees as he views a display containing motion depends strongly upon whether or not he tracks the motion. The difference in interpretation, or perception, may be small or large, as may be the importance of the difference. Nevertheless, the difference is very real and may influence a system design.

The two conditions are: (1) The new information frame rate must not be too large. If the rate exceeds 100 frames per second, corresponding to the 10-ms resolving time of the eye, the display will appear continuous. Most practical systems, however, use new information frame rates between 15 and 30 frames per second, so that this condition is almost always satisfied. (2) The moving object must have a contrasting edge and must be moving at such a velocity that the distance moved between frames is spatially well-resolved by the viewer. Such velocities are commonly seen in communicating humans.

The effects may be isolated and emphasized by use of a target consisting of a luminous line against a dark background to provide the high contrast. If the line is made very thin and is moved at right angles to its length, the spatial resolution is high, and low velocities, which are easy to track, will satisfy condition (2). New frame rate is often the independent variable in an experiment, and this is usually chosen to satisfy condition (1).

The first experiments were done using a modified television system. R. C. Brainard devised an electronic switch to operate with a delay line memory in such a way that a column of dots was generated in the system. By careful adjustment of the delay time, this line of dots could be made to move on the display. However, since only a few rates were available, and phosphor persistence confused the results, most observations were made with an oscilloscope display.

Very few of the effects to be described can be seen under static conditions, so no still pictures can be shown for illustration. However, the illusions seem so real that some observers had to be convinced by either continuously slowing down the display to "see" for themselves what was "really" being displayed, or by taking snapshots of the display.

To facilitate the description of the display, a simple graph is useful, as shown in Fig. 2. Time is plotted on the abscissa, and horizontal position of the target on the ordinate. When the beam is turned on to display the luminous line (the line is vertical, but this dimension will never be shown on the graphs), this will be shown by plotting a point or a line on the graph, depending on the duration of the on-time. Since some on-times are very short and would hardly be visible on the graph, these are stretched slightly to make the graph more legible.

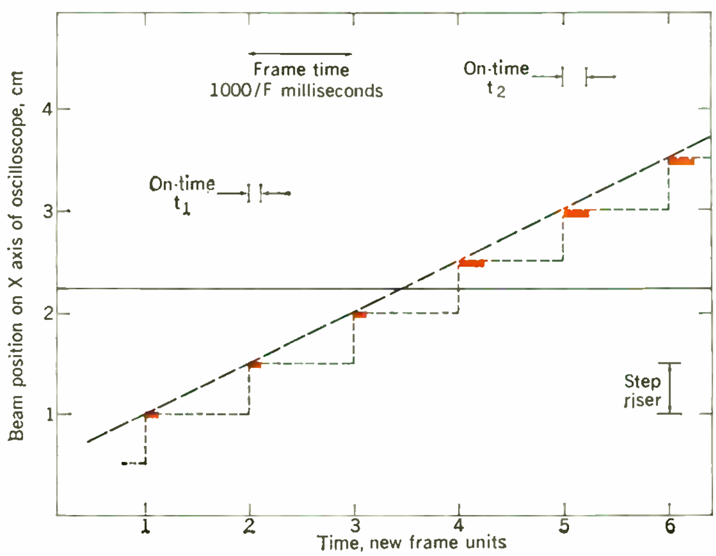
Figure 2 illustrates a target moving with constant apparent velocity. The line at 2¼ cm represents a luminous fixed dot, which serves as a fixation point for the viewer. It is on continuously so that he can keep his eyes fixed in direction. The frame rate is F frames per second, so that the frame time is $(1000/F)$ milliseconds. At the beginning of frames 1, 2, and 3, the beam is turned on for 1/16 of a frame time, indicated here as on-time (or image duration) t_1. For frames 4, 5, and 6, the on-time t_2 is 1/4 of a frame time.

The dotted line is the staircase function generated by a digital stimulus generator. It shows the actual path of the beam, but since the beam is blanked off most of the time, the viewer doesn't know this, and sees only the flashes indicated by the solid bars. The dashed sloping line through the beginning of each flash represents the velocity that an object in the real world would appear to have if illuminated stroboscopically by the flashes. As a specific example, let the frame time be 32 ms, corresponding to a frame rate of about 31 frames per second, which is close to the normal television rate, if we take interlace into account. The apparent velocity is then 15.6 cm per second; the trace crosses the 10-cm CRT face in a little less than one second, so it is easily tracked by the eye.

Suppose an observer fixates on the spot at 2¼ cm. He then receives on his retina a sequence of flashes at equally spaced points. At a viewing distance of 30 cm, the ½-cm step height corresponds to an angle of about 1°. The acuity of the eye drops below 25 percent of that in the fovea when the angle exceeds about 2° from the fovea, so only in frames 2 through 5 does he see a sharp line. In the more peripheral regions he sees blurred flashes, but properly interprets them as in motion. However, the motion is clearly not continuous to him, and he recognizes the discrete positions occupied by the flashes. Furthermore, since the flashes are separated by only 32 ms, well within the 70-ms psychological moment for this type of task, he reports seeing at least two lines at once. Finally, when the target crosses the fixation point and reaches frames 4, 5, 6, etc., he reports that the target suddenly becomes brighter. This is because the image now has a longer on-time at each position, and the retinal summation results in more energy being received within the retinal sumation time of 80 ms.

Next, using exactly the same display, allow the viewer to track the target instead of fixating on a spot. Since the eye has a good velocity predictor, and since the apparent velocity is constant, the eye will "latch onto" the target within three or four frame times and track it accurately. Assuming the target has been tracked for at least a few frames before the one arbitrarily labeled zero in Fig. 2, then the images of the flashes at frames 1, 2, and 3 will fall on the same spot in the retina. Again, these flashes occur within the retinal summation time, so the eye–brain interprets this as a continuously present line (possibly with a slight flicker if the image is bright enough to exceed the critical fusion frequency). Also, because it is tracking, the eye–brain interprets the line as continuously moving as well as continuously visible. However, when the line passes the (now unused) fixation point, it will suddenly appear broader, but not brighter. The formerly sharp line appears to spread into a band. This occurs because the eye's tracking mechanism is capable of following only the apparent or average velocity. The apparent velocity of the line has not changed, so the eye continues to track at its previous velocity. Because the target remains fixed on the CRT during its on-time, the tracking motion of the eye actually spreads out the image on the retina. Since all velocities

Fig. 2. Oscilloscope beam position vs. time for apparent constant velocity. Dotted line shows actual beam trajectory, a staircase function. The viewer sees the beam only during on-times t_1 and t_2.



remain constant, the spreading is constant, so the energy per unit time falling on a given point of the retina, and thus the brightness, remains constant. Instead, the image spreads out, and the observer reports seeing a broader line.

Thus, almost every aspect of what is seen depends upon whether or not the target is tracked. Discontinuity of motion becomes continuity; brightness may or may not vary; and size or width may or may not vary. About the only thing that remains unchanged is that the observer "sees something moving." This is the situation for constant velocity. The interpretation becomes more complex when there is acceleration.

Consider next what is seen under frame repeating conditions. The simplest case, 2-to-1 repetition, is illustrated in Fig. 3. The flash that occurs at the beginning of each new information frame is repeated in the middle of each frame period, at the same position. Each on-time is 1/8 of a frame time. Again assume a new frame time of 32 ms for comparison purposes.

Under nontracking conditions, the observer sees almost exactly the same thing as he did with the display shown in Fig. 2. The only difference is that during frames 1, 2, and 3, the Fig. 3 display appears about log 4 times as bright as the Fig. 2 display. Because of frame repeating and increased on-time for each flash, the luminance is four times larger, and the eye has basically a logarithmic response. During frames 4, 5, and 6, they appear identical, because each position of the flashes is imaged at a separate point on the retina in both cases and the integrated intensity determines the brightness.

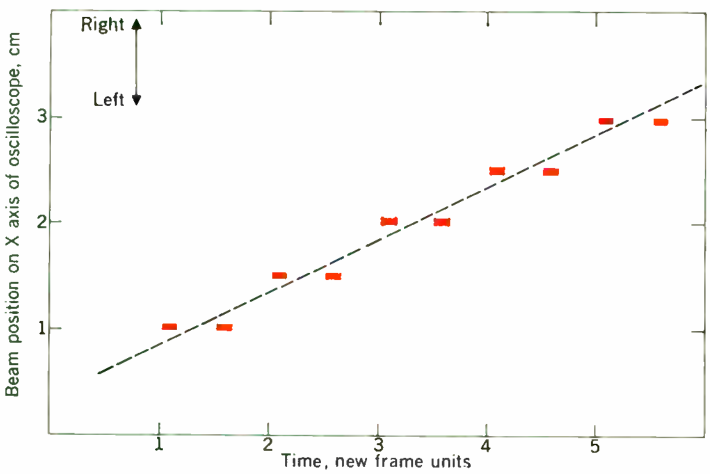
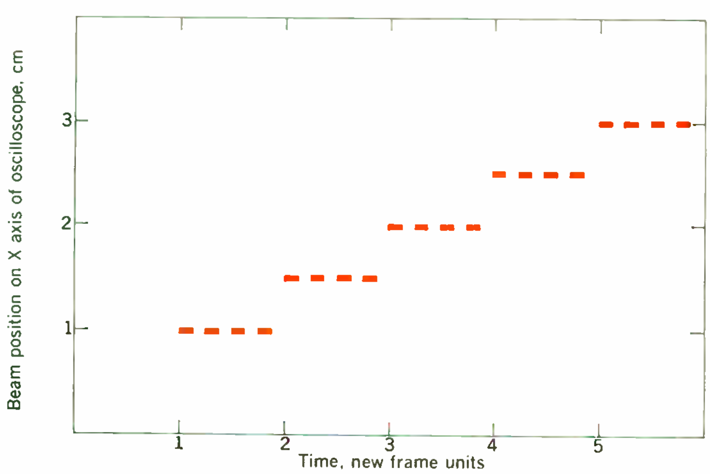


Fig. 3. Oscilloscope beam position vs. time with 2-to-1 frame repeating. The apparent velocity is the same as in Fig. 2.

Fig. 4. Oscilloscope beam position vs. time with 4-to-1 frame repeating and same apparent velocity as in Fig. 2.



Under tracking conditions, however, the appearances are completely different. For the frame-repeated display, the apparent velocity is shown again in Fig. 3 by a dashed line. The tracking mechanism of the eye latches onto this apparent motion quickly and the predictor causes the fixation point to follow the path of the dashed line, or a line parallel to it. Now the first flash of each frame appears slightly to the *right* of the predicted path, so this is seen as a continuously moving, slightly broadened (by the on-time) line or band. The *second* flash of each frame appears slightly to the left of the predicted path, so this is seen as a second, continuously moving, slightly broadened band. The result is that two parallel closely spaced bands are seen moving continuously across the CRT face. This multiple-imaging effect always occurs when the eye tracks a moving frame-repeated target, but does not occur if the target is not tracked.

With 4-to-1 frame repeating, the predictable result occurs as a very compelling illusion; that is, four lines are clearly seen by all observers. But a new effect is seen when the new frame rate is reduced. It is also observed for 2-to-1 frame repeating, but is much more evident at 4-to-1 because of the additional information presented to the eye. Figure 4 shows the display; the (arbitrary) choice of distributing the four flashes equally throughout the frames is quite important, as will be shown later. The effect is most visible at a new frame rate of about 11 frames per second, so assume this value for this discussion.

If the target is not tracked, the observer will again see the same effect as in Figs. 2 and 3, except that the apparent velocity will be smaller, since the beam advances ½ cm in about 91 ms instead of 32 ms. Flicker is negligible, since the effective display rate is 44 flashes per second. However, the motion appears quite jerky. Under tracking conditions, the observer sees a group of four lines moving across the screen. However, these give a strong impression of a rotating barber pole, as if some of the lines were light, some were dark, and the whole pattern rotating.

The following explanation seems both plausible and consistent with other effects discussed previously. Since the frame time of 91 ms is significantly longer than the psychological moment (for flashes) of about 70 ms, the eye-brain does not accept all four flashes in any one moment, but on the average only about three. Hence one of the expected lines will appear dark (or missing). The 70-ms psychological time is certainly not sharply defined, but neither does it shift enough to synchronize with the 91-ms period of the new frames. The missing line is therefore not the same one of the pattern of four in each frame, and the pattern appears to rotate as a barber pole. This phenomenon, in which the brain interprets two-dimensional patterns in terms of three-dimensional objects, is a common one, the source of many illusions.

It is now clear why the uniform distribution of flashes throughout the frame period is important. Were they all grouped together, within say 50 ms, then the pattern of four lines would be seen continuously, without the barber-pole effect. With 2-to-1 frame repeating, the effect is most pronounced at about 24 frames per second, or 12 new frames per second.

Applications to practical systems

Two questions naturally arise: "How do these observations on such a specialized display apply to practical systems of picture display?" and "Do they offer any more than interesting explanations of known phenomena?" The first question can be answered at least in part and the second can at least suggest some precautions in future design.

Film motion pictures typically use a 2-to-1 or 3-to-1 frame repetition to reduce display flicker while simultaneously conserving film. The lowest practical frame rate used is 16 frames per second; and even home movies now have been increased to 18. In his television frame repeating system, F. W. Mounts has found 15 new frames per second to be the lowest "acceptable" rate. Both these rates are just enough larger than the 14-frame-per-second rate, which corresponds to the elementary psychological moment, to avoid the worst of its consequences. It appears that this time is well enough defined for humans that there is little hope of finding lower frame rates useful when we need any significant fidelity in display of motion.

In practical pictures, much more information than just an isolated line is presented. In fact, isolated lines are seldom prominent, although high-contrast edges are. Because the brain is so busy handling the vast amount of information in a typical picture, the effects described in the preceding section are seldom dominant in motion display, but rather lead to a vague dissatisfaction, a feeling that something is wrong. Only in an unusual situation, such as wagon wheels appearing to rotate in the wrong sense, is our perception jolted into conscious annoyance at the illusion being broken. But when a transmission-display system is pushed to the toleration limit, the result is usually that most of the possible margins of acceptance are on the verge of being broken simultaneously. In this case, the variation in subject matter being displayed becomes all-important, for it determines how many limits will be exceeded simultaneously per unit of time.

High-contrast moving edges become seriously reduced in contrast by the phenomenon of edge flicker.[1] Edge flicker is seen at high-contrast moving edges when the new information frame rate is less than the critical fusion frequency, no matter now high the display rate may be. Because of the psychological moment again, edge flicker will be most pronounced and annoying in the vicinity of 12 new frames per second. This is probably associated with an interference effect because of the similar values of the frame period and the psychological moment.

One effect of accurate tracking by the observer does call for care in system design. No matter how sharp each image from the system's camera and display device may be, if the image *display* time is relatively long (say greater than 10 ms) then accurate tracking of a fast-moving target will smear the retinal image and result in blur. The only way to avoid this problem is to use short display times. In television, however, luminance requirements usually make short display times impractical.

Accelerated motion, which is so common in humans, causes another annoying effect. The observer is quite satisfied with the sharply resolved display of a still object. He can become accustomed to the uniform blur of an object moving with constant velocity. But with accelerated motions, the target is constantly changing from a sharply resolved one to a blurred one, and some observers find this disconcerting. As an extreme example, consider the oscilloscope display of an isolated luminous line, using 4-to-1 frame repeating. Under acceleration, this changes from a sharp single line to a group of four moving lines if we track it—and without a single drink.

Several experiments were carried out to test suggestions for television systems, but the results were not encouraging. For example, an advanced digital generator was used to simulate a shutter in front of a storage-type camera, such as a vidicon. It had been suggested that by breaking up a single (blurred) image into several (sharper) images, some improvement in motion rendition could be achieved. There was no indication of such improvement with the simulator, and little if any improvement was observed when a real shutter in front of a vidicon was tested with real pictures. Another approach—that of integrating several successive frames and then displaying them by frame repeating—was simulated. Again, little benefit was found, the major effect being simply increased blurring rather than improvement of motion rendition.

A related experiment, already planned with colleagues, should be quite useful. It is proposed to operate the simulator display equipment in tandem with a television system, so as to supply accurately known targets to the vidicon. The effect of the vidicon memory can then be evaluated fairly easily and accurately by comparing the final television display with the known targets used to expose the vidicon. This technique should also give information about the tandem operation of systems having multiple time constants; the television–eye–brain system is just such a chain. It is important to find out, for example, how various distortions combine when they are passed through two such apparently nonlinear systems.

The suggestions and encouragement of C. C. Cutler and W. T. Wintringham have been most helpful. R. C. Brainard and F. W. Mounts have been excellent instructors in the design and use of logic building blocks. And all of Dept. 1355, Bell Telephone Laboratories, have been generous in exposing the secrets of their eye–brains to the illusions of the oscilloscope.

REFERENCES

1. Eilenberger, R. L., Hempstead, C. F., and Prasada, B., "Study of motion rendition in television using movie simulation," Tech. Memo. MM-65-1355-19, Bell Telephone Laboratories, Murray Hill, N.J., Nov. 15, 1965.

2. Vernon, M. D., *A Further Study of Visual Perception*. New York: Cambridge, 1952.

3. Stark, L., Kupfer, C., Young, L. R., "Physiology of the visual control system," NASA Rept. CR-238, June 1965.

4. van den Brink, G., "The visibility of details of a moving object," *Acta Electronica*, vol. 7, nos. 1 and 2, 1957–1958.

5. Stroud, J. M., "The fine structure of psychological time," in *Information Theory in Psychology*, H. Quastler, ed. New York: Free Press, 1956.

6. White, C. T., "Temporal numerosity and the psychological unit of duration," *Psychol. Monographs*, vol. 77, no. 575, 1963.

7. Julesz, B., Unpublished memorandum.

Experience with on-line monitoring in critical illness

A major responsibility of a physician in treating a critically ill patient is to make quick decisions regarding changes in the patient's vital functions. Experience indicates that an effective automated system may prove to be an invaluable ally in this task

C. Vallbona, W. A. Spencer, L. A. Geddes, W. F. Blose, J. Canzoneri, III *Texas Institute for Rehabilitation and Research*

Even more important than the categorical diagnosis of a disease is the physician's evaluation of the pathophysiological changes that occur in his patient. Automated systems are envisioned that will assist him in his task by providing for monitoring of physiological phenomena, facilitating documentation of clinical observations, aiding in the selection of treatment, establishing rapid communication with hospital service areas, and automatically allocating personnel and equipment. At the Texas Institute for Rehabilitation and Research a dual program of analog monitoring and of computer processing of clinical data collected at bedside is presently under way. This system is described and some of the problems involved are pointed out.

One of the most difficult situations confronting the physician in the medical management of critical illness is the need to make quick and safe decisions. The establishment of a categorical diagnosis of a disease is less important than the assessment of the extent of the pathophysiological changes that take place (shock, asphyxia, cardiac failure, etc.). As we gain greater knowledge of the intimate mechanisms of vital functions and of their derangement in disease states, it becomes more and more obvious that changes may occur rapidly—such as a sudden drop in blood pressure—and that significant events may be of extremely short duration, for example, cardiac arrhythmias.

To facilitate the management of life-threatening situations, it would be desirable to provide the physician with relevant information on the events leading to the patient's present condition. Unfortunately, this information is not always readily available. Extensive use of electronic instrumentation at the bedside has allowed for the recording of some physiological events, but the physician's or nurse's capability of keeping track of other concomitant occurrences is sorely taxed since greater attention must be directed to implementing treatment procedures and to carrying out administrative functions.

System objectives

The experience acquired in the last few years with the use of monitoring systems and computers in hospitals indicates clearly that effective automated systems may be developed to assist the physician in the management of critical illness. Their basic objectives would be

1. To provide for recording and display of physiological phenomena and to establish their significance.
2. To facilitate the documentation of clinical observations and the logging of events.
3. To aid in the selection of treatments.
4. To establish rapid communication with special hospital service areas.
5. To schedule the allocation of personnel and resources available to the intensive care teams.

These objectives are discussed more fully in the following.

Recording and display. Adequate monitoring of physiological phenomena is essential in acute life-threatening situations because changes in vital functions may occur rapidly and unexpectedly. It is necessary that the physician and other members of the intensive care team become aware of the occurrence of these compromising events if corrective action is to be taken. The traditional vital signs—temperature, pulse, respiration, and blood pressure—provide useful information on the activities of vital functions, but in specific instances it may be advantageous to monitor other physiological phenomena, such as electrocardiograms, electroencephalograms, or gastric motility; or to provide for rapid and frequent measurements of biochemical variables.

A complete automation system must provide capability for continuous or intermittent monitoring of these variables, conversion from analog to digital representation, processing of the data, and on-line displays either of the values directly measured, as with readings of systolic or diastolic blood pressure, or of those derived from numerical transformation of the variables recorded, as with values of rate of change of mean blood pressure. It is

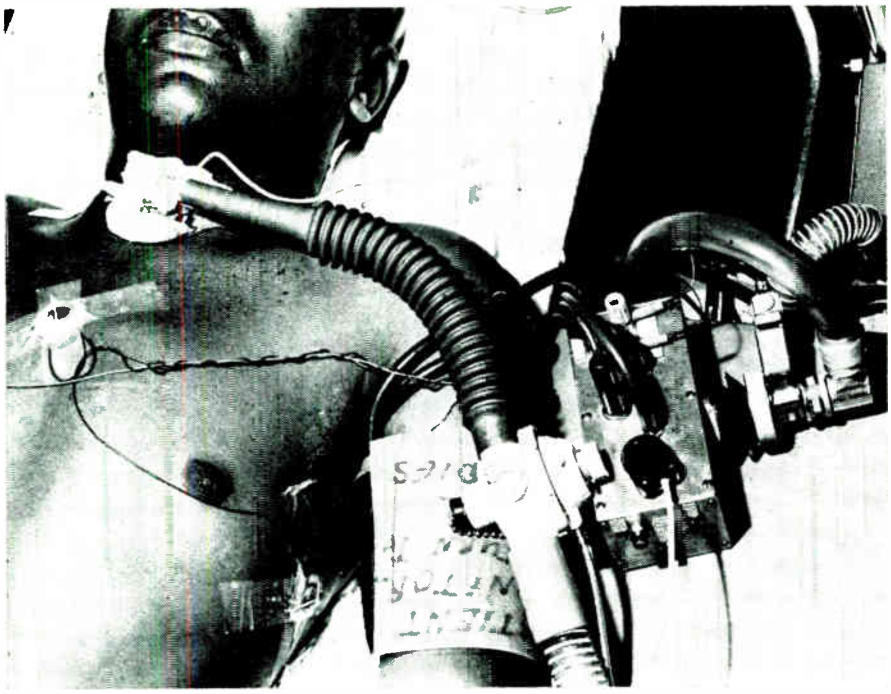


Fig. 1. Electrodes and sensor devices used to detect physiological data on critically ill patients.

desirable that the system establish the degree of significance of the recorded changes and that an appropriate alert signal be given when warranted.

At present, important advances have been achieved with monitoring systems, specifically with those that concern the display of analog data, but on-line connection of these monitoring systems with a digital or hybrid computer has not yet found widespread application.

Documentation. The documentation of clinical observations and the logging of events are of great importance for a retrospective analysis of the clinical situation. Several kinds of observations and numerous types of events may be entered at variable frequency in the course of a patient's critical illness, and it is necessary that these observations and events be classified adequately and presented chronologically in a readable and didactic form. The type of illness, the condition of the patient, and the number of professionals who assist him determine the frequency of these retrospective analyses and the period that they should cover.

Significant experience is being gathered in the use of computers for medical record keeping[1] although few projects deal with the problems of processing the information that originates directly at the bedside.[2]

Selecting treatments. The selection of adequate treatment is a crucial objective in critical illness. Although there is potential application of computer systems for the automatic control of pathophysiological situations, little is known about the requirements and measurements of effectiveness of these systems. Some attempts have been made in this field but they have been limited to the control of specific functions such as the use of anesthesia and artificial respiration.

Until solutions have been found for the numerous problems encountered in this type of activity, automatic systems can be utilized on a smaller scale to assist the physician in selecting adequate medication schedules for his patients and to relieve him of the tedious task of calculating daily fluid and electrolyte requirements for the critically ill.

Communication. Rapid communication is essential between the bedside of a critically ill patient and the parts of the hospital—laboratory, X-ray department, pharmacy—that provide specific services. This is an area of a hospital's functions in which automation systems can be of great value. Unfortunately, the network of communications is extremely complex and does not yield to ready implementation of the well-established computer techniques that have been so effective in industry and business. However, current efforts at the Massachusetts General Hospital,[3] the University of Missouri Medical Center,[4] the Children's Hospital of Akron, Ohio,[5] and other institutions show promise of increasing the efficiency of hospital operations, especially in the intensive care units.

Allocation of resources. Immediate availability of personnel and of special equipment may be urgently needed by the intensive care team, and effective communication with the hospital areas responsible for these services should facilitate their rapid provision. Techniques for the automatic scheduling of personnel and for allocation of resources must be implemented in hospitals to achieve maximum utilization of facilities.

Experience at the Texas Institute

An ongoing investigation at the Texas Institute for Rehabilitation and Research on the usefulness of monitors and computers in the management of critical illness had its origin in 1954 when a system was developed to record analog physiological data in patients in life-threatening situations as a result of poliomyelitis.[6] In 1959, we began to utilize simple electronic data-processing techniques to collect, encode in digital form, store, and retrieve medical information collected at bedside.[7] In the last two years, we have merged this dual program of analog monitoring and of computer processing of clinical bedside data. The program is still in the investigative stage, but numerous elements, including bedside monitoring, reporting of laboratory data, and processing of parts of the nurses' notes, have been in use for several years.[8]

Our experience has evolved from testing in the real situation of an intensive care unit of the Institute (the General Clinical Research Center for Chronic Illness) where chronically ill patients are admitted for clinical investigation. Teleprocessing techniques have been used in the past 12 months to evaluate the usefulness of on-line computer systems. At present, our system has aimed at fulfilling to a great extent the first two objectives indicated previously, and we have begun an evaluation of the system in fulfillment of the third objective.

The monitoring system

Since our report at the 1965 National Telemetering Conference,[9] we have developed and tested a third model of the bedside monitor. Most of the improvements in the new model were made to provide greater reliability for the system and to assure simplicity of operation. Physiological data acquired by this monitor include the body temperature, electrocardiogram, impedance pneumogram, and indirect blood pressure by electrosphygmography.

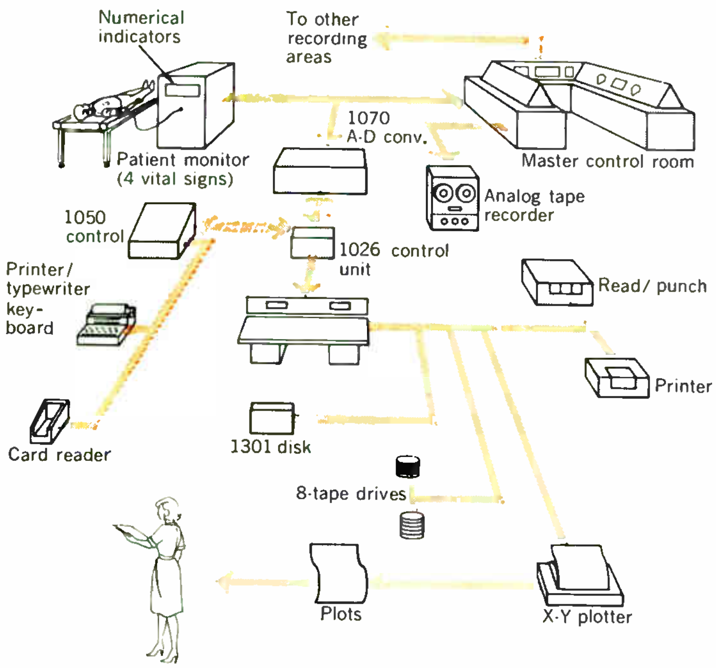
The sensor devices are three electrodes, a thermistor, and a piezoelectric microphone. Two electrodes applied across the chest detect the electrocardiogram and the impedance pneumogram, and a third electrode applied over the manubrium of the sternum serves as a ground electrode (Fig. 1). A thermistor inserted in the rectum allows the measurement of body temperature. The piezoelectric crystal is placed over the area of the brachial artery and underneath the lower edge of a blood pressure cuff applied in the usual fashion. This cuff is automatically inflated by means of a pump and the crystal microphone registers the Korotkoff sounds during the cuff's deflation. The electrodes and sensor devices are connected to the monitor by an eight-foot cable, which is surrounded by a flexible sealed plastic tube that serves as a pneumatic conduit between the blood pressure cuff and the pump contained in the monitoring unit.

A timer located in the back part of the monitor is used to program the frequency and duration of the periods of recording on analog magnetic tape. Activation of the tape recorder also results in setting up the blood pressure pump every 30 seconds. A limit switch turns the pump system motor off when a peak pressure of 220 millimeters of mercury (for an adult) is reached; a constant leak allows for linear bleed-off of the cuff pressure over a period of 15 seconds. The blood pressure pump can also be activated on demand by pressing a button switch on the front panel. The front panel of the monitor unit accommodates the meters that display digital values of temperature, instantaneous heart rate, respiratory rate, and blood pressure, and a lid in the panel can be opened to show a one-beam oscilloscope that displays either the electrocardiogram, the pneumogram, or the electrosphygmogram. The preamplifiers, amplifiers, and ohmmeters (to check the conductivity through all sensors and the patient cable) are accessible through a front door of the monitor.

The monitor is connected to an interface unit that houses frequency-to-voltage converters. Sampling and holding devices are used for readings of instantaneous heart rate and respiratory frequency, as well as for readings of blood pressure (systolic and diastolic) and temperature. The interface unit is connected directly to an IBM 1070 multiplexer and analog-to-digital converter. The digitized data are then transmitted by telephone line, using serial pulse coded modulation, to an IBM 1026 control processing unit attached to an IBM 1410 digital computer installed in the Biomathematics Laboratory of Baylor University College of Medicine, about 3000 feet away from the Institute. The numerical results processed by the computer are then retransmitted to the bedside and printed on an IBM 1050 unit; Fig. 2. Limits of acceptance of vital signs can be set for each patient at any time by sending the proper message to the computer.

All the patient wards of the Institute have panels into which the monitors can be plugged; thus, analog data recorded at any bed are transmitted by direct-wire telemetry to the Institute's Center for Vital Studies and from there to the Physiology Laboratory of the Baylor University College of Medicine. At the Center for Vital Studies the electrocardiogram, pneumogram, electrosphygmogram, and cardiotachrogram are displayed in a four-channel direct analog recorder and stored on a seven-channel magnetic tape.

Fig. 2. Diagram of the monitoring and computer system developed at the General Clinical Research Center for Chronic Illness for processing clinical research data.



Documentation of bedside data

Entries of clinical observations or events occurring at the bedside are made in a transmitter–receiver unit installed at the nurses' station of the General Clinical Research Center for Chronic Illness. Prepunched cards containing specific queries are used to allow the nursing personnel to enter the data in a sequence acceptable for processing and without need for coding. Repetitive data, such as regularly scheduled medications and dosages, are entered automatically by inserting into the transmitter a prepunched card that contains all the information, including name of drug, dosage, and route of administration. The data acceptable for computer processing are classified according to: (1) vital signs; (2) medications; (3) liquid and solid intake and excreta; (4) physical data, including height, weight, and position of the patient; (5) clinical observations—signs and symptoms; (6) artificial respiration devices; and (7) nursing and treatment procedures. The total number of entries of critically ill patients may range from 20 to 75 records for an eight-hour shift, with each record containing from 50 to 120 alphabetical and numerical characters. This number does not include the entries of vital sign data obtained directly from the monitor (one record every minute). Since the transmitter unit is on-line with the computer, automatic editing and correction of the entered data are possible. Immediate filing of the data and on-line retrieval also are accomplished.

The system allows for on-line preparation of chronological listings of all entries made over a prespecified period of time; ordinarily these reports are prepared on a batch basis at the end of each nursing shift and at the end of a 24-hour period. In addition, summary reports are prepared that contain averages and standard deviations of the digital values of the vital signs, totals of volumes of intake and output, times spent by the patient in different positions, and a list of medications administered that also indicates the doses and the starting date for each; Fig. 3.

Graphic plotting of vital signs can be achieved automatically by means of a plotter connected on-line with the computer. These plots can be prepared on a variable time axis so that trends and fluctuations of vital signs over short or long periods of time can be visualized; Fig. 4.

At the Center for Vital Studies, a transmitter–receiver unit is installed for teleprocessing the results of pulmonary function studies. In this application, spirometric readings and instrumental readings of gas test meters can be entered on prepunched cards. A computer program allows for automatic computation of the results and for an interpretative analysis of the significance of the results.

Laboratory reports are prepared in our institution by a computer on a batch-processing basis. Implementation of an on-line capability to retrieve laboratory data at the bedside is under consideration.

On-line retrieval of the clinical data collected at the bedside is possible but has not yet proved practical for medical management although it has been extensively used for clinical research purposes. A major problem that arises is the need for a complex code to specify each type and subtype of entry to be retrieved. The slow speed of the printer device is also a major deterrent.

Automatic selection of treatment

Our experience in the field of automatic selection of treatment has been limited to the development, testing, and implementation of an on-line computer program to calculate the fluid and electrolyte needs of critically ill patients. This is an important function because therapy must be ascertained for each patient each time it is ordered and tedious calculation of estimates is required. Since there is ample background of theory and observations for the formulation of more or less general expressions for water and electrolyte requirements, it is possible to measure safe limits of parenteral fluid therapy based on the patient's body surface area. Specific formulas are used in our program to calculate the volume of fluids required according to estimates of insensible water loss, measurements of gastrointestinal losses, volume of urine excreted, and measurements of other fluid losses such as bleeding.

The computer report expresses the total amount of fluid, glucose, sodium, and potassium to be administered to the patient in 24 hours. It also indicates the number of bottles and the precise composition of standard solutions of fluids to be administered. The rate of administration is clearly expressed in milliliters per hour and drops per minute according to the size of the infusion set; Fig. 5. The recommended therapy for a 24-hour period can be modified at any time upon entry of the latest measurements of fluid losses.

Problems

The experience acquired at our Institute indicates that progress in hospital automation systems, especially for the management of critical illness, is contingent on the availability of adequate systems for entering data in the computer and for displaying computer-prepared results. Typewritten alphabetical or numerical information is not as acceptable to physicians as are graphic displays of the same data. Some of the automatic plotting devices currently available are too slow and thus are uneconomical for use on-line with the computer. Cathode-ray-tube display devices should prove very useful because of their ability to project instantaneously graphic, alphabetic, and numeric information, and light-pen input devices should prove more valuable than the present typewriter input transmitter units.

A second problem area arises because there are too many strict program requirements on a system for processing clinical data collected at the bedside. The requirements include peripheral processing, multiple programming, re-entry capability, dynamic program

Fig. 3. Computer report of the summary of clinical data collected in a 24-hour period on a critically ill patient.

```
..............................   DAILY SUMMARY   ..............................

                         ........  VITAL SIGNS  ........
              MEAN    ST. D.   N      RANGE
TEMPERATURE   98.4     4.7   22     95.0 TO 108.0
PULSE         82.3    10.2  110     64.0 TO 150.0
RESPIRATION   17.4     4.0   90      5.0 TO  27.0
SYST BL.P    118.2     6.5  111    100.0 TO 140.0
DIAST BL.P    79.5     6.4  111     50.0 TO  94.0

                         ........    DIET    ........
REGULAR   SOFT   SPECIAL   LIQUID   TOTAL   AVGE APPET
 ---       ---      3       ---       3        FAIR

                       ........   INTAKE   ........
 LIQUID       IV      BLOOD      OTHER      TOTAL
  ----       1990      ----       ----      1990   CC

                       ........   OUTPUT   ........
     LIQUID                                            SOLID
  URINE  TR.SUCT.      OTHER     TOTAL         NO. BM   AV. SIZE   AV. TEXT.
   1640    ----         195       1835            3      SMALL       LOOSE

                      ........  MEDICATIONS  ........
NAME                  TOTAL DSG   UNITS    N     AVGE DSG   STARTING DATE
KCL                       40.00   MEQ      2        20.00      02/18/66
SOLU B                     5.00   CC       1         5.00      02/22/66
ATROPINE                   1.60   MG       4          .40      02/19/66
PRO PENICILLIN       2400000.00   UNITS    4    600000.00      02/21/66
PROP GLYCOL + N           16.00   CC       4         4.00      02/21/66
VAPONEFRIN                 1.00   CC       2          .50      02/21/66
STREPTOMYCIN               1.00   GRAMS    2          .50      02/21/66
ARAMINE                   10.00   CC       1        10.00      02/22/66
DULCOLAX                   1.00   SUPP     1         1.00      02/22/66

                     ........  RESPIRATOR  ........
                                         HRS MIN
                      AIRSHIELDS          23 59
                      ...TOTAL            23 59


     PT NO.          PT. NAME                                  DATE
     169525          S[redacted] OSCAR E.                      02/22/66
```

relocation, accommodation of tasks in a priority sequence, fail-safe capability (ability to recover in the event of breakdown), multiple process capability, acceptability of tasks from several job streams, including multiple remote stations, and ability to support diverse input–output and storage devices without placing an additional burden on the programmer. The new generation of digital computers and the availability of high-level computer languages should help to solve such problems.

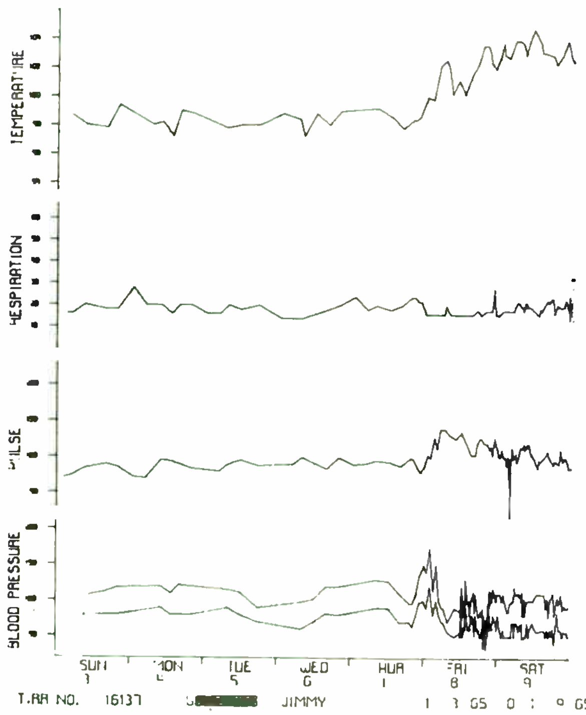


Fig. 4. Graphic display of the vital signs data collected over a week on a patient who developed critical illness in the last three days of the week. Plot is automatically prepared by the computer using on-line data from the monitor.

Fig. 5. Computer report of the fluid and electrolyte requirements for an acutely ill patient.

	BASIC FORMULA	SP. GRAV.	DEHYD.
SWEATING	MODERATE	1015	NONE
G.I. LOSS	500.0 CC		
OTHER LOSS	.0 CC		
CALC. REQMTS. FOR 24 HR.			
FLUID	3208 CC	3073 CC	2459 CC
GLUCOSE	196 GM	210 GM	168 GM
NACL	103 MEQ	105 MEQ	84 MEQ
KCL	123 MEQ	118 MEQ	118 MEQ

RECOMMENDED AMTS. BASED ON
BASIC FORMULA
SP. GRAV.

		CONC. OF SALINE	ADD KCL	DROPS/MIN VENOPAK
1	1000 CC	1/4	40 MEQ	45
2	1000 CC	1/4	40 MEQ	45
3	1000 CC	1/4	40 MEQ	45
4	250 CC	1/4	10 MEQ	45

TOTAL FLUID	EXCESS NACL	EXCESS KCL	CC/HR
= 3250 CC	= 7 MEQ	= 7 MEQ	= 135

SEX	AGE YR.	HEIGHT IN.	HEIGHT CM.	WEIGHT LB.	WEIGHT KG.	BODY SURFACE AREA - SQ.M.
M	37.0	62.0	157.4	148.0	67.1	1.682

HOSP. CODE	DATE	PATIENT	NAME
1	51765	16240	S[redacted], JOSEPH

One major inconvenience with our system has been the inadequacy of the teleprocessing equipment. The system has been adequate for the development of concepts and for maintaining a reasonable amount of remote and central activity, however, it requires a large manual backout procedure and is not readily accessible to a great number of remote stations. The resulting delay is costly when one is attempting to meet a timed schedule for acceptance of digitized collected information from the bedside monitor. Another major disadvantage of the system is the lack of memory protection. Equipment envisioned for our next phase of operation will provide a tenfold increase in the process time in addition to other improvements such as memory protection, priority sequencing, and additional display controls.

At the 1965 Telemetering Conference we reported on the problems encountered with our monitoring system.[9] The majority of these problems have been solved but the need persists to develop analytic techniques for computer processing of data collected at the bedside. As knowledge is gained concerning the significance of the specific patterns of change in vital signs, we should be able to establish the most suitable statistical models to analyze these data (slopes of changes, autocorrelation coefficients, etc.) and the most effective way to display the results at the bedside for utilization by the physician or the intensive care unit team.

Conclusions

It has been shown that continuous monitoring of vital signs and easy retrieval of clinical data must be provided to assist the physician. To achieve this, a physiological monitor and a bedside input terminal are connected with a computer. Vital signs are then merged with other data; the reports obtained after processing this information are available on demand. The automatic computation of fluid and electrolyte therapy is thereby accomplished.

This article is based on a paper presented at the 1966 National Telemetering Conference in Boston, Mass., May 10–12.

The authors appreciate the cooperation of Dr. G. M. Harrison, Dr. P. Harrington, Dr. R. E. Carter, Miss M. McKee, R.N., Mrs. L. Amerine, R.N., and T. Williams. Miss R. Rodrigues, T. O. Townsend, and T. Coulter assisted in the monitoring aspects of the project. Mrs. D. Bellis, Mrs. C. Meigs, S. Pope, B. Hobbs, J. Donovan, and R. Baker participated in the computer programming aspects. The assistance provided by Mrs. S. Gotcher and Mrs. E. Stallworth is gratefully acknowledged.

REFERENCES

1. Schenthal, J. E., Sweeney, J. W., Nettleton, W. J., Jr., and Yoder, R. D., "Clinical application of electronic data processing apparatus. III. System for processing of medical records," *J. Am. Med. Assoc.*, vol. 186, pp. 101–105, 1963.
2. Vallbona, C., "Processing medical information at the bedside," Proc. 4th IBM Medical Symposium, 1962, p. 405.
3. Baruch, J. J., and Barnett, G. O., "Joint venture at Massachusetts General," *Datamation*, pp. 29–33, Dec. 1965.
4. Lindberg, D. A. B., "Electronic processing and transmission of clinical laboratory data," *Missouri Med.*, p. 296, Apr. 1962.
5. Campbell, C. M., "Akron speeds information system slowly," *Mod. Hosp.*, Apr. 1965.
6. Hoff, H. E., Geddes, L. A., and Spencer, W. A., "The physiograph—An instrument in teaching physiology," *J. Med. Educ.*, vol. 32, p. 181, 1957.
7. Spencer, W. A., and Vallbona, C., "Digitation of clinical and research data in serial evaluation of disease processes," *IRE Trans. on Medical Electronics*, vol. ME-7, pp. 296–308, Oct. 1960.
8. Vallbona, C., "Application of computers for hospital usage," *J. Chronic Diseases* (in press).
9. Vallbona, C., Geddes, L. A., Harrison, G. M., Hoff, H. E., and Spencer, W. A., "Physiological monitor: experience, clinical value and problems," Proc. 1965 Natl. Telemetering Conf., pp. 126–129.

Reactive power control in a metro system

In large metropolitan power systems, the magnitude of reactive power can become so great that its control—and the control of system voltage—can demand more than the ordinary requirements in equipment and operating techniques

W. J. Balet, R. L. Webb *Consolidated Edison Company of New York, Inc.*

The reactive power problem will become intensified in the future with the increasing use of EHV overhead lines and high-voltage cable in densely populated areas. Generators alone, with their underexcited limitations during light load periods, may not have sufficient capability to control voltages in some systems in which substantial amounts of high-voltage cable are used. Turbine generators with automatic voltage regulators, however, can be operated successfully in the underexcited region to minimize the problem of excessive system capacitive reactive loading. Although shunt reactors are an effective solution, economics may suggest other methods. Nevertheless, shunt reactors are helpful in controlling voltage during a system start-up. Finally, more experience with the daily switching of 138-kV and 345-kV circuit breakers is necessary before definite conclusions can be drawn as to the advisability of operating them so frequently.

On the Con Edison system, one of the major problems is an excess of charging megavars from high-voltage cables during light load periods. At first glance this would seem to indicate that, with such excesses of capacitive megavars, few capacitors would be required in the distribution area of the system. This, unfortunately, is not the case since it is desirable and economical to have the reactive power for distribution located as near the load as possible to maintain the megawatt capability of distribution stations and equipment at their maximum values. This requires additional capacitance, therefore, in the form of switchable capacitor banks at strategic locations, even though there may be an excess of charging megavars on the high-voltage systems at the same time. Also, voltage control is usually required to be held within relatively narrow limits in metropolitan areas and this, again, aggravates the overall problem. With the installation of approximately 65 miles of 345-kV cable in the past two years, a very large change has been effected on the Con Edison system.

System conditions

Con Edison is a closely knit, high electric load density metropolitan system, with about 6250-MW generating capacity (one hour peak) for 593 square miles of service area in New York City and adjoining Westchester County (Fig. 1). There are load concentrations in several locations in Manhattan that run as high as 500 MW per square mile.

The system's transmission "backbone" consists of a 345-kV installation of high-capacity cables in New York City and 345-kV overhead lines in Westchester County.

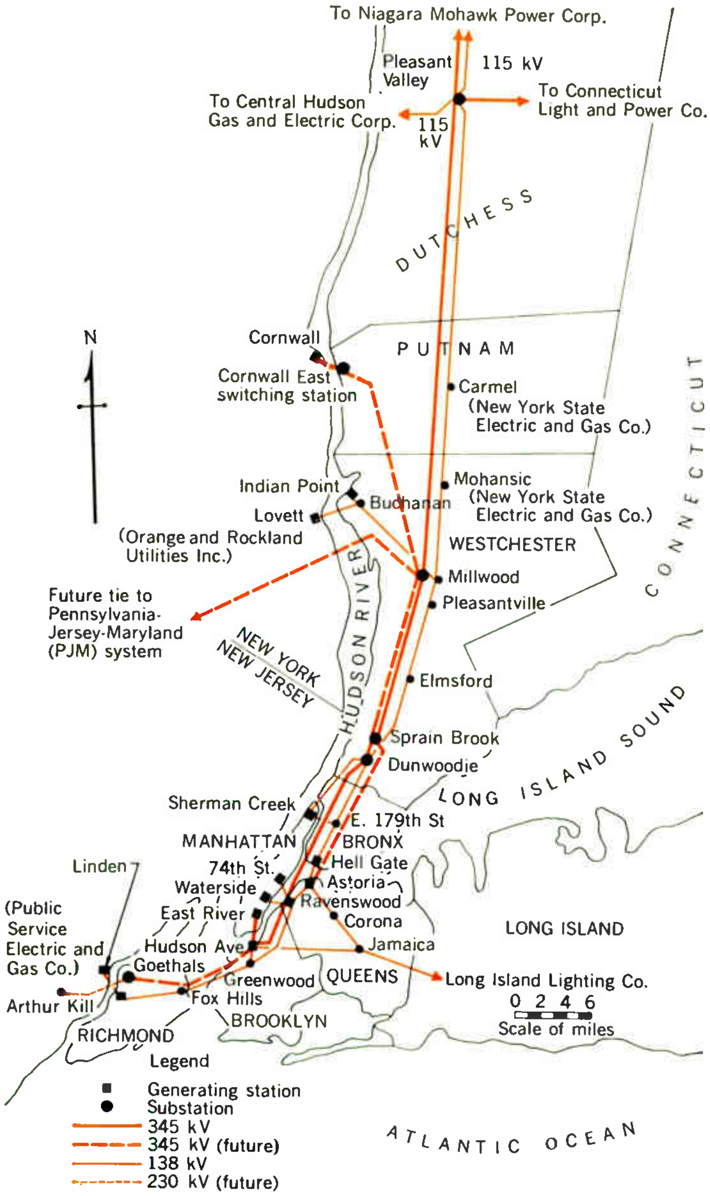


Fig. 1. Geographical arrangement of Con Edison's major 60-Hz transmission system.

There is also a paralleling system at 138 kV. Within New York City limits, both the 138- and 345-kV systems are entirely underground. In addition, in Manhattan and the high load density areas of the other four boroughs, the entire distribution at 27 and 13.8 kV is underground. The preponderance of cable on the system produces a reactive load of about 2700 charging megavars on cable above 33 kV. Although these cable-charging megavars are useful during system peak load conditions, they constitute an engineering and operational problem during light load periods.

The minimum daily load during the off-peak season on the system is about 20 percent of the weekday peak load. On an average day, the minimum load is about 30 percent of the peak load. This wide variation in system daily load creates a problem in the control of system voltage during both heavy and light load periods. During the former periods, about 500 Mvar of switchable shunt capacitors are placed in service both for voltage control on the 4-kV feeders and to reduce the current load on high-voltage cable and substation transformers. During the lighter load periods, the capacitive megavars will far exceed the inductive megavars even though all switched capacitors are disconnected. It is during these periods that control of system voltage becomes an acute problem, requiring the special attention of "engineering and operations" to maintain stable voltage conditions at each voltage level of the system.

The excess capacitive megavars on the system in these light load periods must be absorbed to maintain satisfactory system voltage. This can be accomplished in one—or more—of four ways:

1. Generator operation in the underexcited zone.
2. Use of shunt reactors at strategic locations.
3. Circulation of reactive megavars in selected system circuits.
4. Switching off selected system circuits.

Underexcited generator

For a number of years it has been the practice to operate generators in the underexcited region during light load periods. This is the most important method of absorbing excess system capacitive megavars. To ensure that this can be done without the risk of heating or instability of the generator end-turn area, generator capability curves, as shown in Fig. 2, are provided to operators for all station machines that may be operated during off-peak periods. By the use of these curves, it is possible for the generators to be operated to absorb a maximum number of megavars and for the operators to feel secure in the knowledge that the machines and the system are being operated in a safe and prudent manner. Operating limits in the underexcited region include recognition of the effects of a system disturbance that will thrust sudden load increases upon the unit.

The minimum loads—and hence the greatest surplus of megavars—usually occur in the early morning hours on weekends during the spring and autumn. These are the times when units are most likely to be out of service for scheduled overhaul or weekend maintenance. Thus a number of the largest machines, with the greatest capability of absorbing reactive power, are often not available. The problem is further intensified by the pressure for operating economy that forces some machines, which could absorb sizable capacitive megavars, to be removed from service.

From Table I, it may be seen that, with the two largest machines (Ravenswood nos. 2 and 3) out of service, the ability to absorb megavars is greatly reduced. Should other machines be simultaneously out of service on scheduled overhaul, it will be necessary to switch selected cable circuits out of service, if system voltage is to be adequately controlled.

Shunt reactors

Shunt reactors are a direct means of absorbing excessive charging megavars. By placing them at or near the terminals of high-voltage cable, it is possible not only to absorb the unnecessary megavars but also to maintain a desirable voltage gradient on the 345- and 138-kV transmission systems. At the present time, Con Edison has connected shunt reactors of about 560-Mvar capacity to the tertiary windings of its 345- to 138-kV autotransformers, and reactors of 120-Mvar capacity have been attached to 27-kV substation buses. In all cases, these reactors could be removed from service by switching, as necessary, during heavy load periods.

Eight 13.8-kV, 70-Mvar, outdoor open-type shunt reactors (560-Mvar total capacity) were connected to the tertiary windings of autotransformers. The current rating, and, therefore, the megavar capacity of the reactors, is controlled by the 3000-ampere rating of the 13.8-kV breaker that is used to switch the reactor. In some cases, the phase relationship of the tertiary winding of the autotransformer is such that it also can be used as a source of start-up power for generating units in the vicinity. Because of peculiarities of the secondary network system, there are no plans for using the tertiary windings to supply customer load.

In the two substations in which shunt reactors have been connected to the 27-kV bus, the reactors have been limited in size to 30 Mvar each, so that there will be a voltage variation of less than 2 percent both on the bus and the customers' service when the reactors are switched. At these substations, the same breaker that normally switches the shunt capacitors also switches the shunt reactors. This emphasizes one of the paradoxes of system design: the use of shunt capacitors during the high load period to increase substation megawatt capability, and shunt reactors at night on the same bus for system voltage control.

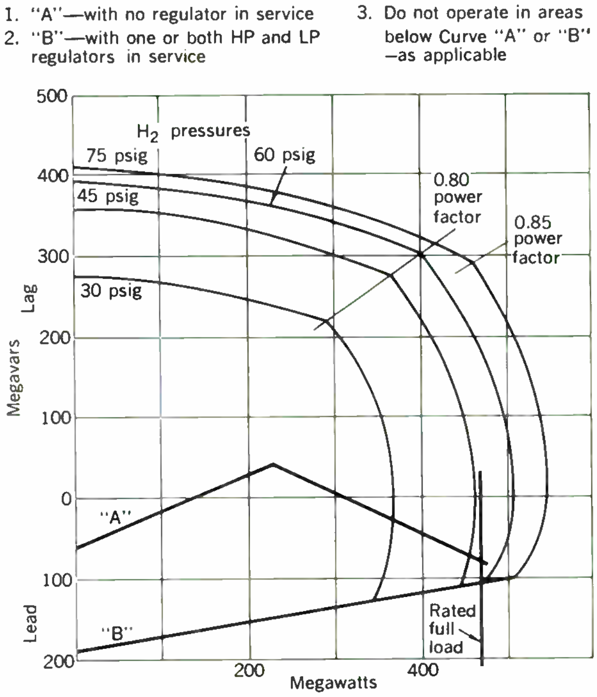


Fig. 2. Turbogenerator capability curves as used for machine operation.

Reactive load and losses

The reactive load and associated inductive losses make up a large "sink" for megavars on the system. The load during the early morning hours, however, has not been growing very rapidly and, therefore, has not matched the peak load growth and the continually increasing charging megavars. The transmission inductive reactive losses and station auxiliary inductive load appear to be relatively independent of system load and are about 600 Mvar on the Con Edison system during early morning hours. Substation and distribution transformer inductive losses, of course, vary with system load, but no breakdown is available for the line, transformer, and station auxiliary losses that make up the total inductive losses into the various components. Efforts to make a generalized calculation of inductive reactive losses have been largely inaccurate; it is more practical to measure the losses at various load levels and to treat them empirically.

Increase of reactive losses by means of controlling devices

Although the transmission system reactive losses do not seem to vary markedly with the load level, on the Con Edison system there are a number of phase-angle regulators (Fig. 3) that control the megawatt flow on the various cable circuits. Also, practically every transformer in the transmission system is equipped with voltage load tap changers. By manipulating these controlling devices, it is possible to force more current to flow than otherwise would occur and thus the inductive megavar losses would be increased.

If all of the four transformers, normally operating in parallel on a major substation bus, had their taps biased so that two would operate at plus 5 percent—and the other two at minus 5 percent—from their schedule, bus voltage would not change, but each transformer would be loaded to 50 percent of its rating. If this were done at all presently available substations, coupled with the effects that may be obtained from the phase-angle regulators, about 200 Mvar in additional losses could be realized. Operating problems associated with this solution are now being studied, but it can be seen that the compensation obtainable in this way is not a total solution to the problem.

De-energizing of cable

To avoid the need for excessive installation of shunt reactors to compensate for all the cable-charging current, another means of controlling the excess reactive losses is to de-energize selected cables during the light load periods. Since, for the most part, each link of the Con Edison high-voltage interconnections is composed of two or more cables in parallel, system reliability would not be seriously affected by certain cable outages. During the early morning hours, most of the cable circuits are

I. Generator absorption capabilities under light load conditions

Station	Generator No.	Underexcited Megavars*
Ravenswood	1, 2	344
Ravenswood	3	335
Astoria	4, 5	348
Astoria	3	167
Astoria	1, 2	184
Indian Point	1	70
East River	7	24
Arthur Kill	2	167
	Total	1639

*At approximately 40 percent of machine rating

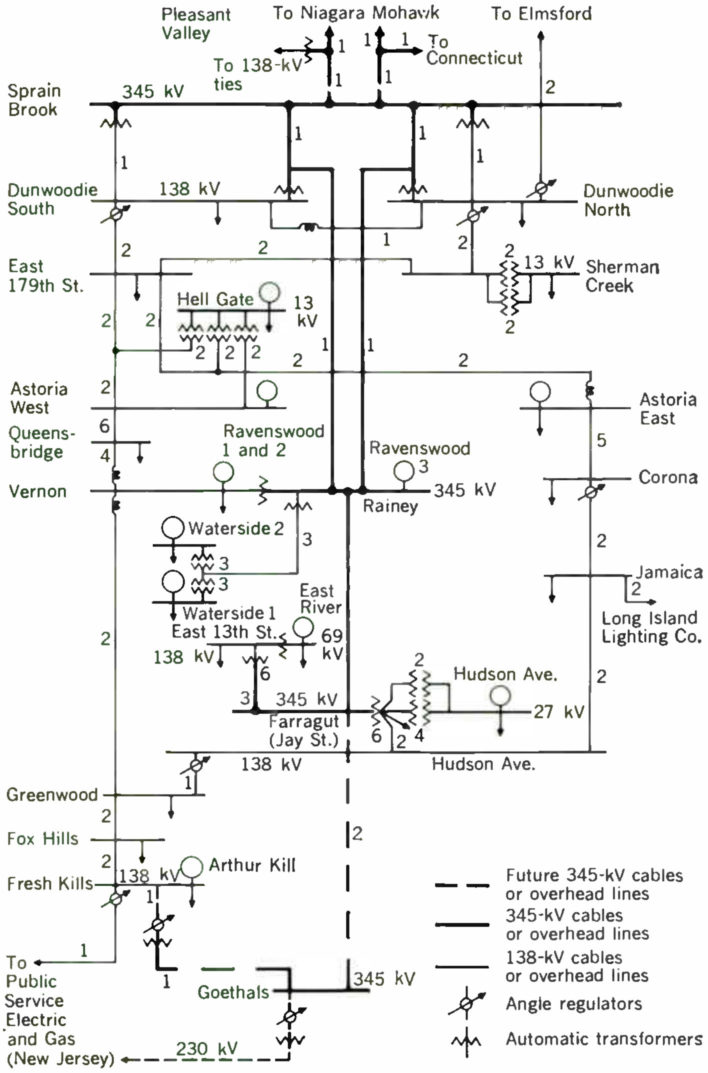


Fig. 3. Block diagram of Con Edison's 60-Hz high-voltage interconnections.

loaded only to a small fraction of their rating; thus there would be little difficulty for the operators to maintain feeder loadings within cable ratings while parallel feeders were out of service. There are, however, two principal disadvantages, which tend to make cable switching less attractive:

Switching surges. With the installation of preinsertion resistors in the air-blast breakers, switching surges are minimized. The maximum surges, even under such severe conditions as reclosing a 345-kV cable on a trapped charge, are no more than 2.3 times the normal voltage. Although this is well within lightning arrester ratings, there is some feeling that it would not be wise to subject equipment to switching-surge voltage on a frequently repeated basis.

Circuit-breaker limitations. High-voltage switchgear is not presently designed for daily operation. The ASA standards stipulate that, following 125 load-switching operations, a 345-kV air-blast breaker should be inspected and maintenance work should be done. It further specifies that the mechanical life test of such a breaker is 750 operations. As may be readily seen, the daily switching of cables would greatly increase the necessary maintenance on circuit breakers and could possibly lead to their premature replacement, or, at least, excessive replacement of parts. Neither of these prospects is tolerable from an operational standpoint.

Nevertheless, the severity of these two disadvantages is being further investigated so that an ultimate decision can be made on the feasibility of reactive control by switching cable on a regular daily basis.

Possible future courses of action

As the high-voltage cable mileage on the system becomes greater, the requirements for underexcited reactive generation during light load periods will increase far more rapidly than the operating capability of new generators.

This problem will be further aggravated if generation is installed outside New York City and relatively long cable circuits are necessary to transport the output to the load center. At that time, it will probably be necessary to adopt one—or a suitable combination—of the steps we have discussed. As of now, however, the most suitable way to neutralize large blocks of capacitive megavars is to use shunt reactors.

Installation of shunt reactors

Although the installation of more shunt reactors is the most direct and simple approach to the problem of reactive control, it is also expensive. Con Edison can foresee the day when almost all high-voltage cable will have to be totally compensated. This would increase the effective cost of the cable by about 10 percent, since the cost of the shunt reactors would be directly associated with the installation of the cable.

A decision has been reached recently to compensate fully two new 345-kV cable circuits from Brooklyn to Staten Island. Each of these circuits is 18.3 miles long and requires approximately 300 Mvar for compensation at 345 kV and 60 Hz. The total of 600 inductive megavars will be provided by three 150-Mvar 345-kV shunt reactors and two 13.8-kV shunt reactors connected to an autotransformer tertiary winding.

The capacitive reactance of uncompensated cables caused considerable difficulty in re-energizing the high-voltage tie-feeder system on the night of November 9, 1965. Overvoltages appeared in several areas and this condition required the careful handling of the 138- and 345-kV cable re-energizing until considerable generation and system load had been restored. Fully compensated cables would have been helpful at that time.

This article is based on a paper presented at the IEEE Summer Power Meeting held in New Orleans, La., July 10–15, 1966.

BIBLIOGRAPHY

Concordia, C., Crary, S. B., and Lyons, J. M., "Stability characteristics of turbine generators," *AIEE Trans.*, vol. 57, pp. 732–744, 1938.

Crossman, G. C., Lindemuth, H. F., and Webb, R. L., "Loss-of-field protection for generators," *AIEE Trans.*, vol. 61, pp. 261–266, 1942.

Roberts, W. J., and Webb, R. L., "Operation of turbine generators during off-peak high power factor periods—practices of one utility in a metropolitan area," *AIEE Trans. on Power Apparatus and Systems*, vol. 74, pp. 461–467, 1955.

Webb, R. L., "The electric power system of Consolidated Edison Company of New York, Inc.," *New York Academy of Sciences Trans.*, ser. II, vol. 26, suppl. to no. 8.

Special Conference Report

International Microwave Symposium

David K. Adams *Stanford Research Institute*

The annual symposium on microwave theory and techniques (MTT), sponsored by the IEEE Group on MTT, is a source of keen stimulation to the 500 or more engineers, scientists, and teachers who look forward to this international gathering.

One outstanding feature of the 1966 International Microwave Symposium, which was held at Palo Alto, Calif., on May 16–19, was derived from its proximity to the Stanford Linear Accelerator (SLAC); see Fig. 1. Two guided tours of SLAC, which has been described as "the world's longest traveling-wave tube," were arranged during the symposium. In addition, three papers were presented in an evening technical session on linear particle accelerators. G. A. Loew explained the RF drive system for the Stanford accelerator. P. B. Wilson presented a paper on microwave applications of superconductivity with particular emphasis on accelerator applications where power dissipation could be reduced. J. N. Weaver and R. Alvarez treated accurate phase-length measurements of large microwave networks.

Under the chairmanship of W. W. Mumford of Bell Telephone Laboratories, an evening session was devoted to summaries of microwave research in England, Japan, Scotland, Scandinavia, and Western Europe. The speakers were Prof. Peter Clarricoats of Leeds University, Prof. Nobuaki Kumagai of Osaka University, Jeffrey Collins of Glasgow University, and Michael Brady of NERA A/S in Oslo, Norway.

The banquet address, "Technical Education in Developing Countries," was given by Prof. John Brown of University College, London. During a leave of absence (1962–1965) Prof. Brown established the electrical engineering department at the Indian Institute of Technology in Delhi. He spoke from personal experience regarding the importance of tailoring technical education to each nation's level of industrial development. Often greater emphasis needs to be placed at the technician level. For example, students in India are often unfamiliar with simple tools prior to enrolling in the engineering curriculum; however, he commended their enthusiasm. If teaching is made relevant to the level of technical practice in the country, this enthusiasm will not be lost. Professional contacts by visiting engineers were encouraged by Brown, who recommended that traveling engineers contact the local engineering society in advance.

Readers specializing in other phases of electronics may wonder about the relevance of microwaves to their field of endeavor. Microwave engineers are usually concerned with wavelengths from 10 cm to 1 mm; however, a more informative viewpoint would be to note that microwaves bridge the gap between low-frequency lumped-constant circuit design and optical design involving many wavelengths. As a result the microwave engineer is often challenged to work to an electrical precision approaching that possible at lower frequencies, but with structures several wavelengths in dimension. Today, the microwave spectrum plays a significant role in voice and video communications, high-energy particle accelerators, telemetry, and radar.

One microwave dream, which has continually eluded us so far, has been the use of microwaves in computer design. However, new microwave applications are already opening up in fields as diverse as nondestructive testing, radiant cooking, and pulse code modulation.

The 1966 International Microwave Symposium audience heard, for the first time, contributed papers on bulk negative resistance (the Gunn effect), avalanche transit-time oscillations, and microwave integrated circuits. Other papers brought us up to date on such diverse subjects as frequency translator design, microwave filters, ferrite devices, and optical waveguide techniques.

Solid-state design

As in other phases of electronics, solid-state devices have had an immense impact on microwave system design. Twenty years ago, solid-state microwave components were limited mainly to resistive-diode mixers and detectors. Today, mixing and detecting requirements can be met with half a dozen different solid-state components.

A. M. Cowley made a quantitative comparison of six classes of solid-state microwave detectors. A procedure was explained whereby different detectors can be compared on the basis of noise equivalent power over a 1-Hz video bandwidth as a function of RF and video output frequency.

Although microwave mixers have long been with us, the greatly improved mixer diodes available today require

David K. Adams is a senior research engineer, Electromagnetic Techniques Laboratory, Stanford Research Institute, Menlo Park, Calif.

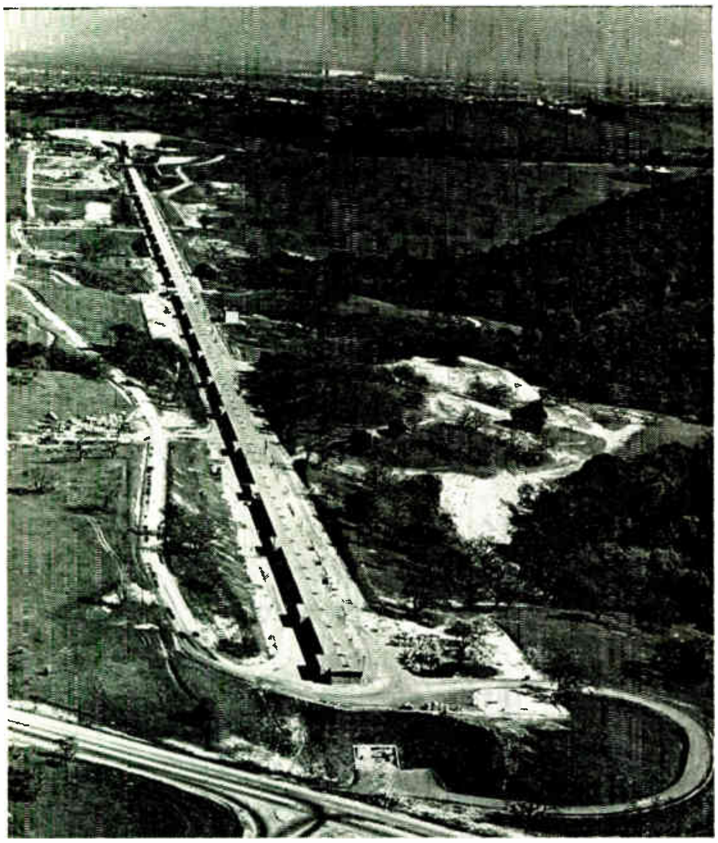


Fig. 1. Aerial photograph of two-mile-long building housing the klystron gallery of the Stanford Linear Accelerator.

the fresh application of mixer analysis. M. R. Barber and R. M. Ryder considered ultimate performance limitations of Schottky diodes in mixers. The value of detailed information about mixer diodes was also evident from the work of J. H. Lepoff and A. M. Cowley, who discussed reduction of intermodulation in microwave mixers.

In addition to papers presented at the session on microwaves abroad, interesting papers were given during the general technical sessions by visiting authors from other countries. L. Lewin summarized recent work done by commercial microwave laboratories in England. Extensive efforts were noted in the areas of bulk negative resistance, parametric and tunnel-diode amplifiers, and frequency multiplication. For example, a new tunnel-diode fabrication technique in use by Standard Telecommunication Laboratories (STL) was described. In this technique, the diode cross section is lapped to reduce its area and to give more control of device parasitics. As a step forward in parametric amplifier operation above 20 GHz, Lewin noted the development of a double-diode package by Ferranti Ltd. The two diodes mounted in a section of waveguide are pumped in parallel, but they provide the idler resonance in series. The speaker also described the G. and E. Bradley work on frequency multipliers to 27 and 36 GHz, from which an output of 50 mW is obtained (tripling with 50 percent efficiency). Another class of multipliers described use ferrites, and have the interesting property of giving not quite times-2 multiplication.

In recent years, each microwave symposium has opened up one or more new areas resulting from newly reported discoveries or device developments. In a timely summary of the ultimate capabilities of solid-state microwave sources, B. DeLoach pointed out that in any semiconductor device, the maximum voltage is determined by sample length, maximum current by saturated carrier velocity, and maximum frequency by the ratio of saturated velocity to sample length. Thus, the product (output power) × (frequency)^2 will not exceed some constant for any given class of microwave source. The constant in this case is best determined by plotting logarithmically power versus frequency for typical devices in the class. These data will define a line falling off approximately with the inverse of frequency squared. Further advances will move some of these lines to higher powers and higher frequencies, but there is an upper limit for each class of device.

Two other papers were concerned with new techniques for microwave generation. B. W. Hakki described a bulk negative-resistance device that works simultaneously as an oscillator, mixer, and IF amplifier. Results on avalanche diode oscillators were reported by F. A. Brand, V. J. Higgins, and J. J. Baranowski, who have obtained 30 mW at 13 GHz with about 5 percent efficiency.

A device that had a similar impact on the microwave community a few years ago is the varactor diode. Useful for frequency multiplication and translation and for parametric amplification, points of interest about varactors continue to come up. Design methods for these varactor circuits are now well known, but more subtle details are emerging. D. B. Leeson discussed instabilities in varactor multiplier chains. Four sources of instability were noted, which produce either spurious oscillation or step discontinuities in frequency response. The importance of using filters with extremely wide stop bands was explained. M. E. Hines and J. G. Ondria discussed the incidence of AM and FM noise in varactor frequency multipliers. They also compared an X-band solid-state multiplier with more conventional magnetron and klystron sources. In this comparison, the solid-state source was superior to the other two types of source in FM noise (up to about 100 kHz from carrier) and superior to the magnetron in AM noise.

A novel frequency-translation system was described by E. Sard, B. Peyton, and S. Okwit. A tunable, varactor up-converter tends to have a low noise figure but poor gain. Therefore, second-stage noise is always significant when an up-converter is used as the input for a broadband receiver. The idea was conceived of using a 6-GHz maser as a low-noise second stage. The entire system was refrigerated and the overall system noise temperature was less than 30°K from 2.0 to 2.4 GHz. Another up-converter design, with tuning range from 300 to 3000 GHz, was described in a paper by F. S. Coale, H. M. Weil, and P. M. LaTourrette, which also gave the design of other broadband frequency translators with novel properties.

Filters and couplers

One phase of microwave research with a long history is microwave filter synthesis and design. In recent years, design methods for maximally flat and equal-ripple bandpass and bandstop filters have been carried to the point where extremely accurate results can be obtained. For example, R. J. Wenzel gave scaling procedures useful for obtaining TEM bandpass filters with reasonable impedance levels. Wenzel's method was applied to a 3.06:1

transformer operating over a 3:1 frequency band.

New challenges continue to present themselves in microwave filter design. With filters having equal ripple in both passband and stopband beginning to emerge, a paper by R. Levy and I. Whiteley gave a general method for realizing distributed elliptic-function filters.

An important element in many microwave systems is the directional coupler; couplers with greater bandwidth and stronger coupling are constantly in demand. E. G. Cristal described the design of directional couplers using coupled lines of unequal characteristic impedances. One advantage of the resulting couplers is that they can act simultaneously as transformers. An alternate coupler design approach was described by C. P. Tresselt, who employed a pair of nonuniform transmission lines for coupling, rather than step discontinuities. Tresselt pointed out that the reactance of step discontinuities degrades coupler directivity, but continuously varying the coupling improves it.

Transistors have climbed into the microwave realm, and now matching and interstage networks require investigation. F. E. Emery, M. O'Hagen, and S. D. Nolte discussed the use of computer programs for the design of microwave transistor amplifiers. The program accounts for changes in transistor properties with power level.

A somewhat similar problem arises in the design of fast microwave switches, calling for very broadband filters both in the RF band and in the drive circuit. R. V. Garver and T. H. Mak addressed themselves to this problem, and pointed out the need to avoid switching transients, which can saturate a receiver if not suppressed.

N. Eberhardt described the use of field displacement in a waveguide for filtering. An absorption filter was discussed in which dielectric was loaded with lossy material to absorb energy coupled into the dielectric. R. W. Dawson and B. C. DeLoach considered the circuit requirements of a fast X-band switch, which employed a newly developed p-i-n diode with an extremely thin "i" region. As a precursor to microwave pulse-code-modulation rates, R. E. Fisher described a 1200-megabit-per-second analog/digital converter.

Integrated circuits

The impact of integrated circuits on the electronics industry is now being felt in the microwave community. In a survey of integrated circuits R. L. Pritchard emphasized the need for innovation, as it is not to be expected that microwave integrated circuits can be successfully made by simply integrating conventional circuits. Pritchard noted that another difficulty in integrating microwave circuits is the useful quarter-wavelength distance often employed in existing circuits. However, the bright future ahead for microwave integrated circuits is signaled by the rapid recent improvement in solid-state microwave components. B. T. Vincent, Jr., described characteristic impedance measurements on ceramic microstrip and a balanced mixer design. Gallium arsenide Schottky barrier diodes were bonded into a hybrid integrated circuit.

The advantages of thin-film circuit technology in making a tunnel-diode device were cited by H. C. Okean. The structure described consists of a beam-lead germanium tunnel diode bonded to the thin-film circuit across a thin-film stabilizing network. The entire circuit is supported by a sapphire substrate. Some interesting points on diode construction for application to hybrid and monolithic integrated circuits were made by C. M. Howell. After forming a Schottky junction on epitaxial silicon or gallium arsenide, the slice is covered with a thin layer of glass encapsulant. Contact is made through an etched hole in the glass by depositing metal in the hole and on top of the glass, which acts as a support. Direct contact to microstrip is now easily accomplished. W. J. Moroney and A. Botka employed p-i-n diodes of similar construction in the fabrication of integrated microwave switches.

Microwave techniques

The competent microwave engineer must be familiar with theoretical and experimental techniques of considerable diversity. Therefore, two sessions were devoted to general microwave techniques, extending from dc through submillimeter wavelengths. For example, the oscilloscope has always been one of the most useful tools in electronics. Until the development of sampling oscilloscopes they were limited typically to the sub-UHF range. Therefore, a paper by W. M. Grove on sampling oscilloscopes from dc through X band was of great interest.

Careful microwave measurement continues to be a subject of importance. High-quality microwave diodes of the varactor type are characterized by cutoff frequencies in the region of hundreds of gigahertz. A comparison of various methods for accomplishing this measurement were given by G. D. Vendelin and S. A. Robinson, who emphasized the usefulness of a reflection measurement for cases where low VSWR (voltage standing-wave ratio) coupling into the diode can be achieved.

Microwaves are often a useful diagnostic tool for measurements in otherwise inaccessible regions. S. Lederman and E. F. Dawson described a microwave technique for measuring electron density and relaxation time behind a reflected shock in a shock tube. A cavity technique was used, with several nearly equal frequencies incident.

A general discussion of microwave sampling cavities was given by R. E. Post and A. G. Potter, who presented data on cavities operating in periodic open-wall TE_{011} and TE_{012} cylindrical-cavity modes. This structure was recommended for measuring the properties of gaseous media. Another microwave plasma measurement was presented by J. B. Chown, W. C. Taylor, and T. Morita, who obtained data on breakdown power levels due to ionization in high-temperature gases. The measurement was made on an X-band slot antenna in the presence of high-temperature air from a shock tube. After the session, an interesting controversy arose over the claim by Post and Potter that the Q of a periodic TE_{011} cavity increases with the fractional wall opening. This question has not yet been resolved.

In the first session on microwave techniques, Prof. John Brown discussed the use of momentum as an analytical tool for microwave boundary value problems. It was noted that when analyzed, some microwave structures have characteristics that suggest simple physical interpretations. It was shown that electromagnetic momentum, often overlooked in microwave analysis, gives added insight into microwave waveguide problems of interest. In another paper, F. C. de Ronde described matching of waveguide discontinuities over a full waveguide band.

The second session on microwave techniques, which was concerned with ultramicrowave and optical techniques, included a paper by E. A. J. Marcatili, who discussed optical transmission media. A comparison was made of dielectric, metallic, and beam waveguides for long-distance optical transmission. Factors contributing to transmission loss in each case were cited. T. Nakahara spoke on applications of TE_{01} mode circular waveguides, leaky waveguides, and beam waveguides. His paper dealt mainly with current work in Japan, including a 4.2-km tandem hybrid waveguide system employing helical waveguides. Another Japanese project is the use of circular waveguides along railroad lines for trunk communication.

The requirements for filters at millimeter wavelengths call for novel techniques. A paper by J. Cohen and J. J. Taub discussed the use of confocal resonators to achieve high-Q filter elements, and reduction of spurious responses. Another basic need at shorter wavelengths is power-measuring devices with good sensitivity and dynamic range. M. Wang and F. Arams described an extremely broadband power-measuring detector for the 300- to 3000-GHz band, with capability for operating over a more extended band. The device has a 50-dB dynamic range and gives linear voltage out as a function of power in. The sensitivity is 240 volts per watt.

As requirements grow for microwave techniques at optical frequencies, consideration is being given to light-beam waveguides. Y. Suematsu, K. Iga, and S. Ito described the lens effect produced when the gas temperature inside a waveguide varies hyperbolically. In another theoretical paper, S. Harrison and W. Kahn analyzed the stability characteristics of an empty curved-mirror optical resonator upon insertion of an inhomogeneous focusing medium. Transfer matrices analogous to those found in electric filter theory were employed.

Ferrite devices

Ferrite materials give nonreciprocal effects and are also used for magnetic tuning. An important application of ferrites today is in variable phase shifters for rapid scanning of large antenna arrays. Generally, one phase shifter is needed for each array element, and linear phase shifts up to about 360° are required. A recent innovation has been the latching phase shifter, which uses remnant magnetization to hold the ferrite permeability at a desired value without applying a holding current. Current pulses are used to change the state of magnetization as required.

Several papers were given on latching ferrite phase shifters and switches. R. Stern and J. Agrios described their design of a latching switch capable of handling 50 kW peak and operating between 33.5 and 36.5 GHz. To change switching states, the phase shift of the switch element changes 90°; and two elements are used with a hybrid to obtain a single-pole double-throw switch. M. Mohr and S. Monaghan described the design of a reflecting-type, circularly polarized phase shifter in which the signal passes twice through the ferrite to give approximately twice the one-pass phase shift. An X-band phased-array antenna system has been designed using 1300 of these phase shifters. Another example of a waveguide phase shifter was described by J. Parks, B. R. Savage, L. Lavendan, Jr., and J. Brown. Using ferrite loading for phase shifting and for increasing the cutoff frequency of the waveguide, a miniaturized C-band device was obtained. In addition, the authors devised a clever transition to couple the input TEM line with the modified TE waveguide mode used in the phase shifter. In contrast to the waveguide phase shifters, R. R. Jones described a ferrite phase shifter using a TEM meander line passing through a ferrite toroid.

In the absence of a slow-wave structure, the phase shift of a TEM wave in ferrite will not change when the direction of magnetization is reversed. Also, the phase shift will be reciprocal in this case. Therefore, the problem of making a reciprocal TEM latching phase shifter requires a novel approach. The method of J. Simon, W. Alverson, and J. Pipin is to change the direction of magnetization by 90° rather than 180° for switching. In a TEM stripline surrounded by ferrite, the magnetization was switched from longitudinal to transverse, which produced a phase-shifting efficiency of about 250° per decibel of insertion loss. Using field displacement, E. Schlomann, M. Harris, and J. Green designed a reciprocal latching phase shifter in a waveguide.

The forerunner of the reciprocal ferrite phase shifter was a structure developed by Reggia and Spencer and first reported in 1957. It employs a longitudinally magnetized ferrite rod in a rectangular waveguide and has often been the source of analytical curiosity. At this year's symposium, the R-S phenomenon was re-examined by C. R. Boyd using the model of a nonreciprocal Faraday-effect coupling of the vertically polarized TE mode to the cross-polarized TE mode in rectangular waveguide.

Material properties have always been of paramount importance to the ferrite designer. Therefore, a discussion, by J. J. Green, of the choice of ferrite materials for phase-shifting applications was extremely valuable. Both spinels and garnets were described.

For several years, a prize has been awarded at the microwave symposium for an outstanding paper on a microwave subject from an IEEE publication. The award this year went to H. Bosma for his contributions to Y-junction circulator analysis and design. Consistent with the quality of his prize-winning paper, Bosma gave an informative paper on three-port junction circulators. The effects of ferrite losses on circulator performance were noted.

Two papers dealt also with problems in the theory of three- and four-port circulators; they were by L. Davis and M. Omori respectively. The latter author treated the case in which a ferrite disk appears on only one side of the transmission line, with a metal disk being used in the image location.

In the latching circulator, which received considerable attention this year, the direction of circulation is changed (and latched) with a reversal of the ferrite magnetization. P. Goodman used latching circulators to accomplish a four-bit digital time delay in the X band. Other latching circulators were described by W. C. Passaro and J. W. McManus for 25 GHz and by F. Betts, D. H. Temme, and J. A. Weiss in the S band. Four-port circulator design was considered by C. Fay and W. Dean.

Conclusion

Attendance at the 1966 International Microwave Symposium certainly left no doubts about the immense challenge remaining to workers in this field. The microwave engineer is being led into solid-state phenomena, sophisticated filter design, optics, and ultrasonics. It is recognized that even the most frequently used microwave devices are often but partially understood. From the many who accept this challenge, a far-reaching set of papers can be expected for next year's symposium.